

# Do larger watersheds respond different to forest cover change? Re-analysing a global data set.

R. Willem Vervoort<sup>\*,a,1</sup>, Eliana Nervi<sup>\*\*,c</sup>, Jimena Alonso<sup>\*\*,d</sup>

<sup>a</sup>ARC Training Centre Data Analytics for Resources and the Environment

<sup>b</sup>School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

<sup>c</sup>Project Manager, FPTA 357 Instituto Nacional de Investigación Agropecuaria, INIA-Uruguay, Ruta 48 km 10, Rincon del Colorado, 90100 Canelones, Uruguay

<sup>d</sup>Institute of Fluid Mechanics and Environmental Engineering, School of Engineering, Universidad de la República, 11200 Montevideo, Uruguay

## Abstract

This is the abstract. It consists of two paragraphs.

## 1. Introduction

### 1.1. Introduction

There has been an long and on-going discussion in the hydrological literature around the impact of forests on streamflow (Andréassian, 2004; Brown et al., 2013, 2005; Filoso et al., 2017; Jackson et al., 2005; Zhang et al., 2017). The historic work highlights a general consensus that if forest areas increase, streamflow decreases and vice-versa. The most dramatic result in relation to this, is Figure 5 in Zhang et al. (2011) indicating (for Australian watersheds) a 100% decrease in streamflow for watersheds with 100% forest cover. However, on the other end of the spectrum, in a series of French watersheds (Cosandey et al., 2005), there was no change in streamflow characteristics in 2 of the three watersheds studied in relation to deforestation.

Several review papers have summarized different studies across the globe, in relation to paired watershed studies (Bosch and Hewlett, 1982; Brown et al., 2005), related to reforestation in particular (Filoso et al., 2017), and more generally (Jackson et al., 2005; Zhang et al., 2017). These studies aim to generalize the individual findings and to identify if there are global trends or relationships that can be developed. The most recent reviews (Filoso et al., 2017; Zhang et al., 2017) developed an impressive global database of watershed studies in relation to changes in streamflow due to changes in forest cover. The Zhang et al. (2017) dataset, which covers over 250 studies, is described in terms of the change in streamflow 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. (2017) 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

---

\*Corresponding Author

\*\*Equal contribution

Preprint submitted to *Water Resources Research* January 9, 2022  
Email addresses: [willerv@hpl.scg.sydney.edu.au](mailto:willerv@hpl.scg.sydney.edu.au) (R. Willem Vervoort), [elianaanervi@gmail.com](mailto:elianaanervi@gmail.com) (Eliana Nervi), [jalonso@fing.edu.uy](mailto:jalonso@fing.edu.uy) (Jimena Alonso)

38 coded as count data and only a subset of 37 studies was analysed for actual  
39 water yield change.

40 The conclusions of the first paper (Zhang et al., 2017) suggest that there is a  
41 distinct difference in the change in flow as a result of forestation or deforestation  
42 between small watersheds, defined as  $< 1000 \text{ km}^2$  and large watersheds  $> 1000$   
43  $\text{km}^2$ . While for small watersheds there was no real change in runoff with changes  
44 in cover, for large watersheds there was a clear trend showing a decrease in runoff  
45 with and increase in forest cover. Their main conclusion was that the response  
46 in annual runoff to forest cover was scale dependent and appeared to be more  
47 sensitive to forest cover change in water limited watersheds relative to energy  
48 limited watershed (Zhang et al., 2017).

49 The second study (Filoso et al., 2017) was a systematic review which classi-  
50 fied the historical research and highlighted gaps in the spatial distribution, the  
51 types of studies and the types of analysis. Their main conclusion was also that  
52 reforestation decreases streamflow, but that there were many interacting fac-  
53 tors. For a subset of quantitative data (37) they showed a relationship between  
54 watershed size and decline in streamflow.

55 A final summary paper that includes much of the same data as Zhang et  
56 al. (2017) and Filoso et al. (2017) is Zhou et al. (2015), which has one author  
57 in common with Zhang et al. (2017). However, this paper aims to explain the  
58 variation in the data using the Fuh model, and in particular aims to link the  
59 variation in the observed data to variations in the exponent  $m$  in the model.  
60 A key observation is that in drier environments, the effects of deforestation are  
61 much greater than in wetter environments, which is also suggested by Figure 4  
62 in Zhang et al. (2017).

63 Encouraged by the work presented by Zhang et al. (2017) and Filoso et  
64 al. (2017) and the fantastic database of studies presented by these authors, we  
65 believe we can add to the discussion. In this paper, the aim is to develop further  
66 analysis of the collected data and expanding and combining the two data sets  
67 to provide further depth.

68 In particular, the main method in the work by Zhang et al. (2017) is using  
69 simple linear regression, and in Filoso et al. (2017) the focus is mainly on  
70 classification. As Zhang et al. (2017) points out, the main assumption in their  
71 work is that the threshold at  $1000 \text{ km}^2$  is a distinct separation between “small”  
72 and “large” watersheds, but the subset of data in Filoso et al. (2017) does  
73 not appear to support this. And while the work Filoso et al. (2017) provides  
74 important insights in study types, analysis types and broad classification, there  
75 is limited quantification of actual impact. This is because the work had a strict  
76 criterion to select quantitative studies. However, given the fantastic data sets  
77 collected, the analyses can be easily expanded to look at interactions between  
78 the terms and to test the assumption of a distinct threshold at  $1000 \text{ km}^2$ .

