

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

3 R. Willem Vervoort^{*,a,1}, Eliana Nervi^b, Jimena Alonso^{**,c}

4 ^a*School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006,*
5 *Australia*

6 ^b*INIA, Uruguay*

7 ^c*Institute of Fluid Mechanics and Environmental Engineering, School of Engineering,*
8 *Universidad de la República, 11200 Montevideo, Departamento de Montevideo, Uruguay*

9 Abstract

This is the abstract. It consists of two paragraphs.

10 Introduction

11 *Introduction*

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

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

*Corresponding Author

**Equal contribution.
Preprint submitted to *Journal of Hydrology* November 14, 2021
Email addresses: willem.vervoort@sydney.edu.au (R. Willem Vervoort),
eliananervi@gmail.com (Eliana Nervi), jalonso@fing.edu.uy (Jimena Alonso)

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

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

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

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

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

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

82 Building on the analyses by Zhang et al. (2017) and Filoso et al. (2017),
83 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 catchment 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 km², 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 (ET0) 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)).

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 as:

$$D = \frac{E0}{Pa} \quad (1)$$

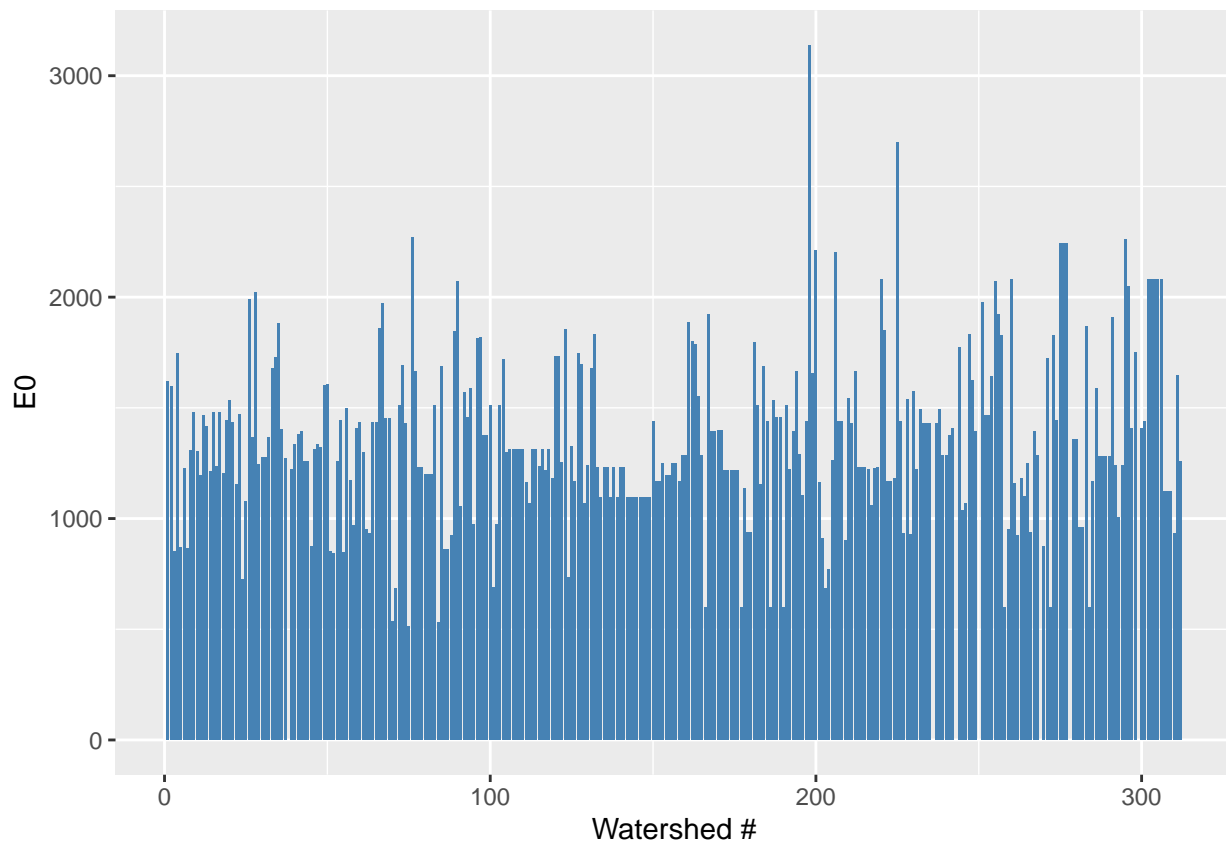
```

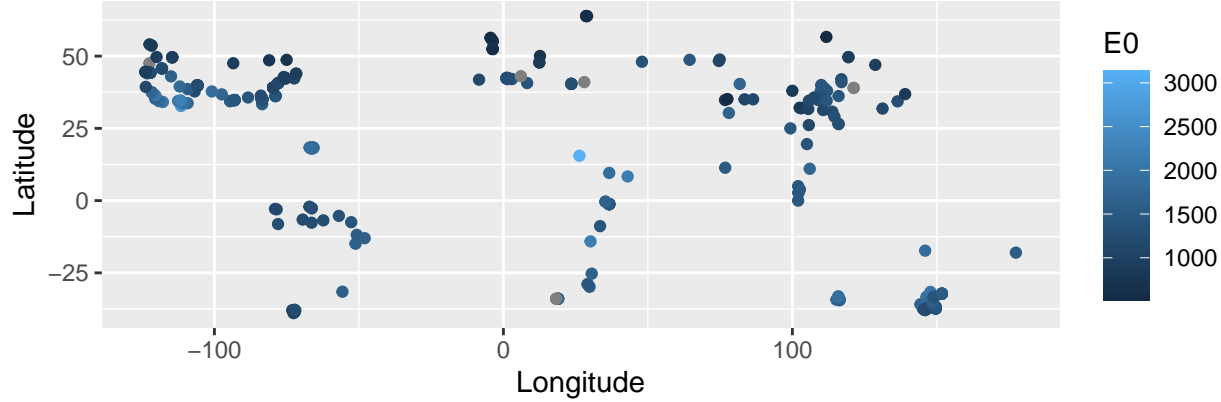
147 Statistical modelling
148 ##
149 ## -- Column specification -----
150 ## cols(
151 ##   Watershed = col_double(),
152 ##   Latitude = col_double(),
153 ##   Longitude = col_double(),
154 ##   rvalue_1 = col_double()
155 ## )

156 ##
157 ## -- Column specification -----
158 ## cols(
159 ##   site_name = col_double(),
160 ##   lat = col_double(),
161 ##   long = col_double(),
162 ##   `Area(km2)` = col_double(),
163 ##   `Forest typ` = col_character(),
164 ##   C = col_character(),
165 ##   rvalue_1 = col_double()
166 ## )

167 ## Warning: Removed 7 rows containing missing values (position_stack).

```





169

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

173 The general model tested is

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

174 In this model, no direct interactions are assumed. The changes in forest
 175 cover contain both positive (forestation) and negative values (deforestation). In
 176 Zhang et al. (2017), these changes were jointly analysed, assuming the effect
 177 on the change in flow was linear and non-hysteretic. However, the impact of
 178 an increase in forest cover can be different from the same fractional decrease in
 179 forest cover. Therefore all the change in forest cover data is converted to positive
 180 values, and an additional column ($sign_{forest\ cover}$) is added that indicates
 181 whether it was a forest cover increase or decrease. A further assumption in the
 182 model is that all continuous variables Z_i (such as annual precipitation (Pa)) can
 183 have a linear or non-linear relationship with $\Delta Q\%$. This means that a smooth
 184 function $s()$ is applied to the Z_i variables.

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

189 More generally the results were analysed to identify:

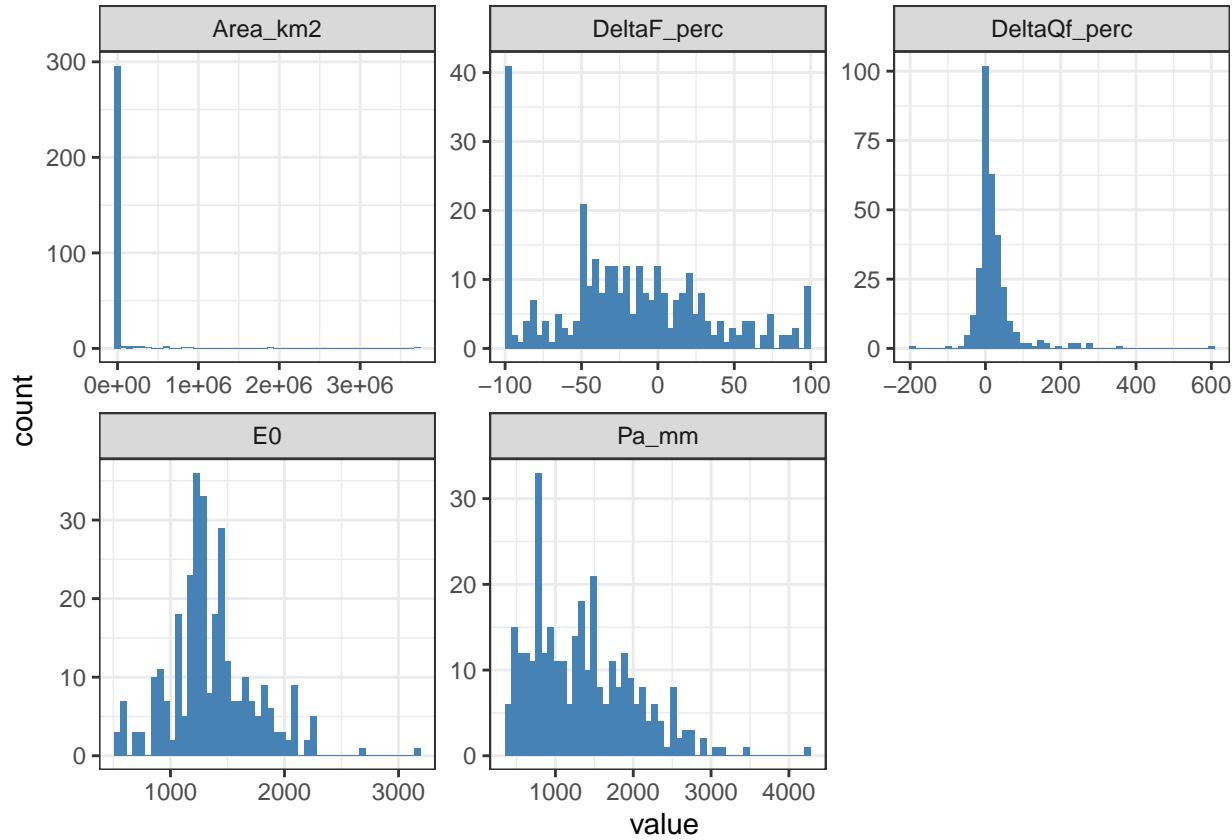
- 190 1. the significance of the different variables
- 191 2. the direction of the categorical or shape of the smooth variables

