

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

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

4 <sup>a</sup>*School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006,*  
5 *Australia*

6 <sup>b</sup>*Instituto Nacional de Investigación Agropecuaria, INIA-Uruguay, Ruta 48 km 10, Rincon*  
7 *del Colorado, 90100 Canelones, Uruguay*

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

10 **Abstract**

This is the abstract. It consists of two paragraphs.

11 **Introduction**

12 *Introduction*

13 There has been an long and on-going discussion in the hydrological litera-  
14 ture around the impact of forests on streamflow (Andréassian, 2004; Brown et  
15 al., 2013, 2005; Filoso et al., 2017; Jackson et al., 2005; Zhang et al., 2017).  
16 The historic work highlights a general consensus that if forest areas increase,  
17 streamflow decreases and vice-versa. The most dramatic result in relation to  
18 this, is Figure 5 in Zhang et al. (2011) indicating (for Australian watersheds) a  
19 100% decrease in streamflow for watersheds with 100% forest cover. However,  
20 on the other end of the spectrum, in a series of French watersheds (Cosandey  
21 et al., 2005), there was no change in streamflow characteristics in 2 of the three  
22 watersheds studied in relation to deforestation.

23 Several review papers have summarized different studies across the globe, in  
24 relation to paired watershed studies (Bosch and Hewlett, 1982; Brown et al.,  
25 2005), related to reforestation in particular (Filoso et al., 2017), and more gen-  
26 erally (Jackson et al., 2005; Zhang et al., 2017). These studies aim to generalize  
27 the individual findings and to identify if there are global trends or relationships  
28 that can be developed. The most recent reviews (Filoso et al., 2017; Zhang  
29 et al., 2017) developed an impressive global database of watershed studies in  
30 relation to changes in streamflow due to changes in forest cover. The Zhang et  
31 al. (2017) dataset, which covers over 250 studies, is described in terms of the  
32 change in streamflow as a result of the change in forest cover, where studies  
33 related to both forestation (increase in forest cover) and deforestation (decrease  
34 in forest cover) were included. In contrast, the paper by Filoso et al. (2017) fo-  
35 cused primarily on reforestation, and covered an equally impressive database of  
36 167 studies using a systematic review. In this case the collected data is mostly  
37 coded as count data and only a subset of 37 studies was analysed for actual  
38 water yield change.

---

\*Corresponding Author  
Preprint submitted to Journal of Hydrology

\*\*Equal contribution

December 3, 2021

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

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

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

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

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

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

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

83 Building on the analyses by Zhang et al. (2017) and Filoso et al. (2017),  
84 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.

## Methods

### *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
	WL	water limited
	EQ	equitant

Factor	Abbreviation	Definition
precipitation data type	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.

#### *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.

148 The final column in the improved data set is a “notes” column, which is not  
 149 further used in the analysis, but gives context to some of the data for future  
 150 research and highlights some of the discrepancies that we found between the  
 151 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:

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

## 152 *Statistical modelling*

153 Removing possible duplicates removes 29 watersheds from the overall data  
 154 set

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

158 The general model tested is:

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

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

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

177 More generally the results were analysed to identify:

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

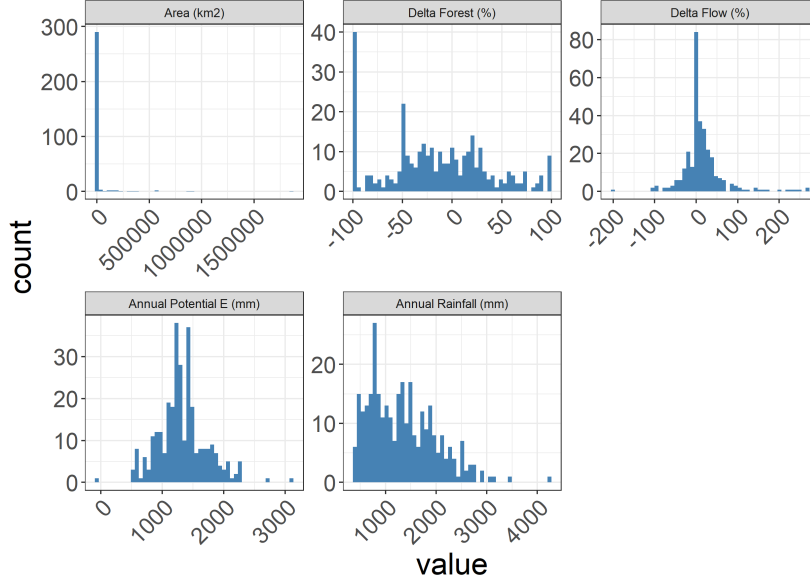


Figure 1: Overview of the distributions of some of the variables in the data set

## Results

### *description of the data*

The overall dataset contains 312 observations of changes in flow. 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 km<sup>2</sup> and 65% of the data for watersheds < 10 km<sup>2</sup>.

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.

### *The initial relationship between change in forest cover and streamflow*

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

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

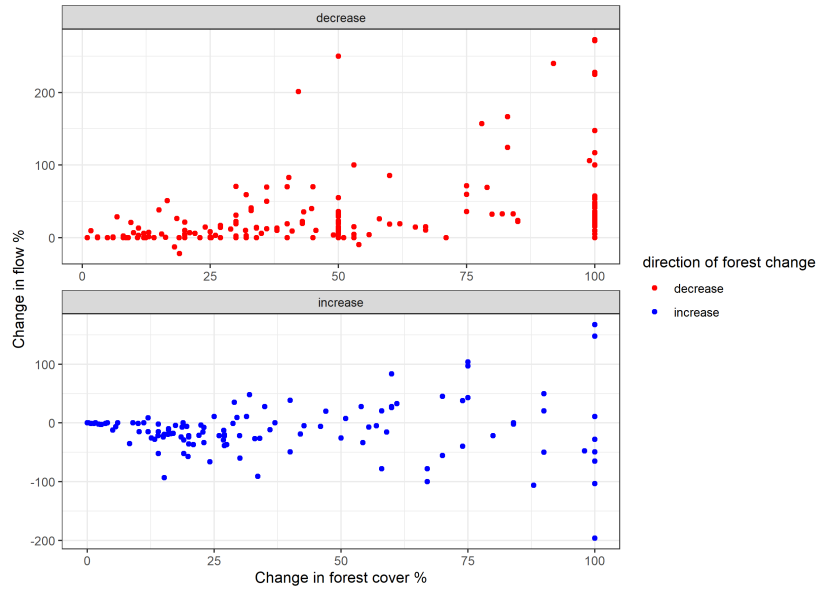


Figure 2: Changes in flow as a function of increases and decreases in forest cover