79 As a result the objective of this paper is to 1) enhance the data set from  
80 Zhang et al. (2017) with further watersheds (such as from Filoso et al. (2017))  
81 and spatial coordinates and 2) to analyse the possibility of non-linear, interac-  
82 tions and partial effects of the different factors and variables in the data using  
83 generalised linear (GLM) and generalised additive models (GAM Wood (2006)).

Building on the analyses by Zhang et al. (2017) and Filoso et al. (2017), and combining their conclusions, the main hypothesis to test is that the change in streamflow is impacted by the change in forest cover. However, this change is clearly modulated by the area under consideration (affecting the length of the flowpaths Zhou et al. (2015)), the length of the study (c.f. Jackson et al. (2005)) and possibly the climate (as indicated by either E0/Pa or latitude and longitude Filoso et al. (2017); Zhou et al. (2015)).

However, there could be further confounding factors, which are eluded to by Filoso et al. (2017):

- the type of analysis, i.e. paired watershed studies, modelling, time series analysis etc.
- the age of the study, assuming that historical studies might not have had the ability to measure at the accuracy that currently is available to researchers, or that more careful historical attention to detail in field studies might have been lost more recently due to reductions in research investment.

Finally, this work aims to point to further research that can expand this area of work, based on the collected data, to better understand the impact of forest cover change on streamflow.

## 2. Methods

### 2.1. The original data sets

The starting point of this paper is the data base of studies which were included in Zhang et al. (2017) as supplementary material. The columns in this data set are the watershed number, the watershed name, the Area in km<sup>2</sup>, 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 streamflow in % ( $\Delta Qf\%$ ), based on equation 1 in Zhang et al. (2017)), the precipitation data type, the assessment technique, and the source of the info, which is a citation. Several of these columns contain abbreviations to describe the different variables, which are summarised in Table 1.

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

Factor	Abbreviation	Definition
precipitation data type	WL	water limited
	EQ	equitant
	OB	observed
	SG	spatial gridded
	MD	modelled
assessment technique	PWE	paired watershed experiment
	QPW	quasi-paired watershed experiment
	HM	hydrological modelling
	EA	elasticity analysis
	SH	combined use of statistical methods and hydrographs

While Zhang et al. (2017) use the dryness index in their analysis, potential or reference evapotranspiration was not originally included as part of the published data set. We combined the tables for small ( $< 1000 \text{ km}^2$ ) and large ( $\geq 1000 \text{ km}^2$ ) watershed data sets 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. (2017). In particular, we focussed on the 37 data points included in the quantitative analysis in Filoso et al. (2017).

In addition, additional variables added were the latitude and longitude for the center of the watershed as an approximation of its spatial location. Using this information reference evapotranspiration ( $E_0$ ) was extracted from the Global Aridity Index and Potential Evapo-Transpiration ( $ET_0$ ) Climate Databasev2 (Trabucco and Zomer, 2018), if a value of  $E_0$  was not available from the original papers. For large watersheds, this value, similar to annual average rainfall, is only an approximation of the climate at the location.

The length of the study can be a variable influencing the change in flow (e.g. Jackson et al., 2005), as for example, more mature plantations are thought to have smaller impacts on flow. 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. (2017).

Several additional data points from watershed studies were extracted from Zhang et al. (2011), Zhao et al. (2010), Borg et al. (1988), Thornton et al. (2007), Zhou et al. (2010), Rodriguez et al. (2010), Ruprecht et al. (1991) and Peña-Arancibia et al. (2012), and these were checked against the existing studies to prevent overlap. In the citation column in the data set, in general the main reference for the calculated change in streamflow was used, because sometimes the original study did not provide the quantification of the change in streamflow (i.e. Table 6 in Zhang et al. (2011)). We also removed one data point from the analysis, which corresponds to Watershed #1 (Amazon) in Zhang et

al. (2017). This is because the cited reference (Roche, 1981) only relates to 1 and 1.5 ha paired watershed studies in French Guyana, and in which the actual change in forest cover is not recorded.

The final column in the improved data set is a “notes” column, which is not further used in the analysis, but 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. (2017).

Similar to Zhang et al. (2017), the “dryness index” was calculated from the reference evapotranspiration and the annual average rainfall (Pa) as:

$$D = \frac{E_0}{Pa} \quad (1)$$

### 2.3. Statistical modelling

Removing possible duplicates removes 29 watersheds from the overall data set

To estimate how the change in streamflow is affected by the change in forest cover while considering the effects of the other variables, we applied generalised additive modelling (GAM) (Wood, 2006).

The general model tested is:

$$\Delta\%Qf \sim \Delta\%forest\ cover_{positive} + sign_{forest\ cover} + \sum X_i + \sum s(Z_i) + \varepsilon \quad (2)$$

Here  $X_i$  are factorial variables, while  $Z_i$  are continuous variables. The model assumes no direct interactions and all variables are additive. The changes in forest cover contain both positive (forestation) and negative values (deforestation). In Zhang et al. (2017), these changes were jointly analysed, assuming the effect on the change in flow was linear and the effect if removing forest cover was the same as an equivalent reforestation. However, the impact of an increase in forest cover can be different from the same fractional decrease in forest cover. Therefore all the change in forest cover data is converted to positive values, and an additional column ( $sign_{forestcover}$ ) is added that indicates whether it was a forest cover increase or decrease. A further assumption in the model is that all continuous variables  $Z_i$  (such as annual precipitation (Pa)) can have a linear or non-linear relationship with  $\Delta Qf\%$ . This means that a smooth function  $s()$  is applied to the  $Z_i$  variables. For the smoothing function we applied thin plate regression splines with an additional shrinkage penalty which means the terms can be shrunk to 0 if not significant (Wood, 2006).