192 Results

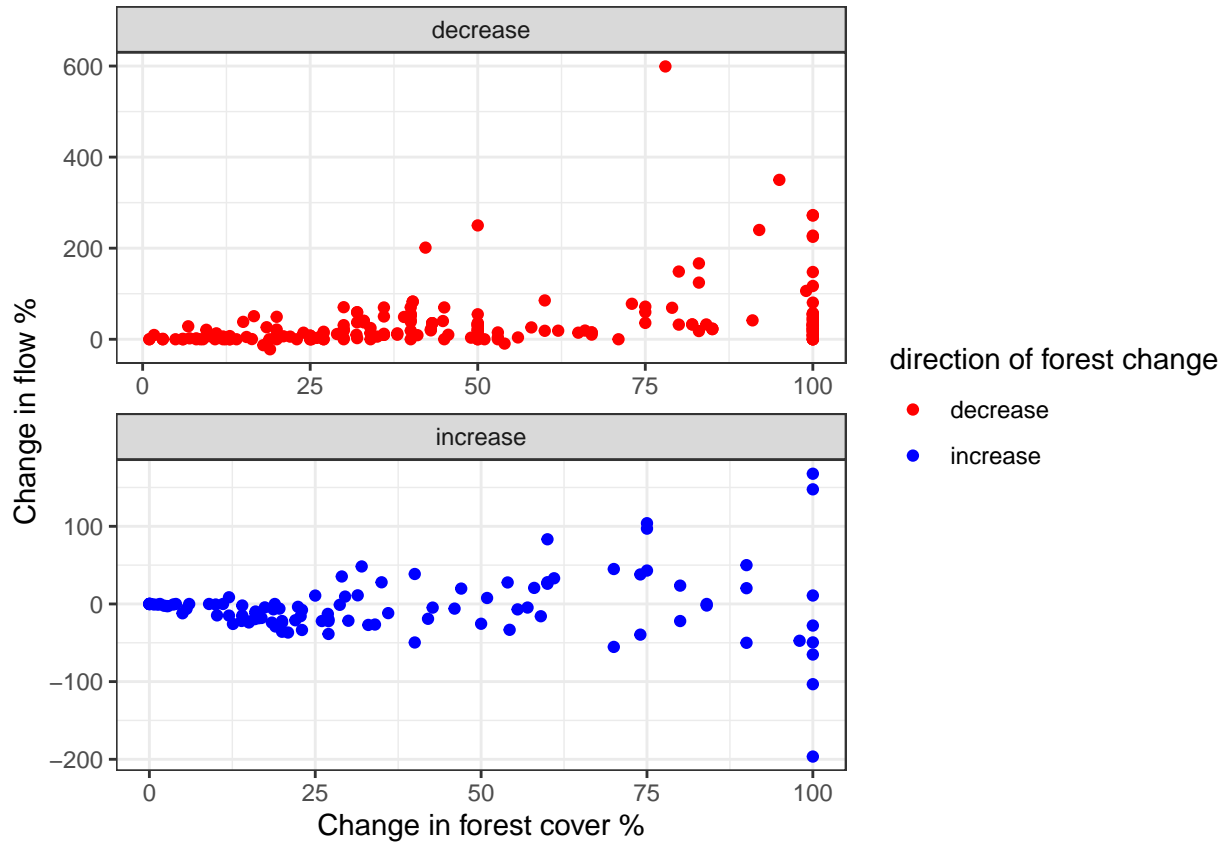
193 *description of the data*

194 The overall dataset contains 312 observations of changes in flow. The overall
 195 distribution of changes in flow is highly skewed as is the distribution of changes
 196 in forest cover and Area. The values of changes in flow greater than 100% and
 197 smaller than -100% clearly create long tails on the change in flow distribution.
 198 Note also the large number of studies with 100% forest cover reduction.

199 **## Warning: Removed 7 rows containing non-finite values (stat_bin).**



200



201
 202 This immediately shows that for the data related to forest decreases, there
 203 is almost always a positive flow change. In other words, flow almost always
 204 increased. However, for increases in forest cover, this is not the case, and flow
 205 can both increase and decrease. However in both cases the variability in the
 206 reported change in flow increases with the increase in forest cover change.

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

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

```

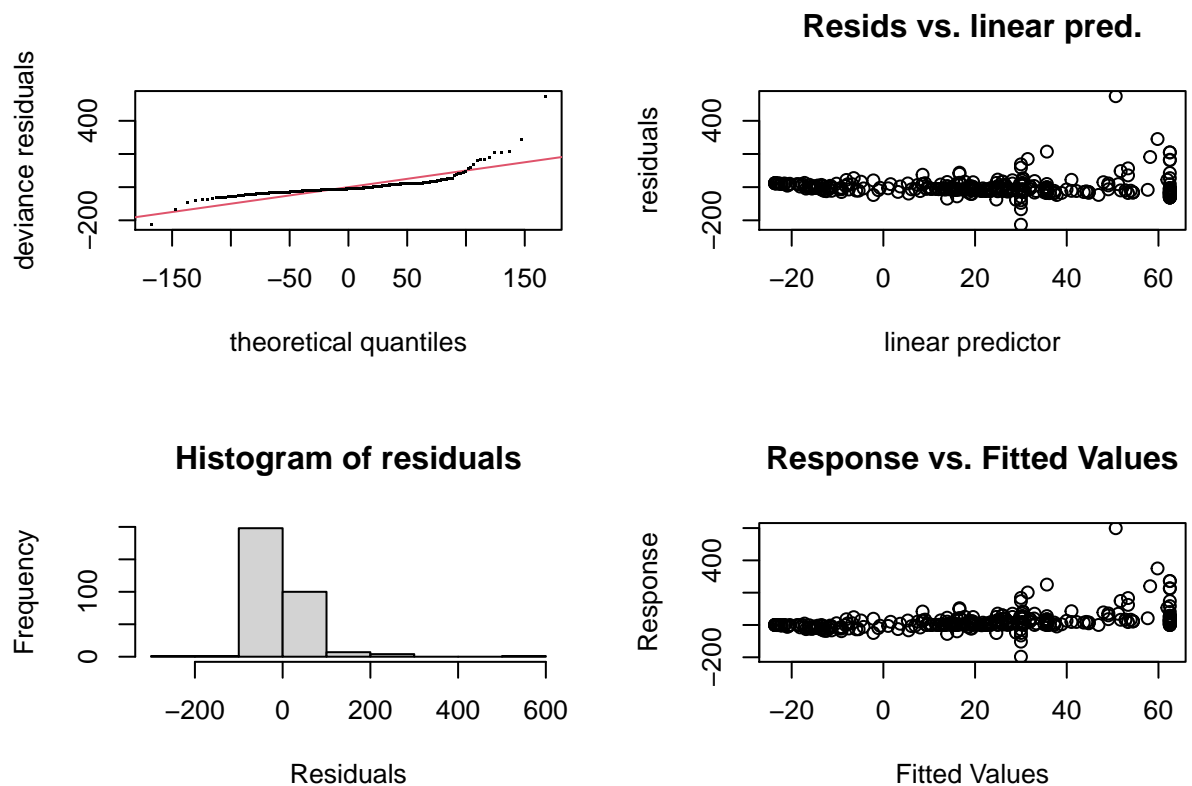
212 ##
213 ## Family: gaussian
214 ## Link function: identity
215 ##
216 ## Formula:
217 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign
218 ##

```

```

219 ## Parametric coefficients:
220 ##               Estimate Std. Error t value Pr(>|t|)
221 ## (Intercept)      8.8466     6.5008   1.361   0.175
222 ## DeltaF_perc_pos  0.5364     0.1017   5.274 2.52e-07 ***
223 ## Forest_Signincrease -32.4760     6.9367  -4.682 4.26e-06 ***
224 ## ---
225 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
226 ##
227 ##
228 ## R-sq.(adj) =  0.16   Deviance explained = 16.6%
229 ## GCV = 3278.3   Scale est. = 3246.8       n = 312