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

199 While the overall variance explained in this model is not high with an ad-  
200 justed  $r^2$  of 0.18, it clearly supports the hypothesized relationship between the  
201 change in forest cover and the change in flow. The model suggests that for every

202 1% change in forest cover, on the average, the flow changes 0.45%. However the  
 203 change in flow is different for forest cover decreases compared to forest cover  
 204 increases. In fact, forest cover increases decrease flow by 29% less than a similar  
 205 decrease in forest cover causes flow to increase. So roughly speaking, a 1% forest  
 206 cover increase on the average decreases flow by  $(1 - 0.29) * 0.45\%$ , while a the  
 207 percentage forest cover decrease will increase flow by 0.45%.

208 Checking this result against including the new data indicates that this mainly  
 209 strengthens the difference between the forest cover increases and decreases, and  
 210 decreases the mean decrease in flow as a result of forest cover change

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

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

214 Where  $Pa_{mm}$  is the annual average rainfall in mm, and a log base 10 trans-  
 215 formation is applied to the area variable given that the distribution of Area is  
 216 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

217 Including area and annual precipitation slightly improves the overall explain-  
 218 ing power of the model. Annual precipitation is in fact not significant. Relative  
 219 to earlier reported studies (Filoso et al., 2017; Zhang et al., 2017), the log base  
 220 10 transformed watershed area indicates a p-value of only 0.07, suggesting a  
 221 marginal impact on the change in stream flow. This supports our approach (in  
 222 contrast to Zhang et al. (2017)) to consider watershed area as a continuous  
 223 variable and making no separation between larger and smaller watersheds The  
 224 main effects remain the change in forest cover and whether this is an increase  
 225 or decrease.

#### 226 *The effect of location on the globe*

227 As indicated, a further hypothesis relates to whether there is a strong spa-  
 228 tial global gradient as captured by latitude and longitude. As the global map  
 229 (@ref(fig:global\_map)) shows, the distribution of case study watersheds covers  
 230 multiple continents and shows some distinct clustering in parts of the world. Of



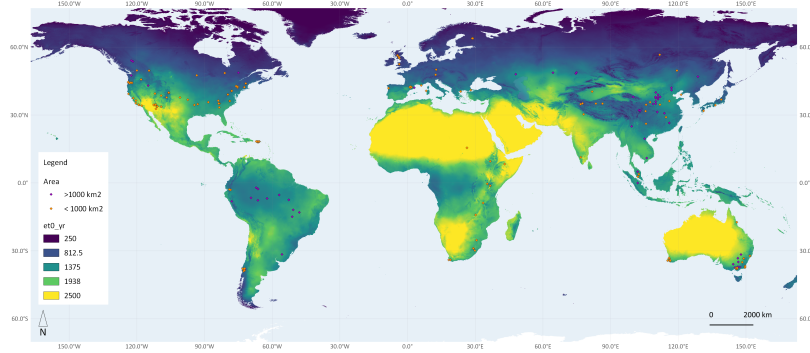


Figure 3: Distribution of included watersheds across the globe based on reported or estimated latitude and longitude

231 interest is whether the spatial clustering also indicates a difference in response  
 232 to forest cover change:

$$\Delta\%Q \sim \Delta\%forest\ cover_{positive} + sign_{forest\ cover} + Pa_{mm} + \log_{10}(Area_{km^2}) + Latitude + Longitude + \varepsilon \quad (5)$$

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

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

233 There appears to be no significant latitudinal gradient. The total explaining  
 234 power of the model is still low with an adjusted  $r^2$  of 0.22 suggesting further  
 235 factors that are currently not included in the model.

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

#### 239 *Impact of the dryness index*

240 Climate, and in particular evapotranspiration would be expected to have a  
 241 significant effect on the streamflow change as represented by the dryness index,

242 which is also indicated in Zhang et al. (2017). Increased evapotranspiration  
 243 could lead to drier watersheds, unless balanced by rainfall (such as possibly in  
 244 the tropics). This model introduces the dryness index as a linear variable and  
 245 drops the annual average precipitation as a variable, as dryness is calculated  
 246 from the precipitation. It also drops the Latitude and Longitude as these are  
 247 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

248 The results from this model confirm that dryness is a significant confounding  
 249 factor of the change in streamflow as function of the change in forest cover  
 250 change. In fact if the dryness index doubles (remembering that Dryness = 1  
 251 when  $E0 = Pa$ , so in this case  $E0 = 2*Pa$ , which is very dry), the change in runoff  
 252 is ~14% greater. However, more interesting, Latitude remains a significant  
 253 predictor with each degree in latitude causing an -0.31% change in runoff. This  
 254 indicates that Dryness (i.e. an increase in radiation) alone does not explain the  
 255 trend in the Latitude and some other unknown confounding factor is captured  
 256 by Latitude.