For the model in equation 2, we initially only used the data from Zhang et al. (2017) to make sure that the additional watersheds added to the data set did not influence the results. Subsequently the analysis was repeated and the additionally identified watersheds were added.

More generally the results were analysed to identify:

1. the significance of the different variables
2. the direction of the categorical or shape of the smooth variables

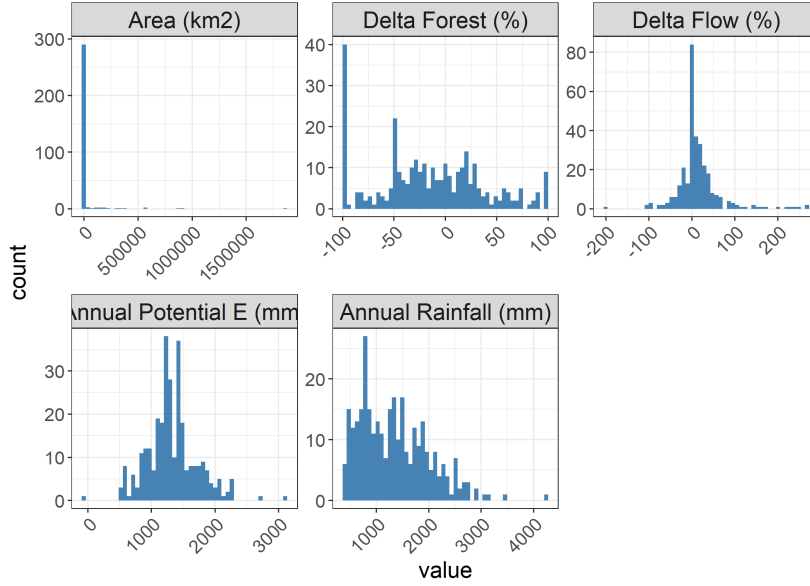


Figure 1: foo

### 3. Results

#### 3.1. description of the data

The overall dataset contains 309 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 dataset from Zhang et al. (2017) contained 312 observations. The overall distribution of changes in flow is highly skewed as is the distribution of changes in forest cover and Area. The values of changes in flow greater than 100% and smaller than -100% clearly create long tails on the change in flow distribution. Note also the large number of studies with 100% forest cover reduction. Smaller watersheds dominate the database with 42% of the data from watersheds  $< 1 \text{ km}^2$  and 65% of the data for watersheds  $< 10 \text{ km}^2$ .

This shows that for the data related to forest decreases, there is almost always a positive flow change. In other words, flow almost always increased. However, for increases in forest cover, this is not the case, and flow can both increase and decrease. However in both cases the variability in the reported change in flow increases with the increase in forest cover change.

#### 3.2. The general relationship between change in forest cover and streamflow

Following Zhang et al. (2017), the first step is to investigate the percent change in flow as a linear effect of the percent change forestry and modulated by the direction of the change, either an increase in forest cover, or decrease in forest cover:

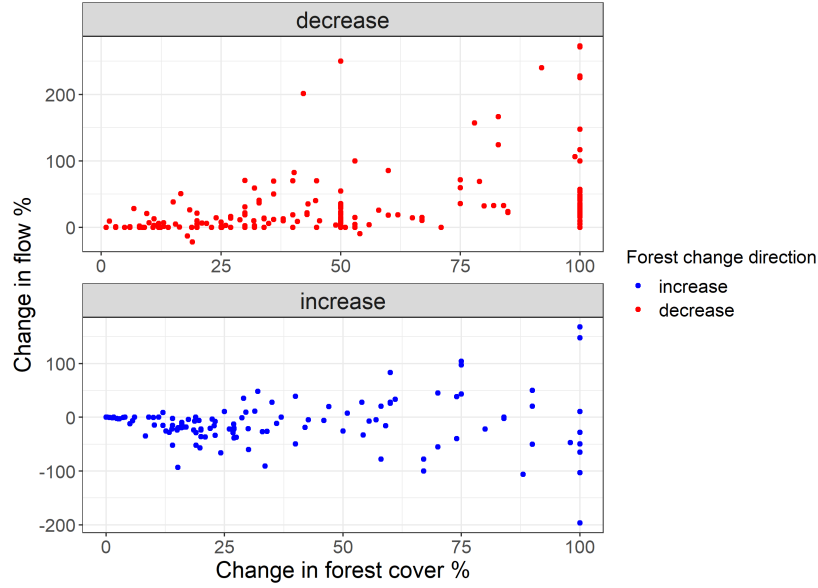


Figure 2: Changes in flow as a function of increases and decreases in forest cover

$$\Delta\%Qf \sim \Delta\%forest\ cover_{positive} + sign_{forest\ cover} + \varepsilon \quad (3)$$

Table 2: Summary results of the first regression model predicting change in streamflow from change in forest cover and accounting for the direction of the change

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	8.65	5.56	1.56	0.12
<b>DeltaF_perc_pos</b>	0.45	0.09	5.26	0
<b>Forest_Signincrease</b>	-29.17	5.79	-5.04	0

Table 3: Summary results of the first regression model predicting change in streamflow from change in forest cover and accounting for the direction of the change including the new data sets

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	9.43	5.54	1.7	0.09
<b>DeltaF_perc_pos</b>	0.44	0.09	5.12	0
<b>Forest_Signincrease</b>	-36.54	5.59	-6.53	0