```



230

```

231 ##
232 ## Method: GCV   Optimizer: magic
233 ## Model required no smoothing parameter selectionModel rank = 3 / 3

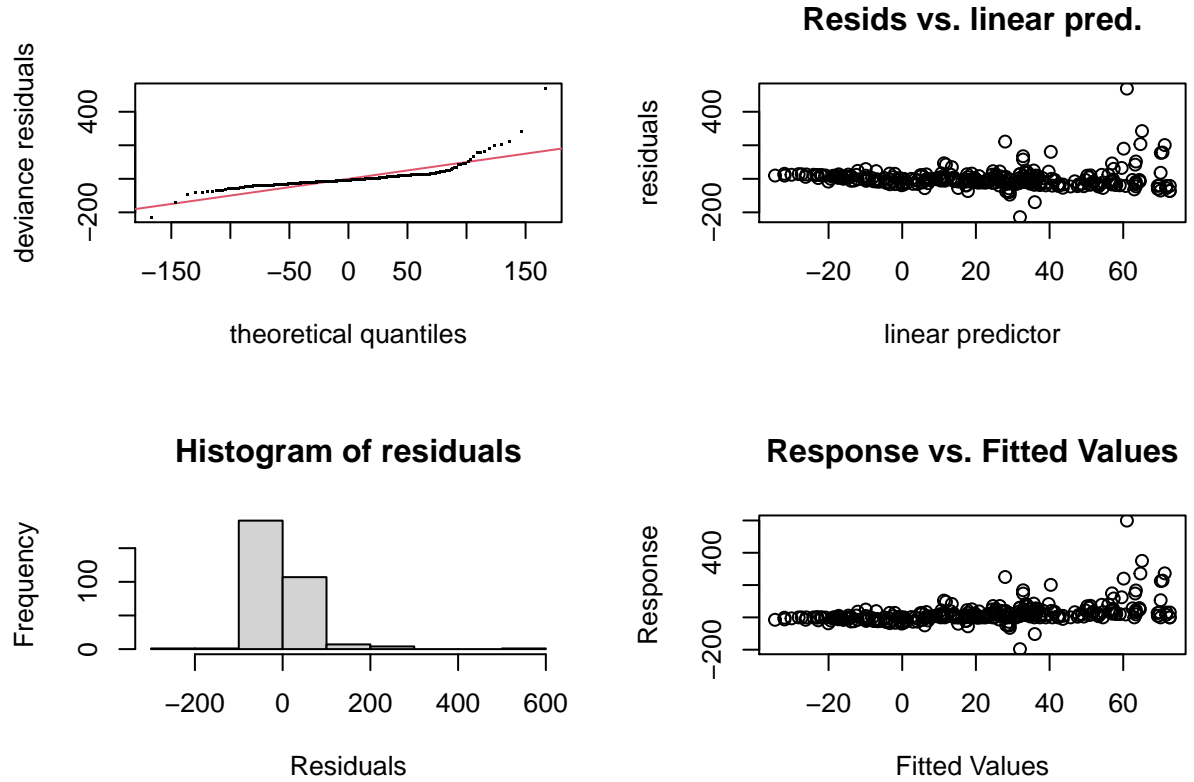
```

Length	Class	Mode
0	NULL	NULL

While the overall variance explained in this model is not high at 0.16, it clearly 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 0.5%. However the change in flow is different for forest cover decreases compared to forest cover increases. In fact, forest cover increases decrease flow by 32% 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 $(1 - 0.32) * 0.5\%$, while a the percentage forest cover decrease will increase flow by 0.5%.

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

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + Area_km2 + Pa_mm
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.323e+01  9.189e+00  2.528   0.0120 *
## DeltaF_perc_pos  5.424e-01  1.030e-01  5.265 2.64e-07 ***
## Forest_Signincrease -3.324e+01  6.955e+00 -4.779 2.73e-06 ***
## Area_km2       -1.867e-06  1.322e-05 -0.141   0.8878
## Pa_mm          -1.067e-02  4.944e-03 -2.158   0.0317 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.168   Deviance explained = 17.9%
## GCV = 3269.3   Scale est. = 3217       n = 312
```



266

```
267 ##
268 ## Method: GCV   Optimizer: magic
269 ## Model required no smoothing parameter selectionModel rank = 5 / 5
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	23.23	9.189	2.528	0.01197
DeltaF_perc_pos	0.5424	0.103	5.265	2.64e-07
Forest_Signincrease	-33.24	6.955	-4.779	2.733e-06
Area_km2	-1.867e-06	1.322e-05	-0.1412	0.8878
Pa_mm	-0.01067	0.004944	-2.158	0.03169

270 Including area and annual precipitation does not really improve the overall
 271 explaining power of the model, in fact, annual precipitation appears to be only
 272 a very small confounding factor, representing only a -0.01/% partial effect in the
 273 change in streamflow, holding all other factors constant. In contrast to earlier
 274 reported studies (Filoso et al., 2017; Zhang et al., 2017), watershed area has no

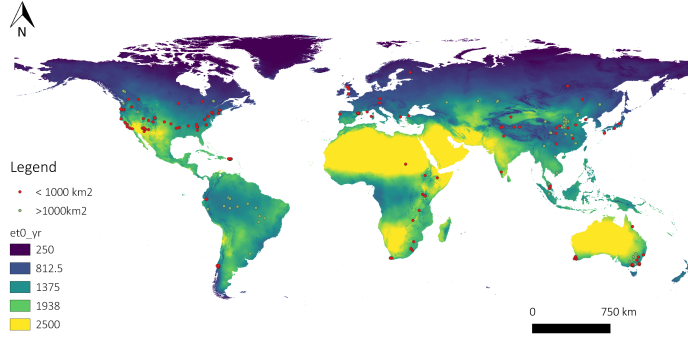


Figure 1: Distribution of included catchments across the globe based on reported or estimated latitude and longitude

effect 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 remain the change in forest cover and whether this is an increase or decrease.

The effect of location on the globe

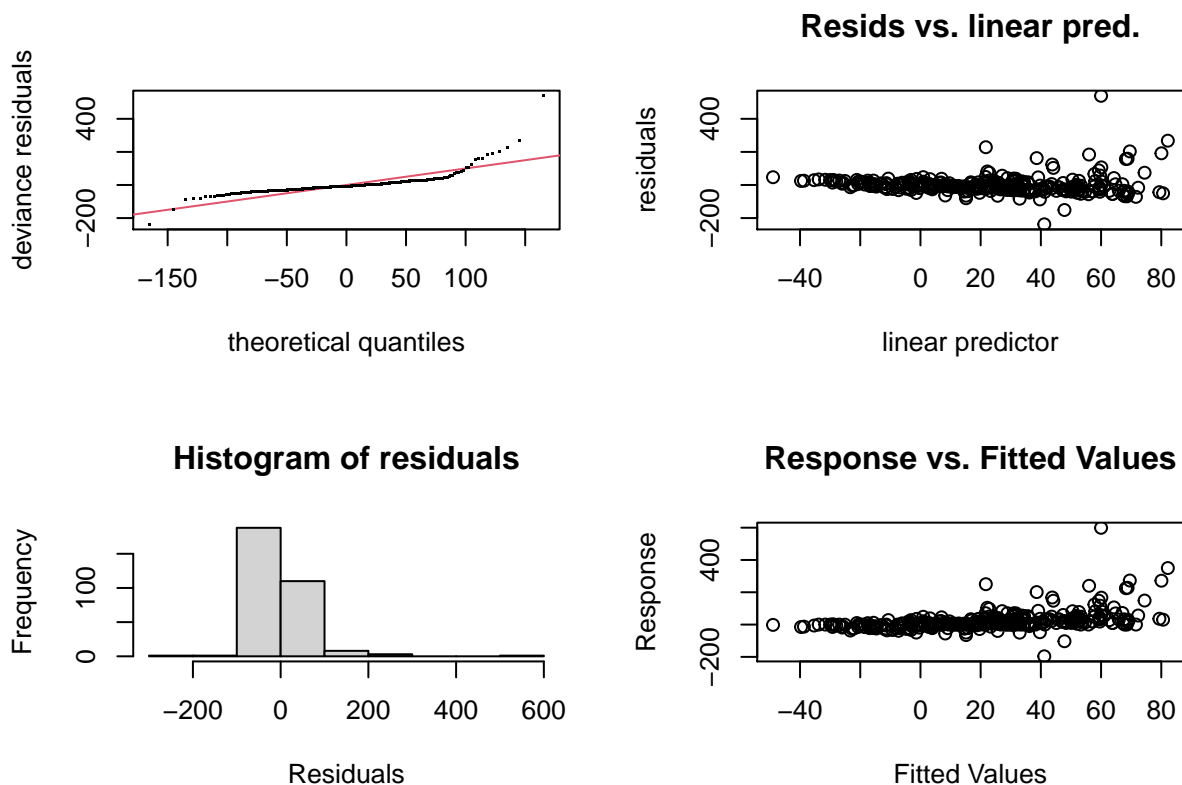
As indicated, a second hypothesis relates to whether there is a strong spatial global gradient as captured by latitude and longitude. As the global map (@ref(fig:global_map)) shows, the distribution of case study catchments covers multiple continents and shows some distinct clustering in parts of the world. Of interest is whether the spatial clustering also indicates a difference in response to forest cover change.

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + Area_km2 + Pa_mm +
##      Latitude + Longitude
##
## Parametric coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.504e+01  1.026e+01   3.415 0.000724 ***
## DeltaF_perc_pos  5.147e-01  1.027e-01   5.014 9.07e-07 ***
## Forest_Signincrease -3.793e+01  7.230e+00 -5.247 2.90e-07 ***
## Area_km2        -6.282e-06  1.327e-05  -0.474 0.636137
## Pa_mm           -1.304e-02  5.055e-03  -2.580 0.010339 *
## Latitude        -3.119e-01  1.255e-01  -2.484 0.013520 *
## Longitude       -1.058e-02  3.965e-02  -0.267 0.789747
```

```

303 ## ---
304 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
305 ##
306 ##
307 ## R-sq.(adj) =  0.183   Deviance explained = 19.9%
308 ## GCV = 3230.5   Scale est. = 3158       n = 312

```



```

309

```

```

310 ##
311 ## Method: GCV   Optimizer: magic
312 ## Model required no smoothing parameter selectionModel rank = 7 / 7

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	35.04	10.26	3.415	0.0007243
DeltaF_perc_pos	0.5147	0.1027	5.014	9.071e-07
Forest_Signincrease	-37.93	7.23	-5.247	2.898e-07
Area_km2	-6.282e-06	1.327e-05	-0.4736	0.6361

	Estimate	Std. Error	t value	Pr(> t)
Pa_mm	-0.01304	0.005055	-2.58	0.01034
Latitude	-0.3119	0.1255	-2.484	0.01352
Longitude	-0.01058	0.03965	-0.2669	0.7897

313 This linear model shows that there is a significant gradient in the Latitude
314 and with annual average rainfall, with watersheds closer to the equator hav-
315 ing lower changes in the runoff compare to watersheds further away from the
316 equator. This suggests an influence of radiation, which will be tested next. In
317 addition, the model suggests an influence of the annual average rainfall, with
318 wetter watersheds having slightly lower changes in runoff. The total explaining
319 power of the model is still low at 0.18 suggesting further confounding factors
320 currently not included in the model.