257 However, the result also indicates possible issues with the data, some of the  
 258 Dryness values are very large ( $> 4$ ) and these values have high leverage in the  
 259 data. These watersheds are listed in Table 5:

260 Quitting from lines 461-462 (Forest\_and\_Water.Rmd) Error in filter(., Dry-  
 261 ness > 4) : object 'Zhang\_all2' not found Calls: ... withVisible -> eval ->  
 262 eval -> pander -> %>% -> select -> filter In addition: Warning messages: 1:  
 263 In eval(expr, envir, enclos) : NAs introduced by coercion 2: In eval(expr, envir,  
 264 enclos) : NAs introduced by coercion

265 *Are some of the variables possibly non-linear?*

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

```

274 ##
275 ## Family: gaussian
276 ## Link function: identity
277 ##
278 ## Formula:
279 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + s(log10(Area_km2),
280 ##      bs = "ts") + s(Dryness, bs = "ts") + s(length, bs = "ts")
281 ##
282 ## Parametric coefficients:
283 ##              Estimate Std. Error t value Pr(>|t|)
284 ## (Intercept)      17.0942      6.0776   2.813  0.00526 **
285 ## DeltaF_perc_pos    0.2606      0.1067   2.442  0.01521 *
286 ## Forest_Signincrease -34.5961      6.1052  -5.667 3.59e-08 ***
287 ## ---
288 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
289 ##
290 ## Approximate significance of smooth terms:
291 ##              edf Ref.df      F p-value
292 ## s(log10(Area_km2)) 1.7399      9 0.782 0.01465 *
293 ## s(Dryness)         0.8731      9 0.570 0.01514 *
294 ## s(length)          8.8276      9 3.097 0.00114 **
295 ## ---
296 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
297 ##
298 ## R-sq.(adj) =  0.287   Deviance explained =  32%
299 ## GCV = 2171.2   Scale est. = 2065.7      n = 297

```

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

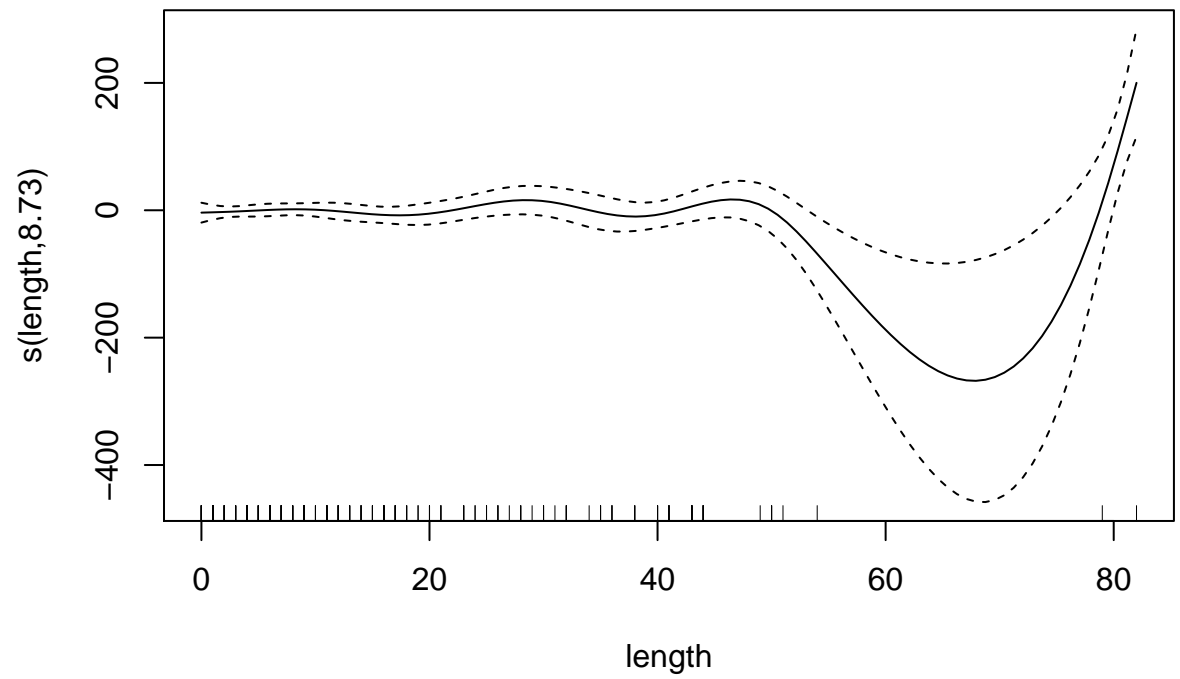
	edf	Ref.df	F	p-value
s(log10(Area_km2))	1.74	9	0.78	0.01
s(Dryness)	0.87	9	0.57	0.02
s(length)	8.83	9	3.1	0

300 Including non-linearity increases the overall explaining power of the model  
301 to an adjusted  $r^2$  of 0.29, but creates few changes in the significance of the  
302 variables. However, it also can create a chance of overfitting, as the smoothing  
303 splines allow significant flexibility, which will be investigated below.

304 However, first a final full model which includes the remaining categorical  
305 variables (Precipitation data type, Assessment technique, Forest type and Hy-  
306 drological regime) are included in the model

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-27.3	19.84	-1.38	0.17
DeltaF_perc_pos	0.31	0.1	3.01	0
Forest_Signincrease	-21.54	7.29	-2.96	0
Precip_data_typeOB	-8.13	15.17	-0.54	0.59
Precip_data_typeSG	17.33	17.29	1	0.32
Assessment_techniqueEA, HM	13.93	46.74	0.3	0.77
Assessment_techniqueHM	32.96	13.92	2.37	0.02
Assessment_techniquePWE	53.24	12.81	4.16	0
Assessment_techniquePWE, HM	33.54	47.07	0.71	0.48
Assessment_techniqueQPW	39.8	22.06	1.8	0.07
Assessment_techniqueQPW, EA	45.97	26.25	1.75	0.08
Assessment_techniqueSH	44.89	13.49	3.33	0
Forest_typeCF	-2.2	8.52	-0.26	0.8
Forest_typeMF	-6.56	8.96	-0.73	0.46
Hydrological_regimeSD	4.59	10.25	0.45	0.65