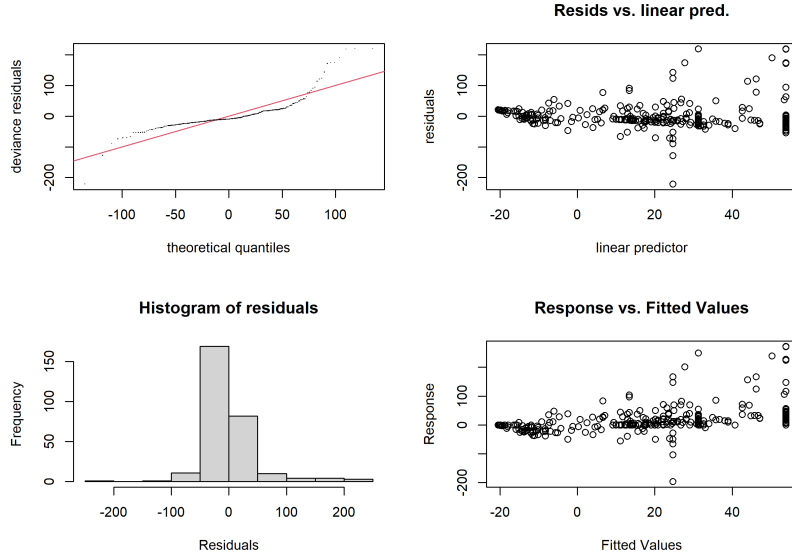


Figure 3: Residual plots for the first simple regression model indicating a slightly fat-tailed residual distribution

The overall variance explained in this model is not high with an adjusted  $r^2$  of 0.22, it generally supports the hypothesized relationship between the change in forest cover and the change in flow. The model suggests that for every 1% change in forest cover, on the average, the flow changes 0.45%. However the change in flow is different for forest cover decreases compared to forest cover increases. In fact, forest cover increases decrease flow by 29% less than a similar decrease in forest cover causes flow to increase. So roughly speaking, a 1% forest cover increase on the average decreases flow by  $(1 - 0.29) * 0.45\%$ , while a the percentage forest cover decrease will increase flow by 0.45%.

Of importance here is to highlight the residuals of this regression. These are approximately normal, although there is still significant skew on the upper and lower parts of the distribution (Figure 3). In other words, the distribution of the residuals is somewhat fat-tailed. We will discuss this later.

Including the data from some of the newly identified studies indicates that this mainly strengthens the difference between the forest cover increases and decreases (Table 2 and Table 3), and the result indication a reduction in the mean decrease in flow as a result of forest cover change if the new data is included.

It is however it is clear from the lack of explaining power for the model, that there could be confounding factors, as alluded to in the methods. The obvious ones being watershed dryness and area (following Zhang et al. (2017)):



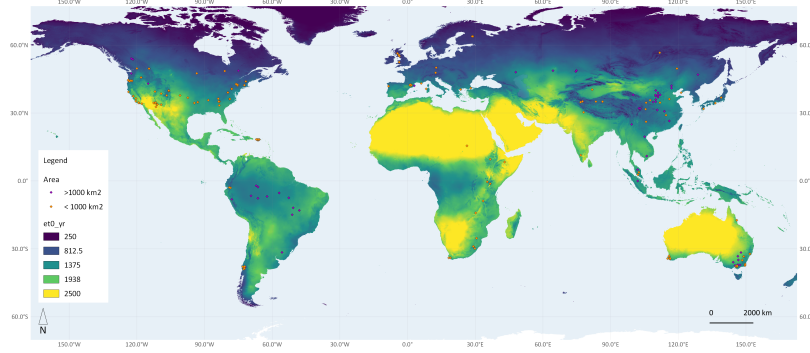


Figure 4: Distribution of included watersheds across the globe based on reported or estimated latitude and longitude

$$\Delta\%Qf \sim \Delta\%forest\ cover_{positive} + sign_{forest\ cover} + Pa_{mm} + \log_{10}(Area_{km^2}) + \varepsilon \quad (4)$$

Where  $Pa_{mm}$  is the annual average rainfall in mm, and a log base 10 transformation is applied to the area variable given that the distribution of Area is highly skewed (Figure 1)

Table 4: Summary of the second model, taking into account the annual rainfall and the area of the watershed

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	22.55	9.16	2.46	0.01
DeltaF_perc_pos	0.34	0.1	3.26	0
Forest_Signincrease	-35.52	5.67	-6.27	0
log10(Area_km2)	-3.12	1.72	-1.81	0.07
Pa_mm	0	0	-1	0.32

Including area and annual precipitation slightly improves the overall explaining power of the model (Table 4). Annual precipitation is in fact not significant. Relative to earlier reported studies (Filoso et al., 2017; Zhang et al., 2017), the log base 10 transformed watershed area indicates a p-value of only 0.07, suggesting a marginal impact on the change in stream flow. This supports our approach (in contrast to Zhang et al. (2017)) to consider watershed area as a continuous variable and making no separation between larger and smaller watersheds. The main effects in the model remain the change in forest cover and whether this is an increase or decrease.