321 There is no relationship with Longitude, suggesting that the different conti-
322 nents do not show a trend in the East-West direction.

323 *Impact of the dryness index*

324 The previous analysis suggests an influence of radiation on evapotranspira-
325 tion, and most likely related to the dryness index, as also indicated in Zhang et
326 al. (2017). Increased evapotranspiration could lead to drier catchments, unless
327 balanced by rainfall (such as possibly in the tropics). This model introduces the
328 dryness index as a linear variable and drops the annual average precipitation as
329 a variables.

```

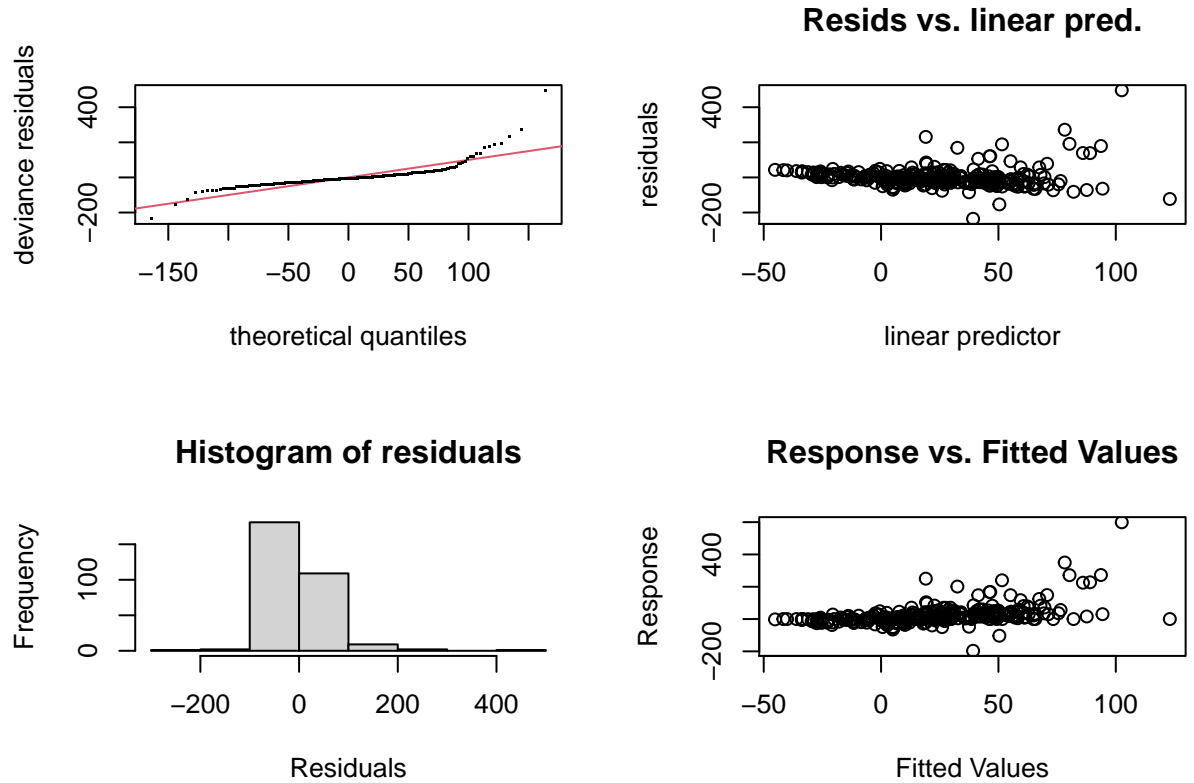
330 ##
331 ## Family: gaussian
332 ## Link function: identity
333 ##
334 ## Formula:
335 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + Area_km2 + Latitude +
336 ##      Longitude + Dryness
337 ##
338 ## Parametric coefficients:
339 ##              Estimate Std. Error t value Pr(>|t|)
340 ## (Intercept)    6.031e-01  8.582e+00  0.070 0.944022
341 ## DeltaF_perc_pos  5.006e-01  1.036e-01  4.831 2.18e-06 ***
342 ## Forest_Signincrease -3.898e+01  7.253e+00 -5.375 1.55e-07 ***
343 ## Area_km2         -7.227e-06  1.317e-05 -0.549 0.583604
344 ## Latitude         -3.121e-01  1.249e-01 -2.500 0.012974 *
345 ## Longitude        -6.775e-03  3.958e-02 -0.171 0.864199
346 ## Dryness          1.370e+01  3.628e+00  3.776 0.000193 ***
347 ## ---
348 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
349 ##
350 ##

```

```

351 ## R-sq.(adj) = 0.204   Deviance explained = 21.9%
352 ## GCV = 3199.2   Scale est. = 3125.7   n = 305

```



353

```

354 ##
355 ## Method: GCV   Optimizer: magic
356 ## Model required no smoothing parameter selectionModel rank = 7 / 7

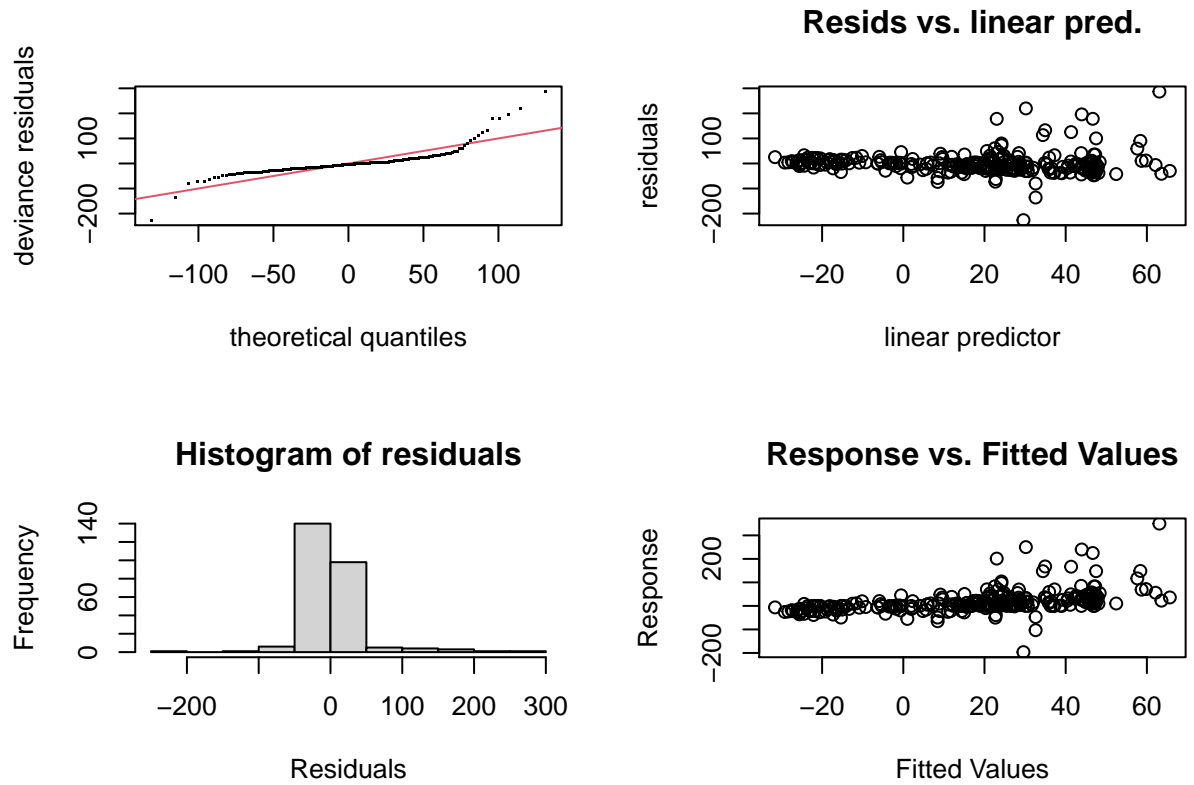
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.6031	8.582	0.07027	0.944
DeltaF_perc_pos	0.5006	0.1036	4.831	2.178e-06
Forest_Signincrease	-38.98	7.253	-5.375	1.549e-07
Area_km2	-7.227e-06	1.317e-05	-0.5487	0.5836
Latitude	-0.3121	0.1249	-2.5	0.01297
Longitude	-0.006775	0.03958	-0.1712	0.8642
Dryness	13.7	3.628	3.776	0.0001927

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