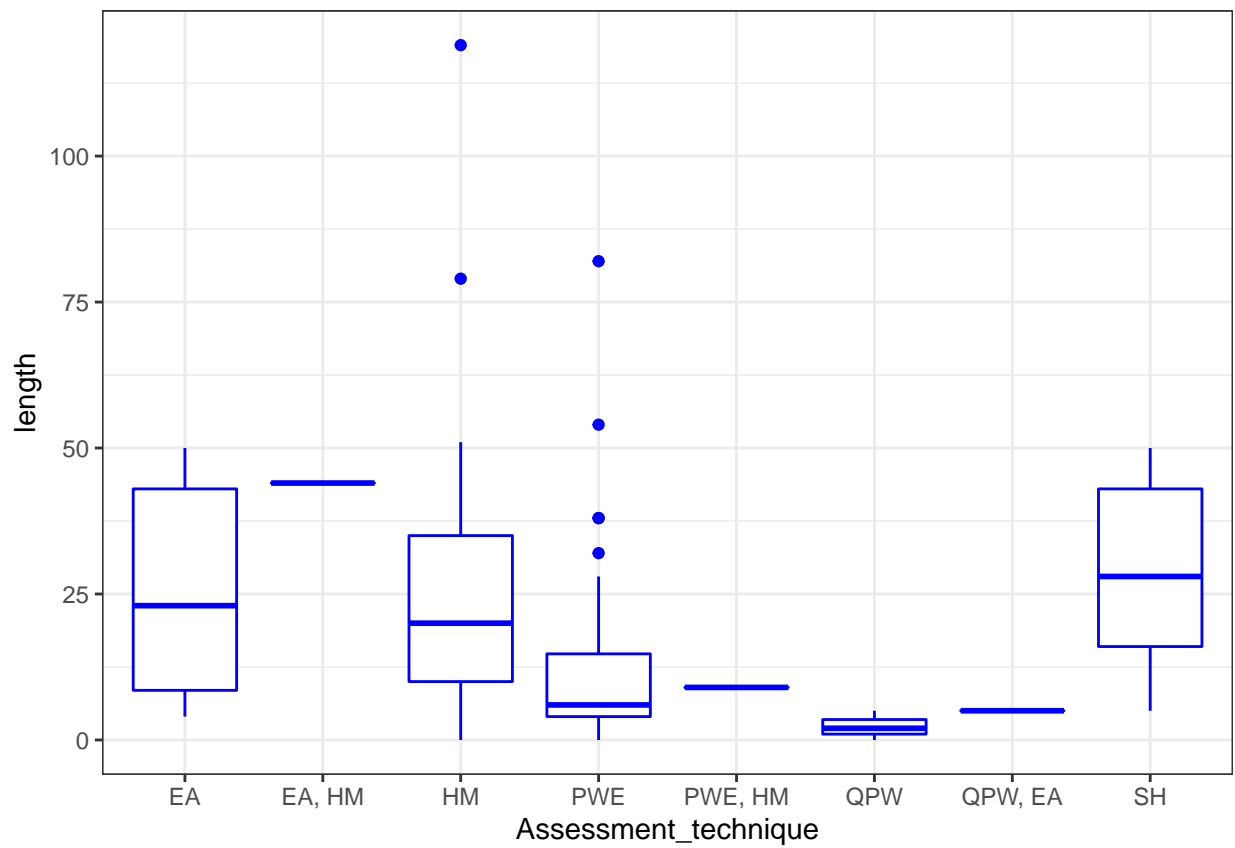
	edf	Ref.df	F	p-value
s(log10(Area_km2))	0.31	9	0.07	0.16
s(Dryness)	3.45	9	1.39	0.01
s(length)	8.73	9	3.04	0



```

307
308     Clearly Latitude is masking other factors including the assessment technique
309     and the forest type
310 ## Warning: Removed 5 rows containing non-finite values (stat_boxplot).

```



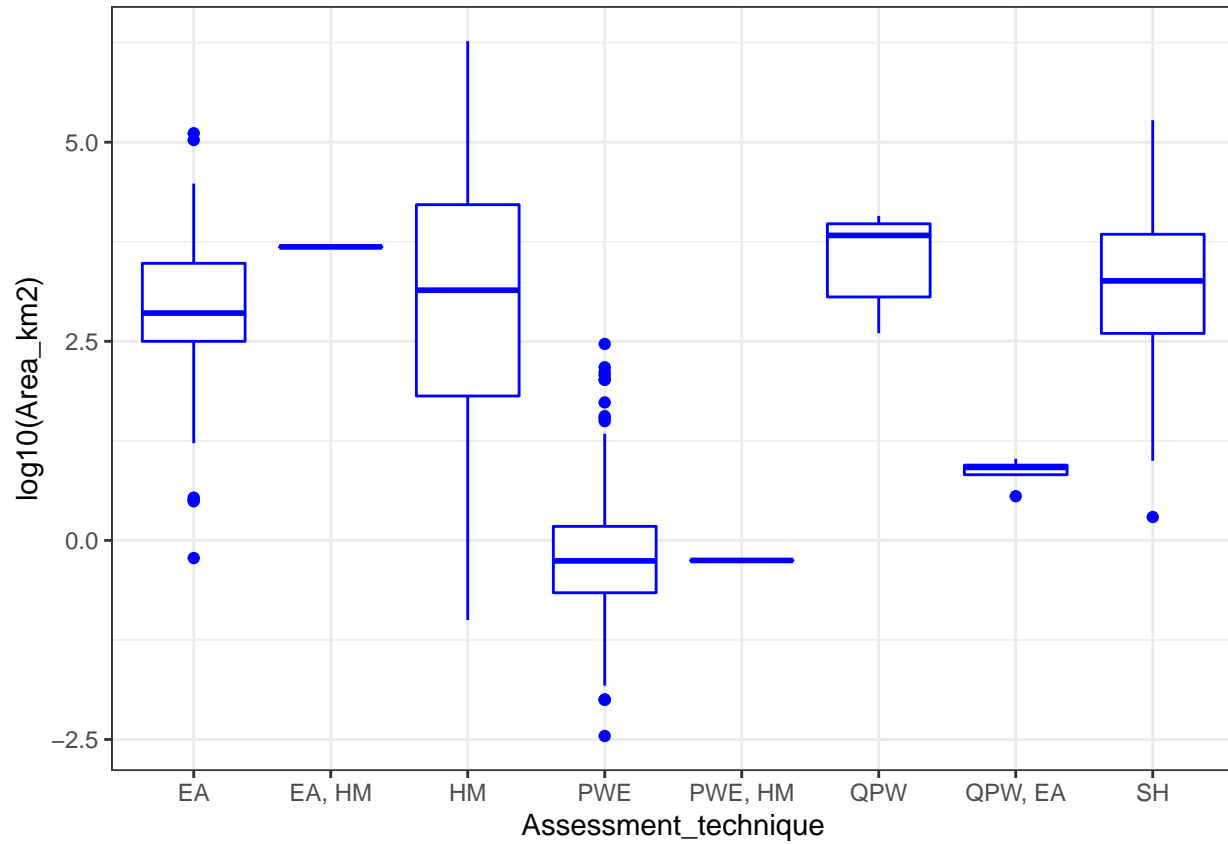
311

```

312 ## # A tibble: 8 x 2
313 ##   Assessment_technique     n
314 ##   <chr>                <int>
315 ## 1 EA                    32
316 ## 2 EA, HM                 1
317 ## 3 HM                    53
318 ## 4 PWE                   185
319 ## 5 PWE, HM                1
320 ## 6 QPW                    7
321 ## 7 QPW, EA                4
322 ## 8 SH                    26