### 3.3. The effect of location on the globe

As indicated, a further hypothesis relates to whether there is a strong spatial global gradient as captured by latitude and longitude. As the global map

(Figure 4) shows, the distribution of case study watersheds covers multiple continents and shows some distinct clustering in parts of the world. Of interest is whether the spatial clustering also indicates a difference in response to forest cover change:

$$\Delta\%Qf \sim \Delta\%forest\ cover_{positive} + sign_{forest\ cover} + Pa_{mm} + \log10(Area_{km^2}) + Latitude + Longitude + \varepsilon \quad (5)$$

Table 5: Results of the model including Latitude and Longitude including new data

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	23.29	9.96	2.34	0.02
<b>DeltaF_perc_pos</b>	0.35	0.1	3.32	0
<b>Forest_Signincrease</b>	-37.21	6	-6.2	0
<b>log10(Area_km2)</b>	-3.28	1.73	-1.89	0.06
<b>Pa_mm</b>	0	0	-0.83	0.41
<b>Latitude</b>	-0.05	0.09	-0.55	0.58
<b>Longitude</b>	0.01	0.03	0.23	0.82

There appears to be no significant gradient in either latitude or longitude (Table 5), suggesting that the distribution of the watersheds across the globe has little influence on the overall result. The total explaining power of the model is still low with an adjusted  $r^2$  of 0.22 suggesting further factors influencing the change in streamflow that are currently not included in the model.

Note that in this case the significance of the Area variable increases, and generally indicates that larger watersheds would be expected to have a lower change in streamflow, as also indicated in Zhang et al. (2017).

### 3.4. Impact of the dryness index

Climate, and in particular evapotranspiration can have a significant effect on the streamflow change as represented by the dryness index, which is also highlighted by both Zhang et al. (2017) and Jackson et al. (2005). Increased evapotranspiration could lead to drier watersheds, unless balanced by rainfall (such as possibly in the tropics). This model introduces the dryness index as a linear variable and drops the annual average precipitation as a variable, as dryness is calculated from the precipitation. It also drops the Latitude and Longitude as these are indicated not to be significant.

Table 6: Results of the model replacing the annual precipitation with the dryness index

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	10.46	7.66	1.36	0.17
<b>DeltaF_perc_pos</b>	0.31	0.11	2.93	0
<b>Forest_Signincrease</b>	-35.67	5.73	-6.23	0
<b>log10(Area_km2)</b>	-3.5	1.76	-1.99	0.05
<b>Dryness</b>	6.54	3.08	2.12	0.03

261 The results from this model (Table 6) confirm that dryness is a significant  
262 confounding factor related to the change in streamflow as a function of the  
263 change in forest cover change. In fact if the dryness index doubles (remembering  
264 that Dryness = 1 when  $E0 = Pa$ , so in this case  $E0 = 2*Pa$ , which is very  
265 dry), the change in runoff is ~14% greater. However, more interesting, Latitude  
266 remains a significant predictor with each degree in latitude causing an -0.31%  
267 change in runoff. This indicates that Dryness (i.e. an increase in radiation) alone  
268 does not explain the trend in the Latitude and some other unknown confounding  
269 factor is captured by Latitude.

Table 7: Watersheds for which the dryness index  $> 4$

Latitude	Longitude	Watershed name
34.67	-111.7	Beaver Creek, AZ #3-2
36.4	-120.4	Cantua
34.43	-112.3	White Spar, Ariz., U.S.A, B
32.74	-111.5	Natural DRDages, Ariz., U.S.A, A

270 However, the result also indicates possible issues with the data, some of the  
271 Dryness values are very large ( $> 4$ ) and these values have high leverage in the  
272 data. These watersheds are listed in Table 7.

#### 273 3.4.1. Are some of the variables possibly non-linear?

274 The work by Filoso et al. (2017) and earlier by Jackson et al. (2005) has  
275 indicated that the length of the study might influence the response. This links  
276 to the idea from Kuczera (1987) that the effect of logging or deforestation or  
277 reforestation reduces with the length of time post intervention (see also Jackson  
278 et al. (2005)). In addition to adding the length as a variable, all continuous  
279 variables (Dryness, Area, length and Latitude) were considered non-linear in  
280 this model and as indicated a shrinkage smoothing spline (Wood, 2006) was  
281 applied to these variables.

Table 8: Statistical summary for the linear terms in the model with non-linear terms

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	17.09	6.08	2.81	0.01
<b>DeltaF_perc_pos</b>	0.26	0.11	2.44	0.02
<b>Forest_Signincrease</b>	-34.6	6.11	-5.67	0

Table 9: Statistical summary for the smooth terms in the model with non-linear terms

	edf	Ref.df	F	p-value
<b>s(log10(Area_km2))</b>	1.74	9	0.78	0.01
<b>s(Dryness)</b>	0.87	9	0.57	0.02
<b>s(length)</b>	8.83	9	3.1	0

282 Including non-linearity increases the overall explaining power of the model  
283 to an adjusted  $r^2$  of 0.29 and deviance explained of 0.32, but creates few changes  
284 in the significance of the variables (Table ??). However, it also increases the  
285 chance of over fitting, as the smoothing splines allow significant flexibility, which  
286 will be investigated later.

287 A final full model which includes the remaining categorical variables (Precip-  
288 itation data type, Assessment technique, Forest type and Hydrological regime)  
289 is tested.