```

366 ##
367 ## Family: gaussian
368 ## Link function: identity
369 ##
370 ## Formula:
371 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + s(Area_km2) +
372 ##      s(Dryness) + length + Assessment_technique + Latitude
373 ##
374 ## Parametric coefficients:
375 ##               Estimate Std. Error t value Pr(>|t|)
376 ## (Intercept)      19.80962   20.90807   0.947  0.34433
377 ## DeltaF_perc_pos    0.34576    0.10727   3.223  0.00144 **
378 ## Forest_Signincrease -33.34803    7.26563  -4.590 7.06e-06 ***
379 ## length           -0.16503    0.26850  -0.615  0.53935
380 ## Assessment_techniqueEA, HM  6.97930   47.05195   0.148  0.88220
381 ## Assessment_techniqueHM    1.39799   16.63165   0.084  0.93308
382 ## Assessment_techniquePWE    4.12984   19.29598   0.214  0.83070
383 ## Assessment_techniquePWE, HM -5.44499   49.38888  -0.110  0.91230
384 ## Assessment_techniqueQPW   -9.50504   26.72374  -0.356  0.72239
385 ## Assessment_techniqueSH     2.52746   18.97865   0.133  0.89416
386 ## Latitude           -0.30515    0.09629  -3.169  0.00172 **
387 ## ---
388 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
389 ##
390 ## Approximate significance of smooth terms:
391 ##               edf Ref.df      F p-value
392 ## s(Area_km2)    1      1 0.513  0.474
393 ## s(Dryness)     1      1 0.002  0.964
394 ##
395 ## R-sq.(adj) =  0.185   Deviance explained = 22.2%
396 ## GCV = 2181.8   Scale est. = 2072.8      n = 260
  
```

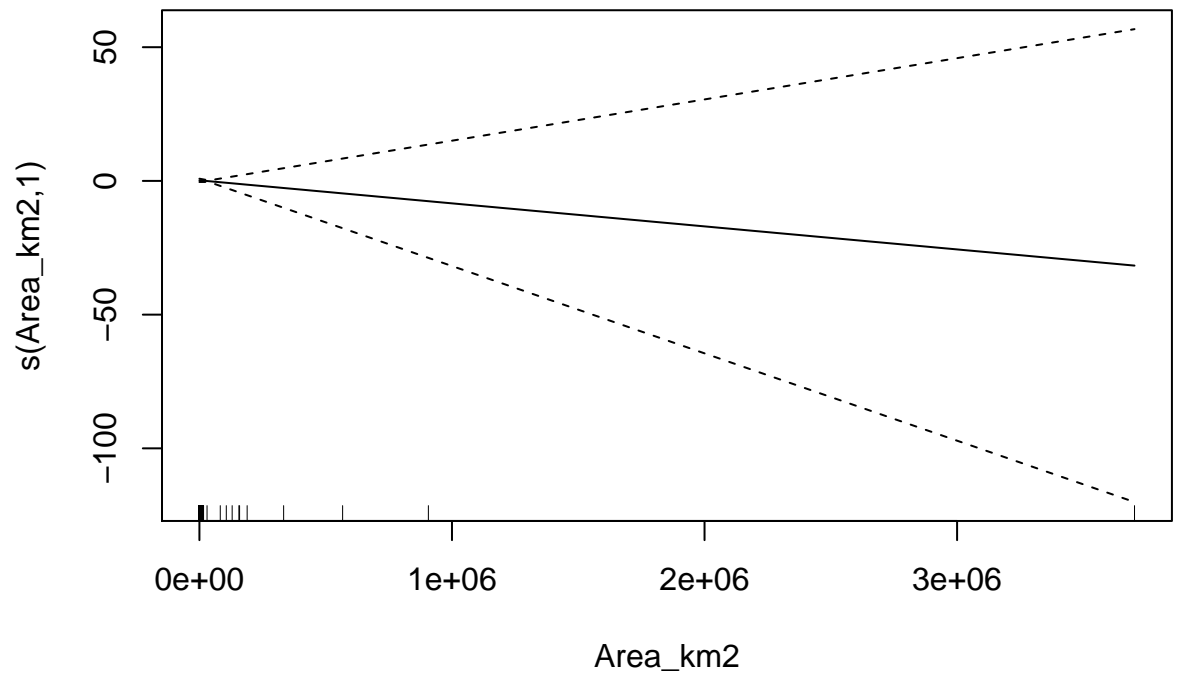


397

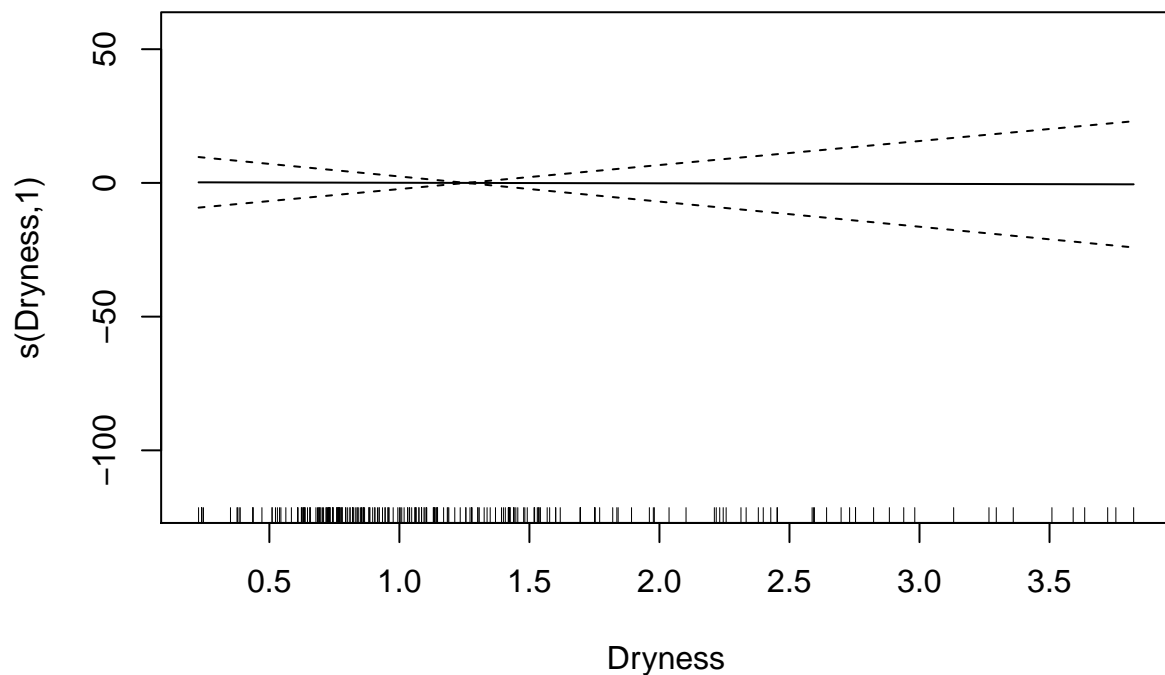
```

398 ##
399 ## Method: GCV   Optimizer: magic
400 ## Smoothing parameter selection converged after 15 iterations.
401 ## The RMS GCV score gradient at convergence was 0.0001863276 .
402 ## The Hessian was positive definite.
403 ## Model rank = 29 / 29
404 ##
405 ## Basis dimension (k) checking results. Low p-value (k-index<1) may
406 ## indicate that k is too low, especially if edf is close to k'.
407 ##
408 ##           k' edf k-index p-value
409 ## s(Area_km2) 9  1   1.06  0.825
410 ## s(Dryness)  9  1   0.86  0.015 *
411 ## ---
412 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```



413



414

```

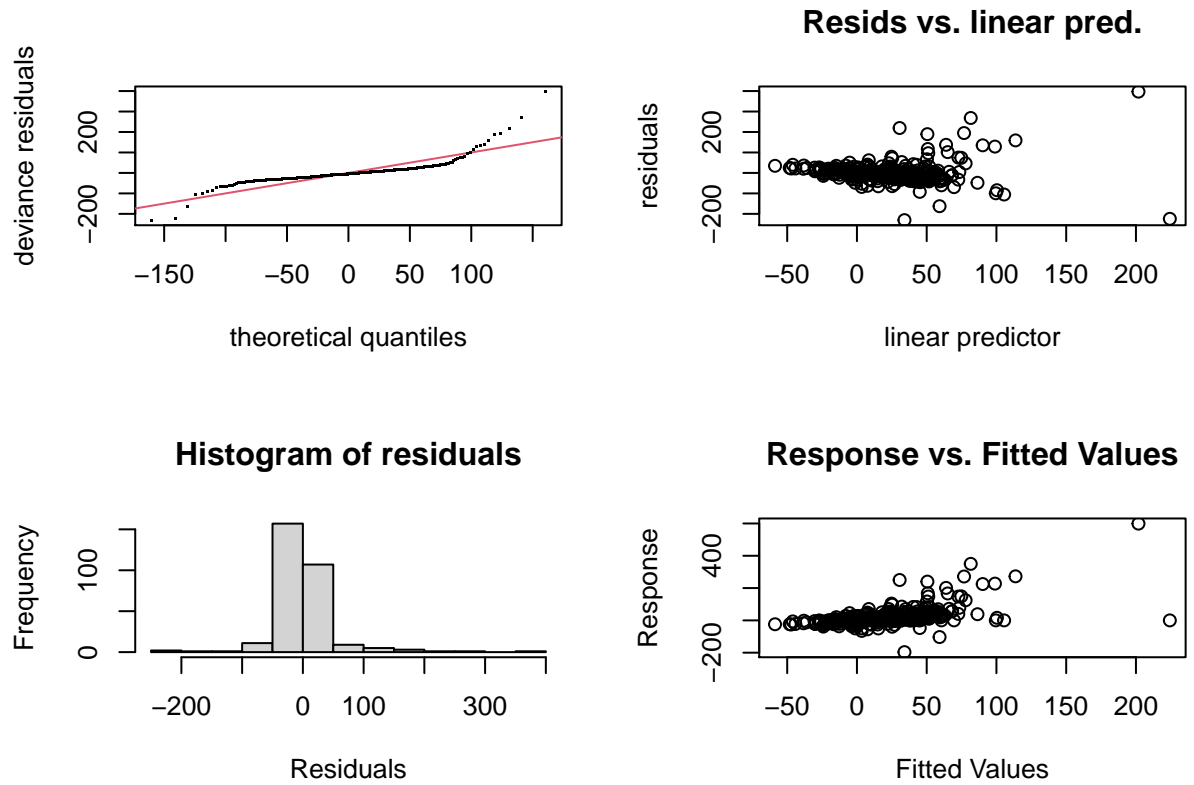
415 ##
416 ## Family: gaussian
417 ## Link function: identity
418 ##
419 ## Formula:
420 ## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign + te(Dryness, Latitude,
421 ##      k = 5) + Precip_data_type + Assessment_technique + Forest_type +
422 ##      Hydrological_regime
423 ##
424 ## Parametric coefficients:
425 ##               Estimate Std. Error t value Pr(>|t|)
426 ## (Intercept)    -15.0968   26.7857  -0.564 0.573480
427 ## DeltaF_perc_pos    0.4516    0.1187   3.805 0.000175 ***
428 ## Forest_Signincrease -27.5255    9.0389  -3.045 0.002552 **
429 ## Precip_data_typeOB  -5.2815   18.9655  -0.278 0.780856
430 ## Precip_data_typeSG   11.3166   21.3449   0.530 0.596417
431 ## Assessment_techniqueEA, HM  45.5460   57.7615   0.789 0.431077
432 ## Assessment_techniqueHM   21.7074   19.8327   1.095 0.274689