```

323 Maybe remove data for low numbers of assessment techniques and rerun the  
324 analysis (<10)

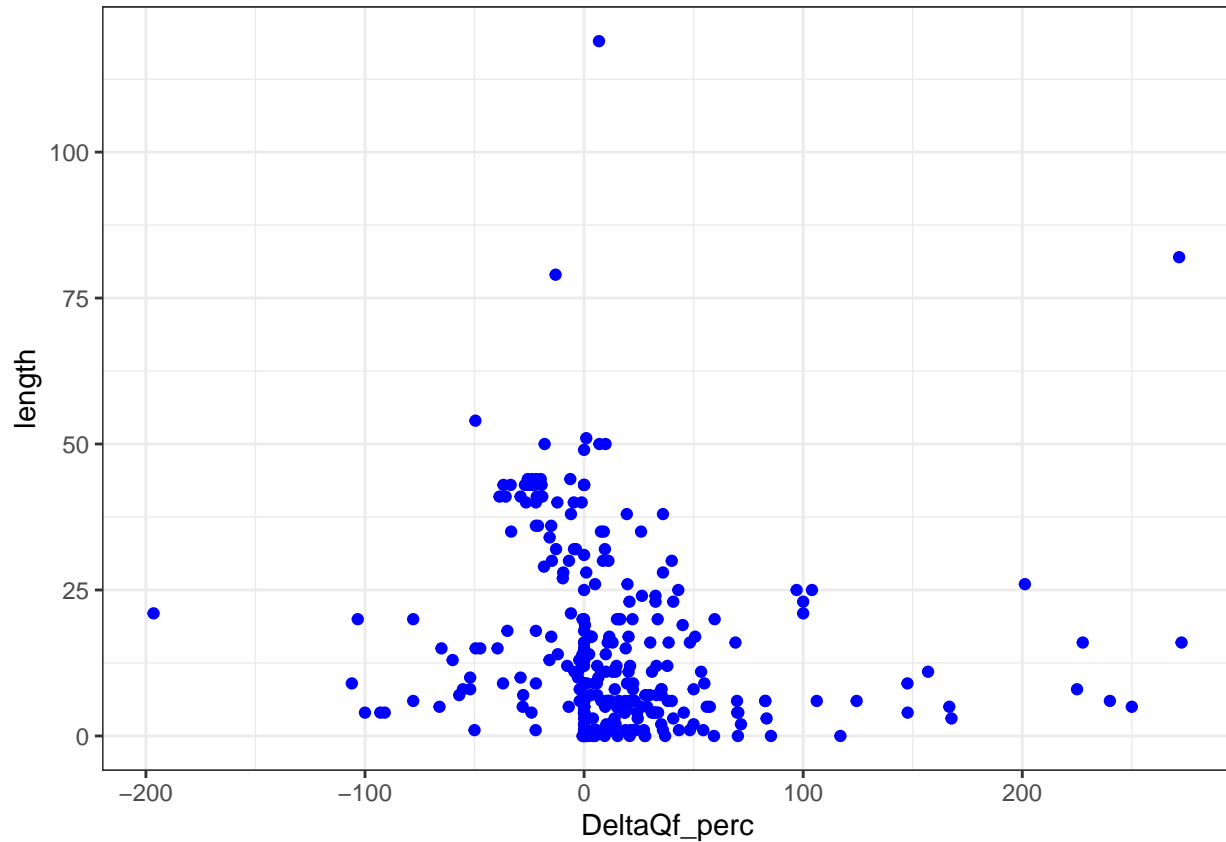


325

326 Clearly all have at least some relationship with Latitude, therefore are being  
 327 masked if Latitude is included in the model.

328 The flexible nature of the splines and the shrinkage applied to the splines  
 329 means that the length variable can capture variation in the data, but it is unclear  
 330 what exactly is captured. The shape of the resulting conditional response does  
 331 not reflect the reponses highlighted in Filoso et al. (2017) and Jackson et al.  
 332 (2005). Reducing the flexibility of the splines, or fitting a linear term results in  
 333 “length” not being significant, which suggests that this variable is not helpful  
 334 in explaining the variation in the data.

335 ## Warning: Removed 5 rows containing missing values (geom\_point).