Table 10: Statistical summary for the linear terms the full model

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	-27.3	19.84	-1.38	0.17
<b>DeltaF_perc_pos</b>	0.31	0.1	3.01	0
<b>Forest_Signincrease</b>	-21.54	7.29	-2.96	0
<b>Precip_data_typeOB</b>	-8.13	15.17	-0.54	0.59
<b>Precip_data_typeSG</b>	17.33	17.29	1	0.32
<b>Assessment_techniqueEA,</b>	13.93	46.74	0.3	0.77
<b>HM</b>				
<b>Assessment_techniqueHM</b>	32.96	13.92	2.37	0.02
<b>Assessment_techniquePWE</b>	53.24	12.81	4.16	0
<b>Assessment_techniquePWE,</b>	33.54	47.07	0.71	0.48
<b>HM</b>				
<b>Assessment_techniqueQPW</b>	39.8	22.06	1.8	0.07
<b>Assessment_techniqueQPW,</b>	45.97	26.25	1.75	0.08
<b>EA</b>				
<b>Assessment_techniqueSH</b>	44.89	13.49	3.33	0

	Estimate	Std. Error	t value	Pr(> t )
<b>Forest_typeCF</b>	-2.2	8.52	-0.26	0.8
<b>Forest_typeMF</b>	-6.56	8.96	-0.73	0.46
<b>Hydrological_regimeSD</b>	4.59	10.25	0.45	0.65

```
pander(round(summary(model6_all)$s.table,2), caption = "Statistical summary for the smooth t
```

Table 11: Statistical summary for the smooth terms for the full model

	edf	Ref.df	F	p-value
<b>s(log10(Area_km2))</b>	0.31	9	0.07	0.16
<b>s(Dryness)</b>	3.45	9	1.39	0.01
<b>s(length)</b>	8.73	9	3.04	0

```
#plot(model6_all)
```

290 This model explains more of the variance, but the improvement is marginal  
291 compared to the previous model with a deviance explained of 0.38. This indi-  
292 cates that the categorical variables explain a limited amount of the variance.  
293 However, interesting to note from Table ?? that several of the assessment meth-  
294 ods are significant. In particular Paired Watersheds experiments (PWE), Hydro-  
295 logical modelling (HM) and Statistical techniques (SH) are strongly significant  
296 ( $p < 0.05$ ).

297 Figure 5 highlights that the relationship between  $\log_{10}(\text{Area})$  and the change  
298 in flow is essentially linear and does not need to be smoothed (this is the value  
299 of using penalized smooths following Wood (2006)). It is still a negative slope,  
300 indicating that in larger watersheds the impact of changes in forest cover on  
301 streamflow is less than for smaller watersheds. In addition, both the length and  
302 Dryness variables show non-linearity, but this does not show a clear trend.

303 Quitting from lines 645-646 (Forest\_and\_Water.Rmd) Error in summary(model7\_nolongLength)  
304 : object 'model7\_nolongLength' not found Calls: ... withCallingHandlers -  
305 > withVisible -> eval -> eval -> pander -> summary In addition: Warning  
306 messages: 1: In eval(expr, envir, enclos) : NAs introduced by coercion 2: In  
307 eval(expr, envir, enclos) : NAs introduced by coercion

308 The flexible nature of the splines means that the length variable captures  
309 some substantial variation in the data, but it is unclear what exactly is captured.  
310 The shape of the conditional response (Figure 5) also does not reflect the type  
311 of response highlighted in Filoso et al. (2017) and Jackson et al. (2005). One  
312 reason could be the few data points with very long data series, but highly  
313 variable responses (Figure 5). Reducing the flexibility of the splines, removing  
314 any studies longer than 60 years (Table ??), or fitting a linear term results in

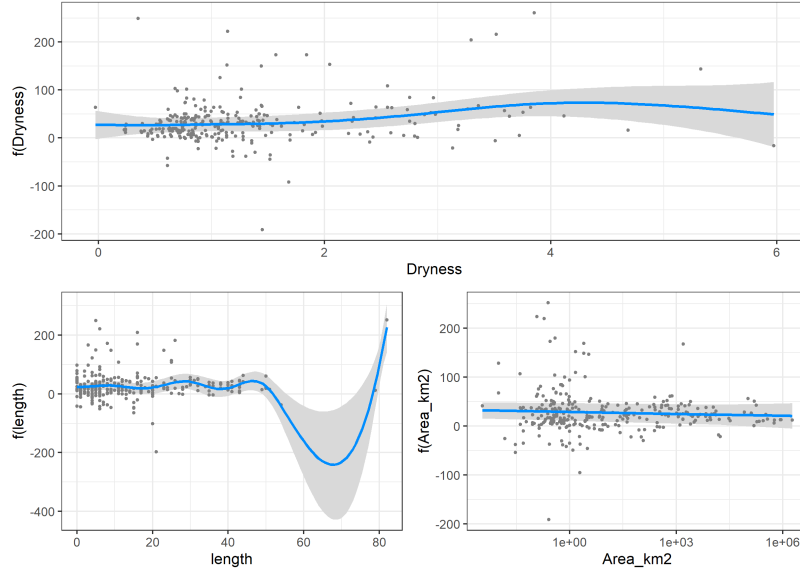


Figure 5: Visualisation of the smooth variables in the model

315 “length” not being significant, suggesting that this variable is not helpful in  
 316 explaining the variation in the data.

```
Zhang_all12 %>%
  ggplot(aes(length)) + geom_histogram(fill="blue", bins =50)
```

#### 317 4. Discussion

318 Essentially, the analysis shows at the moment that in contrast to Zhang et  
 319 al. (2017) there is no evidence that the size of a watershed influences the change  
 320 in the streamflow as a result of changes in forestry. If anything the scatter in  
 321 the data (in the change in flow) is greater for the smaller watersheds than for  
 322 the larger watersheds. In other words, the response to changes in forest cover  
 323 is more consistent for larger watersheds than it is for smaller watersheds.