```

```

433 ## Assessment_techniquePWE      29.7823      20.4186      1.459 0.145827
434 ## Assessment_techniquePWE, HM  51.4437      62.8574      0.818 0.413831
435 ## Assessment_techniqueQPW      19.1277      31.5696      0.606 0.545090
436 ## Assessment_techniqueQPW, EA  29.9396      35.2621      0.849 0.396591
437 ## Assessment_techniqueSH       28.3985      22.2674      1.275 0.203273
438 ## Forest_typeCF                 5.0733      11.4701      0.442 0.658619
439 ## Forest_typeMF                -13.2933      10.8133     -1.229 0.219997
440 ## Hydrological_regimeSD         13.3327      13.2135      1.009 0.313855
441 ## ---
442 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
443 ##
444 ## Approximate significance of smooth terms:
445 ##                edf Ref.df      F  p-value
446 ## te(Dryness,Latitude) 10.7  12.97 3.432 6.26e-05 ***
447 ## ---
448 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
449 ##
450 ## R-sq.(adj) =  0.249   Deviance explained = 31.1%
451 ## GCV = 3281.3   Scale est. = 2999.2      n = 299

```

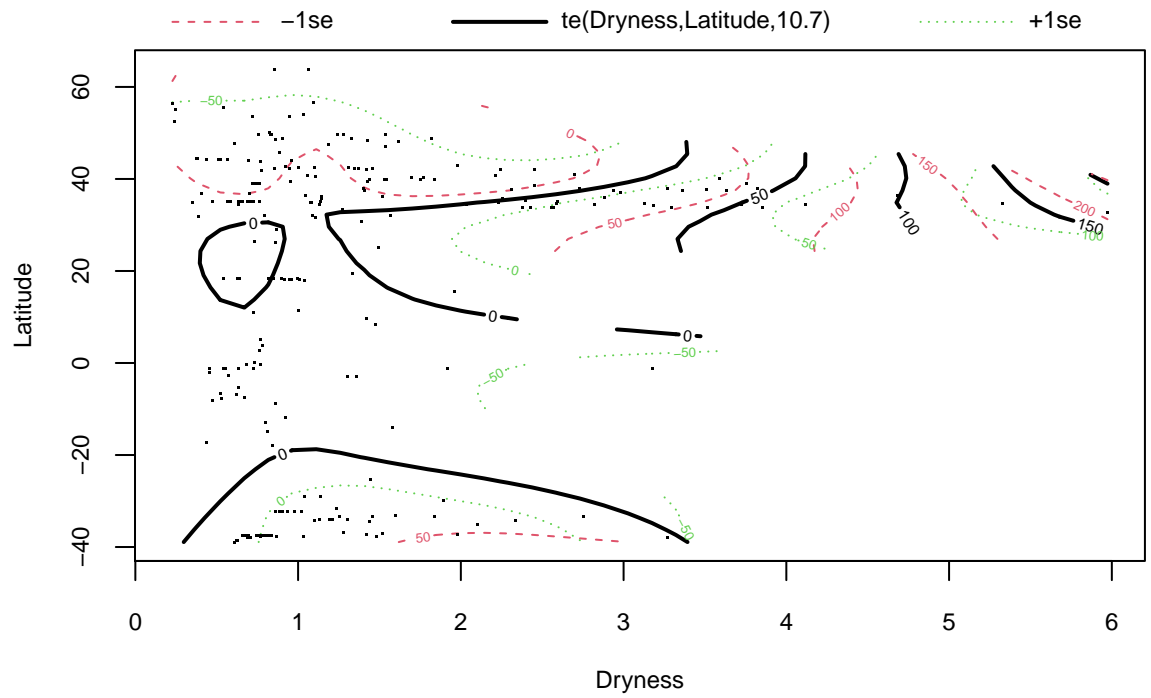


452

```

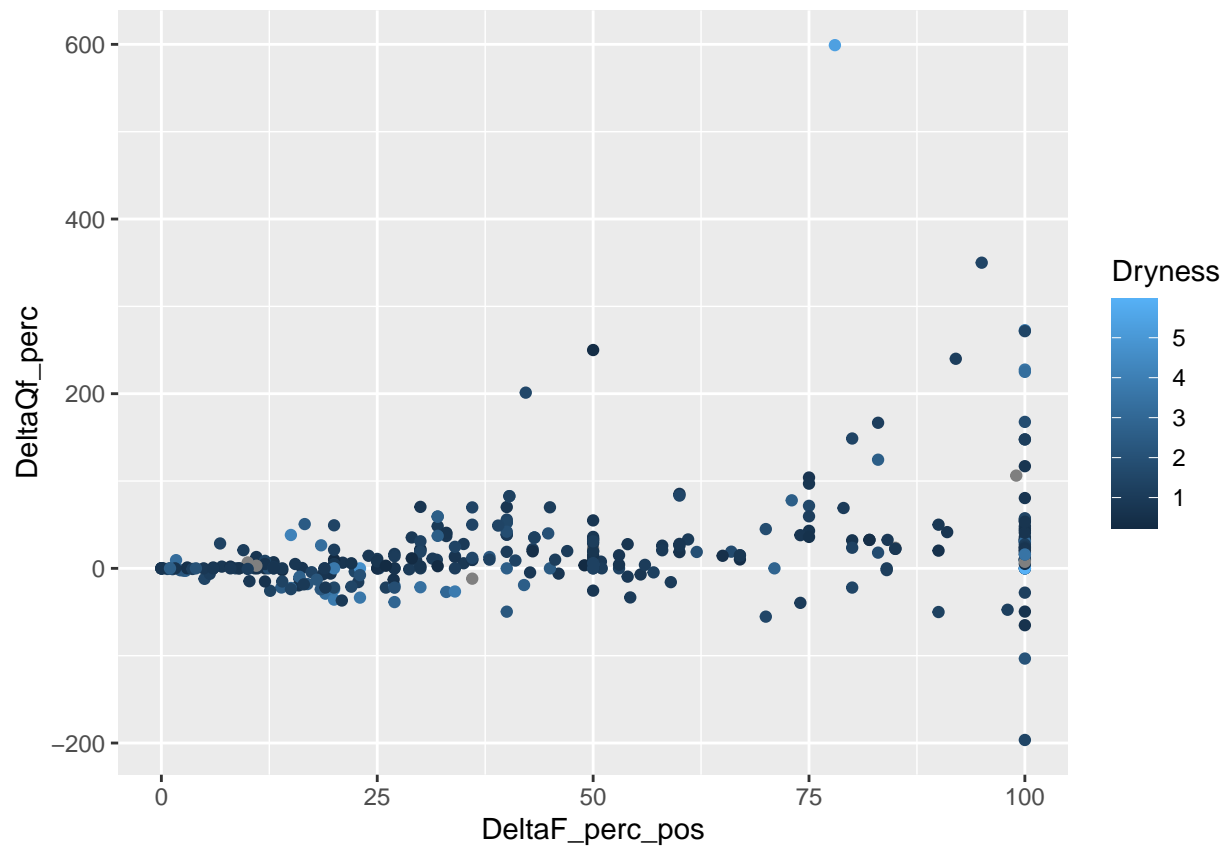
453 ##
454 ## Method: GCV   Optimizer: magic
455 ## Smoothing parameter selection converged after 7 iterations.
456 ## The RMS GCV score gradient at convergence was 0.002397792 .
457 ## The Hessian was positive definite.
458 ## Model rank = 39 / 39
459 ##
460 ## Basis dimension (k) checking results. Low p-value (k-index<1) may
461 ## indicate that k is too low, especially if edf is close to k'.
462 ##
463 ##           k'   edf k-index p-value
464 ## te(Dryness,Latitude) 24.0 10.7      1    0.52