336

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

343 In contrast to Filoso et al. (2017), we also did not identify an effect of the

344 Given how skewed Dryness is due to the few watersheds that have very high  
 345 dryness values, it is worth investigating what excluding these 4 watersheds from  
 346 the data means for the relationships.

```
model7_noLatb <- gam(DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign +
  s(Dryness, bs="ts" ) + #s(Latitude, bs="ts") +
  s(log10(Area_km2), bs="ts") + s(length,bs="ts") +
  Precip_data_type + Assessment_technique + Forest_type +
  Hydrological_regime, data = All_data2 %>% filter(length< 60))
summary(model7_noLatb)
```

347 ##

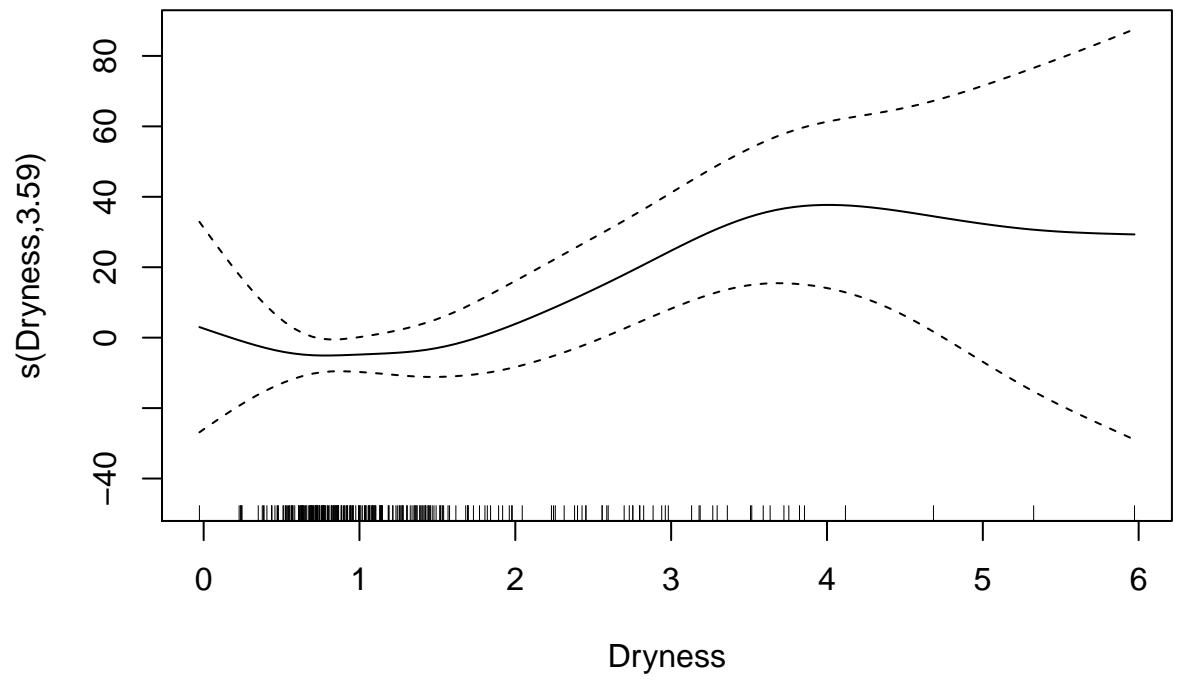


```

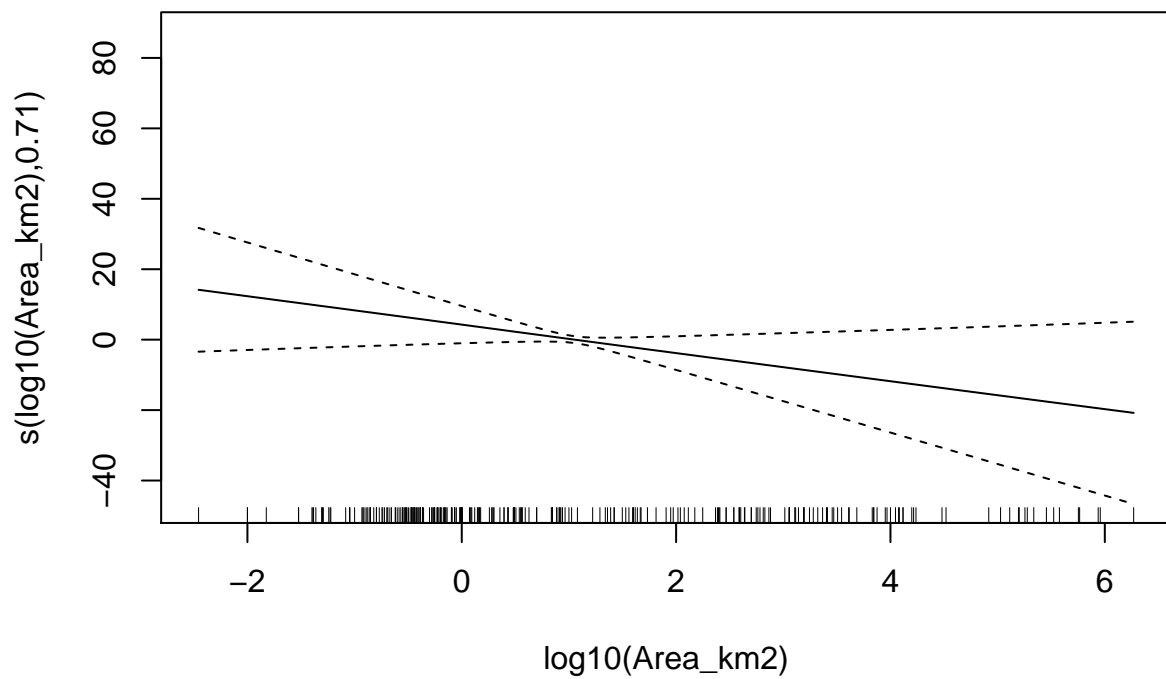
348 ## Family: gaussian
349 ## Link function: identity
350 ##
351 ## Formula:
352 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + s(Dryness, bs = "ts") +
353 ##       s(log10(Area_km2), bs = "ts") + s(length, bs = "ts") + Precip_data_type +
354 ##       Assessment_technique + Forest_type + Hydrological_regime
355 ##
356 ## Parametric coefficients:
357 ##               Estimate Std. Error t value Pr(>|t|)
358 ## (Intercept)      -12.5901    18.2493  -0.690 0.490850
359 ## DeltaF_perc_pos      0.2764     0.1021   2.708 0.007197 **
360 ## Forest_Signincrease -23.5034     6.9108  -3.401 0.000772 ***
361 ## Precip_data_typeOB  -11.8739    14.1588  -0.839 0.402419
362 ## Precip_data_typeSG   16.9396    16.3313   1.037 0.300543
363 ## Assessment_techniqueEA, HM  -8.2703     3.5091  -2.357 0.019143 *
364 ## Assessment_techniqueHM    28.0181    12.3042   2.277 0.023557 *
365 ## Assessment_techniquePWE    39.2429    12.0840   3.247 0.001310 **
366 ## Assessment_techniquePWE, HM -14.4779     5.5623  -2.603 0.009753 **
367 ## Assessment_techniqueQPW    33.8032    20.7827   1.627 0.105002
368 ## Assessment_techniqueQPW, EA  35.5789    25.2225   1.411 0.159506
369 ## Assessment_techniqueSH     42.8100    12.7078   3.369 0.000864 ***
370 ## Forest_typeCF          -3.8554     8.1687  -0.472 0.637324
371 ## Forest_typeMF          -4.0151     8.6538  -0.464 0.643047
372 ## Hydrological_regimeSD      6.4841    10.0569   0.645 0.519641
373 ## ---
374 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
375 ##
376 ## Approximate significance of smooth terms:
377 ##               edf Ref.df      F p-value
378 ## s(Dryness)      3.586e+00     9 1.665 0.0132 *
379 ## s(log10(Area_km2)) 7.053e-01     9 0.293 0.0510 .
380 ## s(length)       2.302e-10     9 0.000 0.9953
381 ## ---
382 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
383 ##
384 ## Rank: 39/42
385 ## R-sq.(adj) = 0.273 Deviance explained = 31.4%
386 ## GCV = 2119.6 Scale est. = 1992.7 n = 289