324 As shown earlier, most of the smaller watersheds are “real observed data”  
 325 using paired watershed studies, while for larger watersheds, the analysis are  
 326 mostly based on modelling approximations using either elasticity analysis (EA),  
 327 Hydrological modelling (HM) or a combined use of statistical methods (SH) or  
 328 quasi paired watershed analysis (QPW), thus all providing an approximation of  
 329 the effect of forestry on streamflow rather than a direct comparison of water-  
 330 sheds. This is a confounding factor that is not easily addressed in the regression  
 331 modelling attempted here. Furthermore, the catchments analysed using EA,  
 332 are concentrated in the drier end of the Dryness index scale compared to the

other methods, with only the paired watershed experiment (PWE) assessment technique covering the full range of dryness indices.

#### 4.0.1. remove the assessment techniques with very small numbers

One concern is that there are a few Assessment techniques in the original dataset with a very low number of observations and this might skew 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).

Assessment_technique	n
PWE	185
HM	53
EA	32
SH	26
QPW	7
QPW, EA	4
EA, HM	1
PWE, HM	1

## pdf  
## 2

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-18.71	19.02	-0.98	0.33
DeltaF_perc_pos	0.29	0.1	2.83	0
Forest_Signincrease	-23.26	7.15	-3.25	0
Precip_data_typeOB	-11.25	14.19	-0.79	0.43
Precip_data_typeSG	18.43	16.43	1.12	0.26
Assessment_techniqueHM	32.02	12.76	2.51	0.01
Assessment_techniquePWE	44.48	12.97	3.43	0
Assessment_techniqueQPW	38.12	21.21	1.8	0.07
Assessment_techniqueSH	46.79	13.17	3.55	0
Forest_typeCF	-3.3	8.23	-0.4	0.69
Forest_typeMF	-3.88	8.69	-0.45	0.66
Hydrological_regimeSD	5.73	10.11	0.57	0.57

	edf	Ref.df	F	p-value
s(Dryness)	3.51	9	1.96	0
s(log10(Area_km2))	0.7	9	0.24	0.08
s(length)	0	9	0	0.99

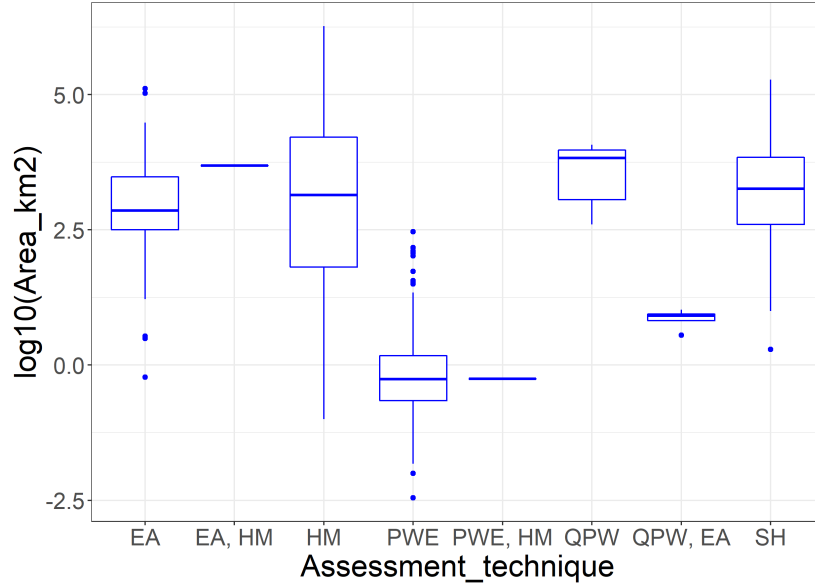


Figure 6: Boxplot of the log base 10 of the watershed area (in km<sup>2</sup>) for the different assessment techniques, showing the dominance of small watersheds in the paired watershed experiments

In drier watersheds, changes in forest cover have greater impact on flow, which is similar to Zhang et al. (2017). This is most likely because in these watersheds the overall flow is surface flow dominated and therefore the buffering that is afforded by the groundwater inputs is not as great. As we don't have a separate variable for groundwater inputs (although this effect is estimated in many studies), we cannot analyse this effect separately.

In contrast to Filoso et al. (2017), we also did not identify an effect of the Given how skewed Dryness is due to the few watersheds that have very high dryness values, it is worth investigating what excluding these 4 watersheds from the data means for the relationships.

#### 4.0.2. Other possible bias in the data

There are further confounding factors in the data, which were also classified by Filoso et al. (2017) and these create biases in the data set that can impact the overall assessment. For example, snow dominated hydrological regimes (SD), which are weakly significant, are dominated by Coniferous Forests (CF), while the majority of the rain dominated regimes are all broadleaf forests (BF). However, the forest type classification is very coarse and does not fully capture possible physiological differences that could affect evapotranspiration and therefore changes in streamflow.

Apart from a difficulty of analysing complex confounding factors in the data, a general limitation of the type of analysis presented is that this work does not consider the spatial arrangement of the forest clearing in the catchments. While



for fully or almost fully cleared smaller catchments this might not be an issue, it is perceivable that for larger catchments being partially cleared, a interaction between spatial location and clearing could be a factor in determining the change in streamflow. Clearing head water catchments on shallower soils might have a larger impact than clearing in downstream areas on deeper soils.