```



465

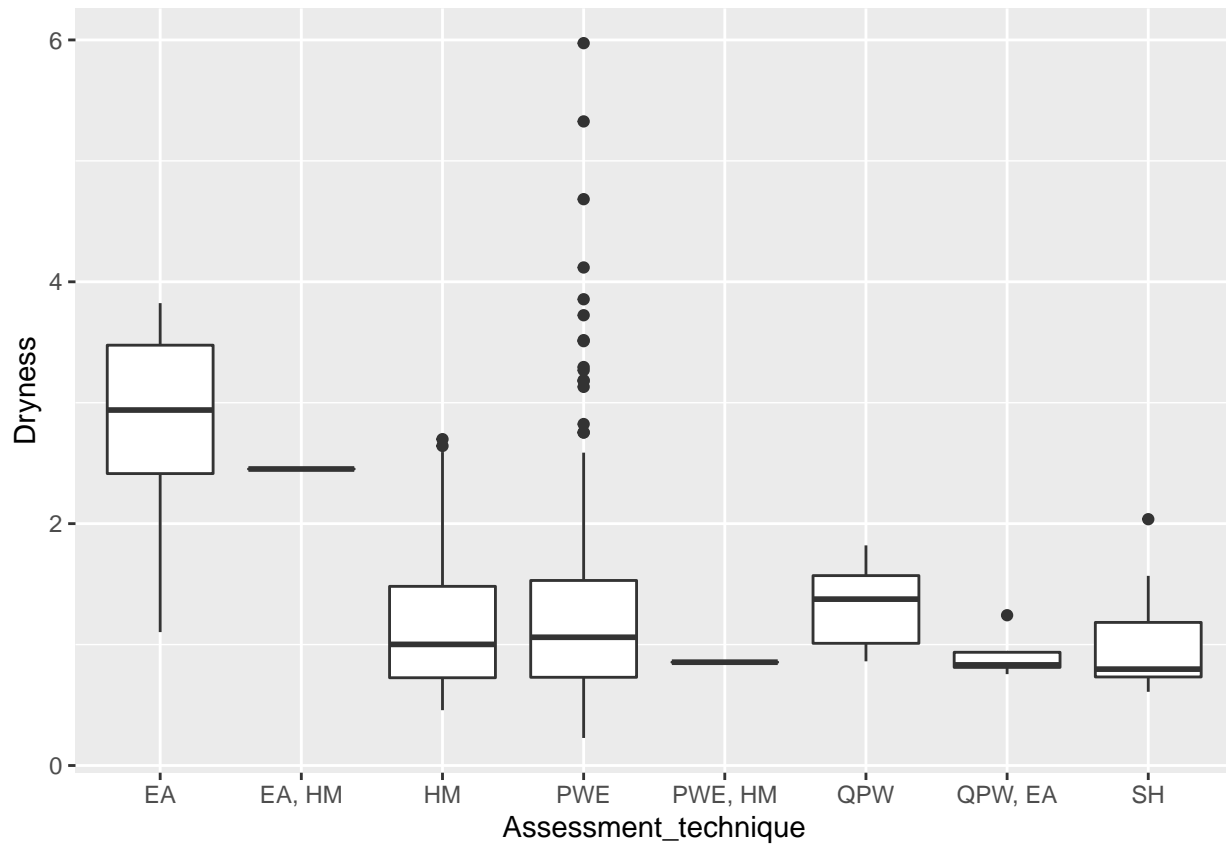
```
Zhang_all12 %>%  
  ggplot(aes(DeltaF_perc_pos, DeltaQf_perc, colour = Dryness)) + geom_point()
```



466

```
Zhang_all12 %>%  
  ggplot(aes(Assessment_technique,Dryness)) + geom_boxplot()
```

467 ## Warning: Removed 7 rows containing non-finite values (stat_boxplot).



468

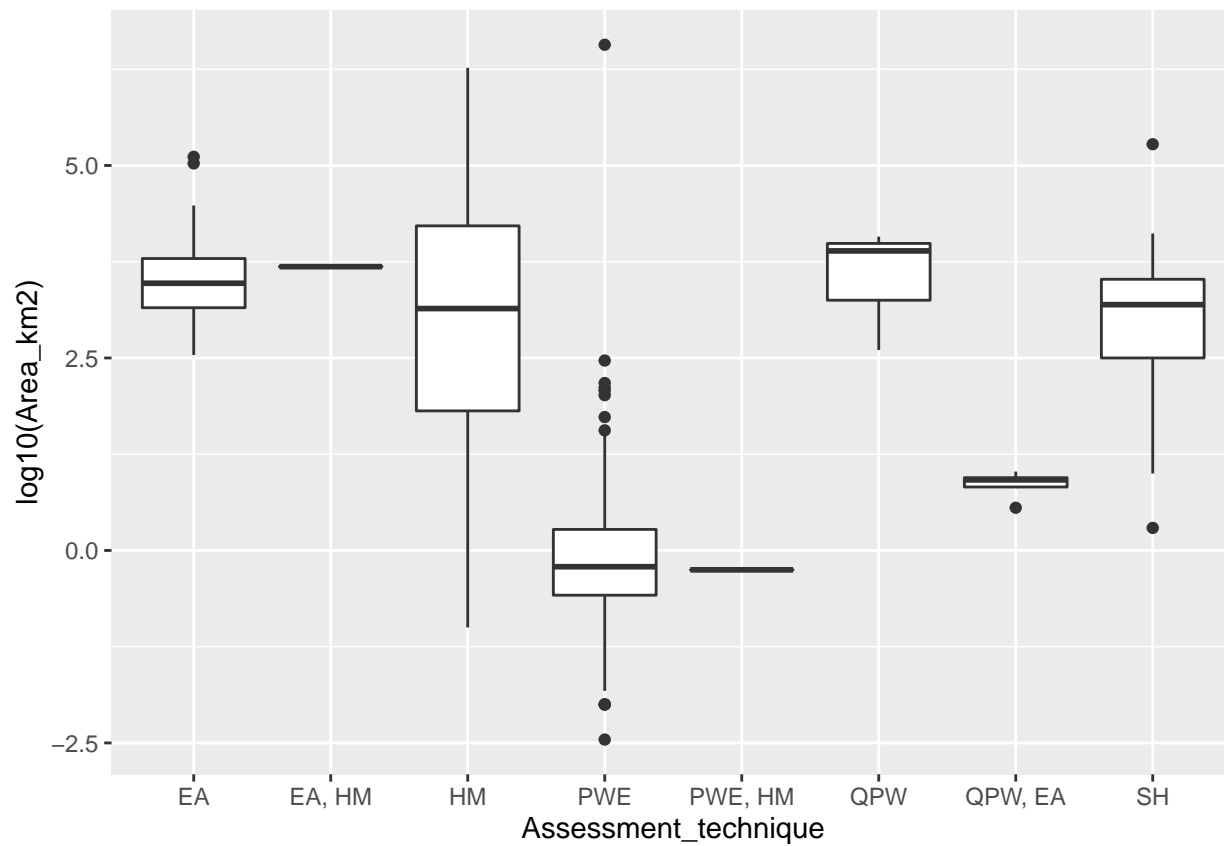
```
Zhang_all2 %>%
  filter(Dryness > 4)
```

```
469 ## # A tibble: 4 x 25
470 ##   `Watershed #` `Watershed name`      Area_km2 Pa_mm Forest_type Hydrological_re~
471 ##           <dbl> <chr>                <dbl> <dbl> <chr>          <chr>
472 ## 1           76 Beaver Creek, AZ ~      1.47  426 CF           SD
473 ## 2           90 Cantua                119    442 MF           RD
474 ## 3          295 White Spar, Ariz., ~      1    549 BF           RD
475 ## 4          225 Natural DRDages, Ar~    0.05  452 BF           RD
476 ## # ... with 19 more variables: DeltaF_perc <dbl>, DeltaQf_perc <dbl>,
477 ## # Precip_data_type <chr>, Assessment_technique <chr>, Source of info <chr>,
478 ## # Length of study (years) <dbl>, Notes <chr>, From <dbl>, To <dbl>,
479 ## # Location Details AL=aprox location given by author, PL=point location given by author,
480 ## # notes <chr>,
481 ## # Location Details AL=aprox location given by author, PL=point location given by author,
482 ## # Latitude <dbl>, Longitude <dbl>, E0 <dbl>, Dryness <dbl>,
483 ## # Forest_Sign <chr>, DeltaF_perc_pos <dbl>, length <dbl>
```