```

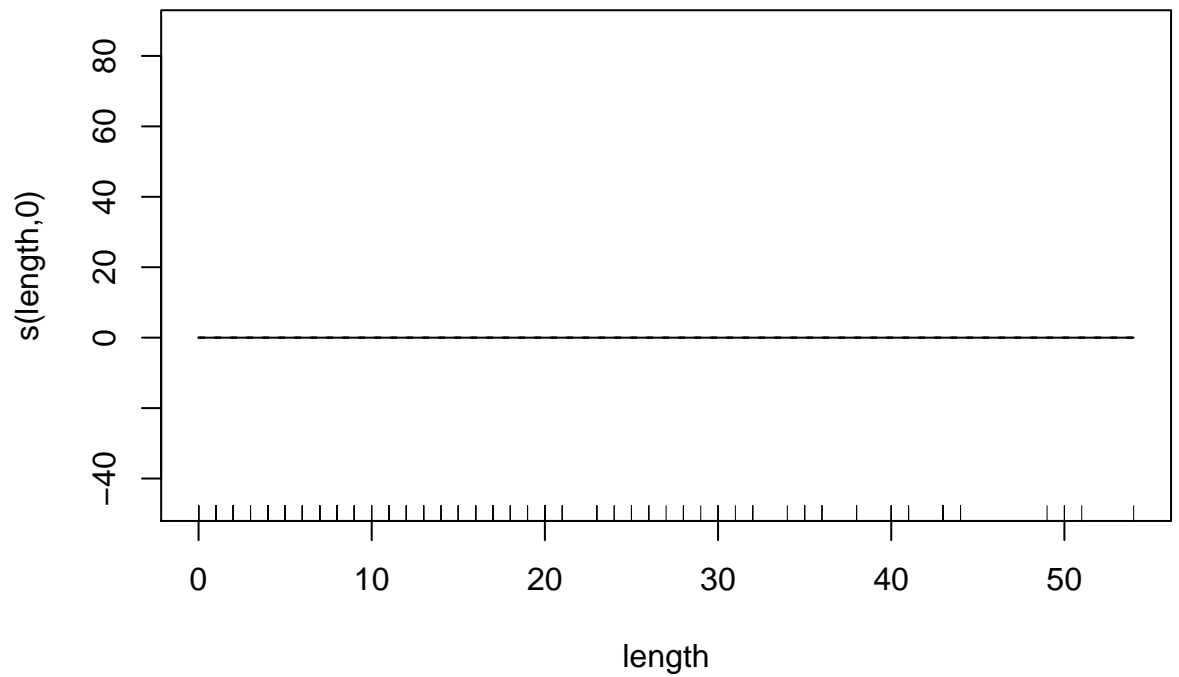
```
plot(model17_noLatb)
```



387



388



389

390 ## pdf

391 ## 2

```
Zhang_all2 %>%
```

```
  ggplot(aes(Longitude, Latitude, colour = DeltaF_perc, size = DeltaQf_perc/100 )) + geom_p
```

```
Zhang_all %>%
```

```
  ggplot(aes(Area_km2)) + geom_histogram(fill="blue", bins =50) +
```

```
  scale_x_log10()
```

```
total <- nrow(Zhang_all)
```

```
length(Zhang_all$Area_km2[Zhang_all$Area_km2<10])/total
```

```
Zhang_all2 %>%
```

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

## Discussion

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

As shown earlier, most of the smaller watersheds are “real observed data” using paired watershed studies, while for larger watersheds, the analysis are mostly based on modelling approximations using either elasticity analysis (EA), Hydrological modelling (HM) or a combined use of statistical methods (SH) or quasi paired watershed analysis (QPW), thus all providing an approximation of the effect of forestry on streamflow rather than a direct comparison of watersheds. This is a confounding factor that is not easily addressed in the regression modelling attempted here. Furthermore, the catchments analysed using EA, 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.