## References

- Andréassian, V., 2004. Waters and forests: From historical controversy to scientific debate. *Journal of Hydrology* 291, 1–27. doi:<https://doi.org/10.1016/j.jhydrol.2003.12.015>
- Borg, H., Bell, R.W., Loh, I.C., 1988. Streamflow and stream salinity in a small water supply catchment in southwest western australia after reforestation. *Journal of Hydrology* 103, 323–333. doi:[https://doi.org/10.1016/0022-1694\(88\)90141-2](https://doi.org/10.1016/0022-1694(88)90141-2)
- Bosch, J.M., Hewlett, J.D., 1982. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology* 55, 3–23.
- Brown, A.E., Western, A.W., McMahon, T.A., Zhang, L., 2013. Impact of forest cover changes on annual streamflow and flow duration curves. *Journal of Hydrology* 483, 39–50. doi:<http://dx.doi.org/10.1016/j.jhydrol.2012.12.031>
- Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W., Vertessy, R.A., 2005. A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology* 310, 28–61.
- Cosandey, C., Andréassian, V., Martin, C., Didon-Lescot, J.F., Lavabre, J., Folton, N., Mathys, N., Richard, D., 2005. The hydrological impact of the mediterranean forest: A review of french research. *Journal of Hydrology* 301, 235–249. doi:<https://doi.org/10.1016/j.jhydrol.2004.06.040>
- Filoso, S., Bezerra, M.O., Weiss, K.C.B., Palmer, M.A., 2017. Impacts of forest restoration on water yield: A systematic review. *PLOS ONE* 12, e0183210. doi:[10.1371/journal.pone.0183210](https://doi.org/10.1371/journal.pone.0183210)
- Jackson, R.B., Jobbagy, E.G., Avissar, R., Roy, S.B., Barrett, D.J., Cook, C.W., Farley, K.A., Maitre, D.C. le, McCarl, B.A., Murray, B.C., 2005. Trading water for carbon with biological carbon sequestration. *Science* 310, 1944–1947. doi:[10.1126/science.1119282](https://doi.org/10.1126/science.1119282)
- Kuczera, G., 1987. Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest. *Journal of Hydrology* 94, 215–236. doi:[10.1016/0022-1694\(87\)90054-0](https://doi.org/10.1016/0022-1694(87)90054-0)
- Peña-Arancibia, J.L., Dijk, A.I.J.M. van, Guerschman, J.P., Mulligan, M., Bruijnzeel, L.A., McVicar, T.R., 2012. Detecting changes in streamflow after partial woodland clearing in two large catchments in the seasonal tropics. *Journal of Hydrology* 416–417, 60–71. doi:<https://doi.org/10.1016/j.jhydrol.2011.11.036>
- Roche, M., 1981. Watershed investigations for development of forest resources of the amazon region in french guyana. *Tropical Agricultural Hydrology*. J 75–82.

409 Rodriguez, D.A., Tomasella, J., Linhares, C., 2010. Is the forest conversion to  
 410 pasture affecting the hydrological response of amazonian catchments? Sig-  
 411 nals in the ji-paraná basin. *Hydrological Processes* 24, 1254–1269. doi:https:  
 412 //doi.org/10.1002/hyp.7586  
 413 Ruprecht, J.K., Schofield, N.J., Crombie, D.S., Vertessy, R.A., Stoneman, G.L.,  
 414 1991. Early hydrological response to intense forest thinning in southwestern  
 415 australia. *Journal of Hydrology* 127, 261–277. doi:https://doi.org/10.1016/  
 416 0022-1694(91)90118-2  
 417 Thornton, C.M., Cowie, B.A., Freebairn, D.M., Playford, C.L., 2007. The briga-  
 418 low catchment study: II\*. Clearing brigalow (*acacia harpophylla*) for crop-  
 419 ping or pasture increases runoff. *Australian Journal of Soil Research* 45,  
 420 496–511. doi:doi:10.1071/SR07064  
 421 Trabucco, A., Zomer, R.J., 2018. Global aridity index and potential evapo-  
 422 transpiration (ET0) climate database v2. CGIAR consortium for spatial  
 423 information(CGIAR-CSI).  
 424 Wood, S., 2006. Generalized additive models: An introduction with r. CRC  
 425 Press, Boca Raton, FL.  
 426 Zhang, L., Zhao, F., Chen, Y., Dixon, R.N.M., 2011. Estimating effects of plan-  
 427 tation expansion and climate variability on streamflow for catchments in aus-  
 428 tralia. *Water Resources Research* 47, W12539. doi:10.1029/2011wr010711  
 429 Zhang, M., Liu, N., Harper, R., Li, Q., Liu, K., Wei, X., Ning, D., Hou, Y.,  
 430 Liu, S., 2017. A global review on hydrological responses to forest change  
 431 across multiple spatial scales: Importance of scale, climate, forest type and  
 432 hydrological regime. *Journal of Hydrology* 546, 44–59. doi:https://doi.org/  
 433 10.1016/j.jhydrol.2016.12.040  
 434 Zhao, F., Zhang, L., Xu, Z., Scott, D.F., 2010. Evaluation of methods for esti-  
 435 mating the effects of vegetation change and climate variability on streamflow.  
 436 *Water Resources Research* 46, W03505. doi:10.1029/2009wr007702  
 437 Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., Sun, G., Scott,  
 438 D.F., Zhou, S., Han, L., Su, Y., 2015. Global pattern for the effect of  
 439 climate and land cover on water yield. *Nature Communications* 6, 5918.  
 440 doi:10.1038/ncomms6918  
 441 Zhou, G., Wei, X., Luo, Y., Zhang, M., Li, Y., Qiao, Y., Liu, H., Wang, C.,  
 442 2010. Forest recovery and river discharge at the regional scale of guangdong  
 443 province, china. *Water Resources Research* 46. doi:https://doi.org/10.1029/  
 444 2009WR008829