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)

433 Bosch, J.M., Hewlett, J.D., 1982. A review of catchment experiments to de-  
434 termine the effect of vegetation changes on water yield and evapotranspiration.  
435 *Journal of Hydrology* 55, 3–23.

436 Brown, A.E., Western, A.W., McMahon, T.A., Zhang, L., 2013. Impact of  
437 forest cover changes on annual streamflow and flow duration curves. *Journal of*  
438 *Hydrology* 483, 39–50. doi:http://dx.doi.org/10.1016/j.jhydrol.2012.12.031

439 Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W., Vertessy, R.A.,  
440 2005. A review of paired catchment studies for determining changes in water  
441 yield resulting from alterations in vegetation. *Journal of Hydrology* 310, 28–61.

442 Cosandey, C., Andréassian, V., Martin, C., Didon-Lescot, J.F., Lavabre,  
443 J., Folton, N., Mathys, N., Richard, D., 2005. The hydrological impact of the  
444 mediterranean forest: A review of french research. *Journal of Hydrology* 301,  
445 235–249. doi:https://doi.org/10.1016/j.jhydrol.2004.06.040

446 Filoso, S., Bezerra, M.O., Weiss, K.C.B., Palmer, M.A., 2017. Impacts  
447 of forest restoration on water yield: A systematic review. *PLOS ONE* 12,  
448 e0183210. doi:10.1371/journal.pone.0183210

449 Jackson, R.B., Jobbagy, E.G., Avissar, R., Roy, S.B., Barrett, D.J., Cook,  
450 C.W., Farley, K.A., Maitre, D.C. le, McCarl, B.A., Murray, B.C., 2005. Trading  
451 water for carbon with biological carbon sequestration. *Science* 310, 1944–1947.  
452 doi:10.1126/science.1119282

453 Kuczera, G., 1987. Prediction of water yield reductions following a bushfire  
454 in ash-mixed species eucalypt forest. *Journal of Hydrology* 94, 215–236. doi:Doi:  
455 10.1016/0022-1694(87)90054-0

456 Peña-Arancibia, J.L., Dijk, A.I.J.M. van, Guerschman, J.P., Mulligan, M.,  
457 Bruijnzeel, L.A., McVicar, T.R., 2012. Detecting changes in streamflow after  
458 partial woodland clearing in two large catchments in the seasonal tropics. *Jour-*  
459 *nal of Hydrology* 416–417, 60–71. doi:https://doi.org/10.1016/j.jhydrol.2011.11.036

460 Roche, M., 1981. Watershed investigations for development of forest re-  
461 sources of the amazon region in french guyana. *Tropical Agricultural Hydrology.*  
462 *J* 75–82.

463 Rodriguez, D.A., Tomasella, J., Linhares, C., 2010. Is the forest conversion  
464 to pasture affecting the hydrological response of amazonian catchments? Signals  
465 in the ji-paraná basin. *Hydrological Processes* 24, 1254–1269. doi:https://doi.org/10.1002/hyp.7586

466 Ruprecht, J.K., Schofield, N.J., Crombie, D.S., Vertessy, R.A., Stoneman,  
467 G.L., 1991. Early hydrological response to intense forest thinning in southwest-  
468 ern australia. *Journal of Hydrology* 127, 261–277. doi:https://doi.org/10.1016/0022-  
469 1694(91)90118-2

470 Thornton, C.M., Cowie, B.A., Freebairn, D.M., Playford, C.L., 2007. The  
471 brigalow catchment study: II\*. Clearing brigalow (acacia harpophylla) for crop-  
472 ping or pasture increases runoff. *Australian Journal of Soil Research* 45, 496–  
473 511. doi:doi:10.1071/SR07064

474 Trabucco, A., Zomer, R.J., 2018. Global aridity index and potential evapo-  
475 transpiration (et0) climate database v2. CGIAR consortium for spatial information(CGIAR-  
476 csi).

477 Wood, S., 2006. Generalized additive models: An introduction with r. CRC  
478 Press, Boca Raton, FL.

479 Zhang, L., Zhao, F., Chen, Y., Dixon, R.N.M., 2011. Estimating effects of  
 480 plantation expansion and climate variability on streamflow for catchments in  
 481 australia. *Water Resources Research* 47, W12539. doi:10.1029/2011wr010711  
 482 Zhang, M., Liu, N., Harper, R., Li, Q., Liu, K., Wei, X., Ning, D., Hou, Y.,  
 483 Liu, S., 2017. A global review on hydrological responses to forest change across  
 484 multiple spatial scales: Importance of scale, climate, forest type and hydrological  
 485 regime. *Journal of Hydrology* 546, 44–59. doi:https://doi.org/10.1016/j.jhydrol.2016.12.040  
 486 Zhao, F., Zhang, L., Xu, Z., Scott, D.F., 2010. Evaluation of methods for  
 487 estimating the effects of vegetation change and climate variability on streamflow.  
 488 *Water Resources Research* 46, W03505. doi:10.1029/2009wr007702  
 489 Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., Sun, G., Scott,  
 490 D.F., Zhou, S., Han, L., Su, Y., 2015. Global pattern for the effect of climate and  
 491 land cover on water yield. *Nature Communications* 6, 5918. doi:10.1038/ncomms6918  
 492 Zhou, G., Wei, X., Luo, Y., Zhang, M., Li, Y., Qiao, Y., Liu, H., Wang,  
 493 C., 2010. Forest recovery and river discharge at the regional scale of guangdong  
 494 province, china. *Water Resources Research* 46. doi:https://doi.org/10.1029/2009WR008829