484

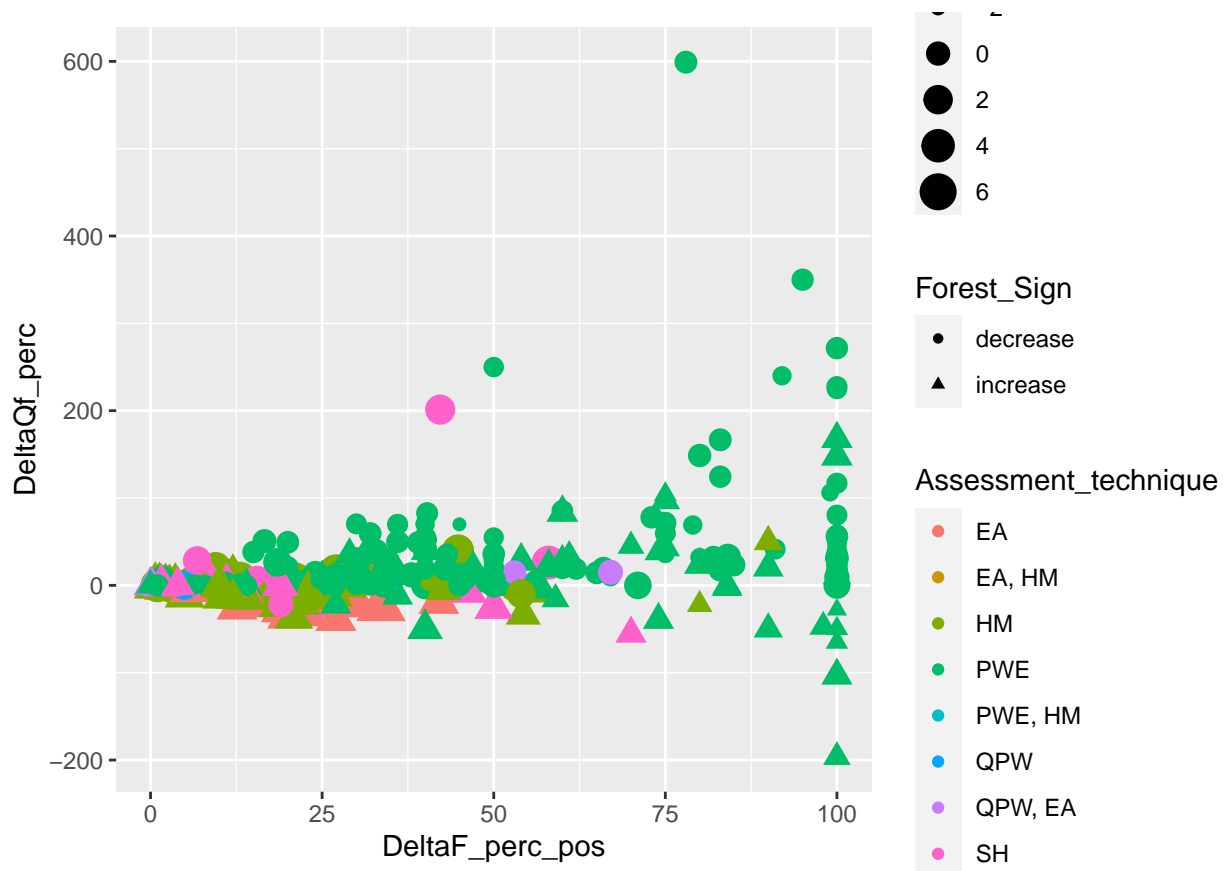
No evidence of effect of area

```
Zhang_all %>%
  ggplot(aes(Assessment_technique, log10(Area_km2))) + geom_boxplot()
```



485

```
Zhang_all12 %>%
  ggplot(aes(DeltaF_perc_pos, DeltaQf_perc, colour = Assessment_technique, size = log10(Area_
```



486

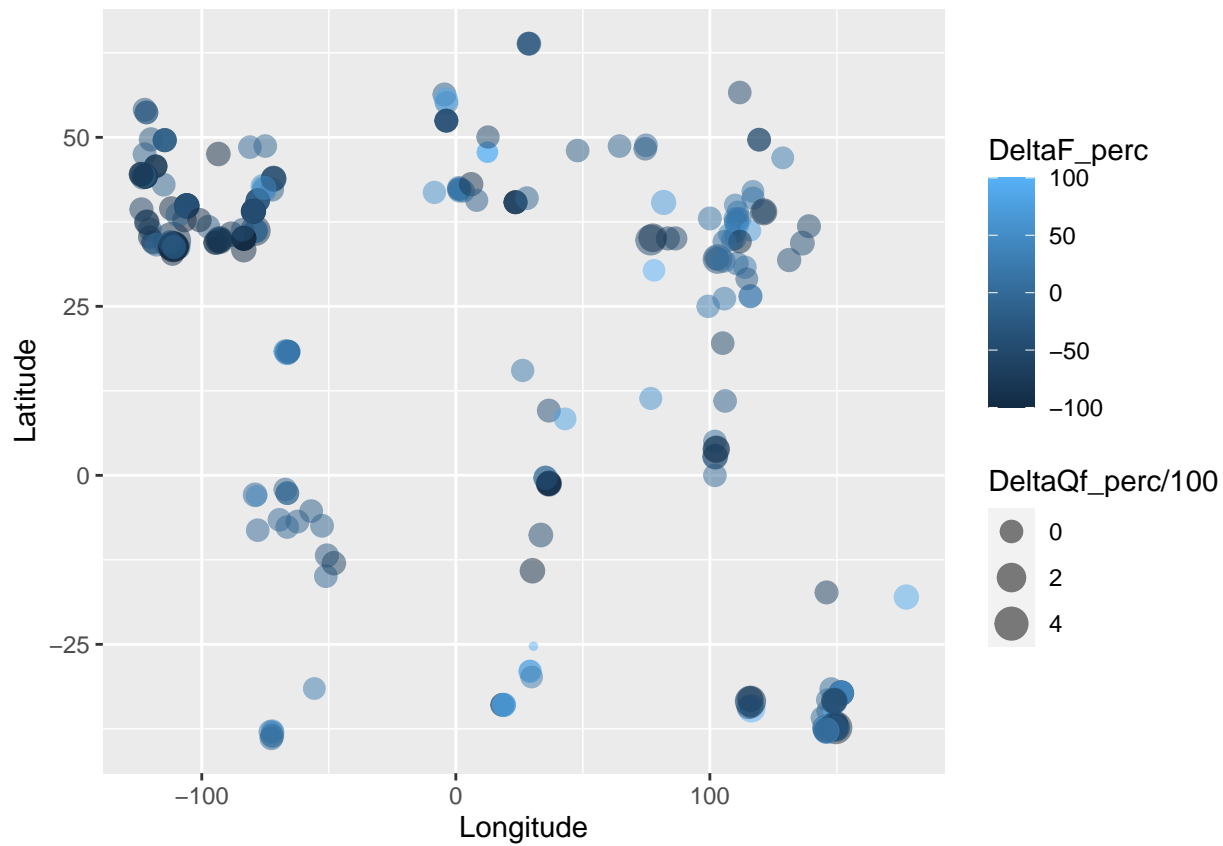
```
tiff("flow_forest_byArea.tiff", width = 2500, height = 1800, res = 300)
Zhang_all12 %>%
  ggplot(aes(DeltaF_perc_pos, DeltaQf_perc, colour = Assessment_technique, size = log10(Area),
             shape = Forest_Sign)) + geom_point(alpha = 0.5) +
  theme_bw() + ylab("% change in flow") +
  theme(axis.title = element_text(size = rel(2)),
        axis.text = element_text(size = rel(1.5))) +
  xlab("% change in forestry") + #scale_y_log10() +
  scale_size_continuous(name = "log10(Area in km2)") +
  scale_colour_discrete(name = "Assessment Technique") +
  scale_shape_discrete(name = "Forest cover direction")
dev.off()
```

487 ## pdf

488 ## 2

```
Zhang_all2 %>%
```

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



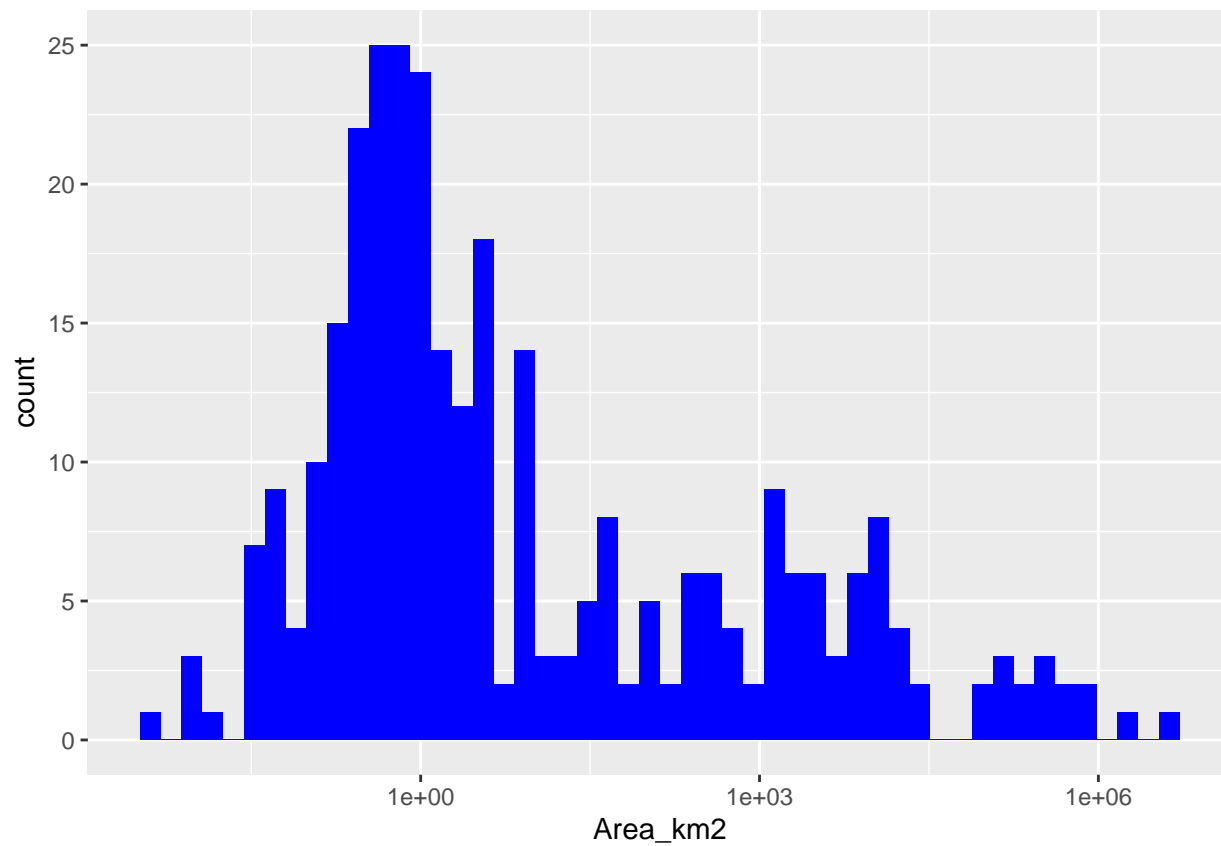
489

490

Check the size distribution of the catchments

```
Zhang_all %>%
```

```
ggplot(aes(Area_km2)) + geom_histogram(fill="blue", bins =50) +  
scale_x_log10()
```



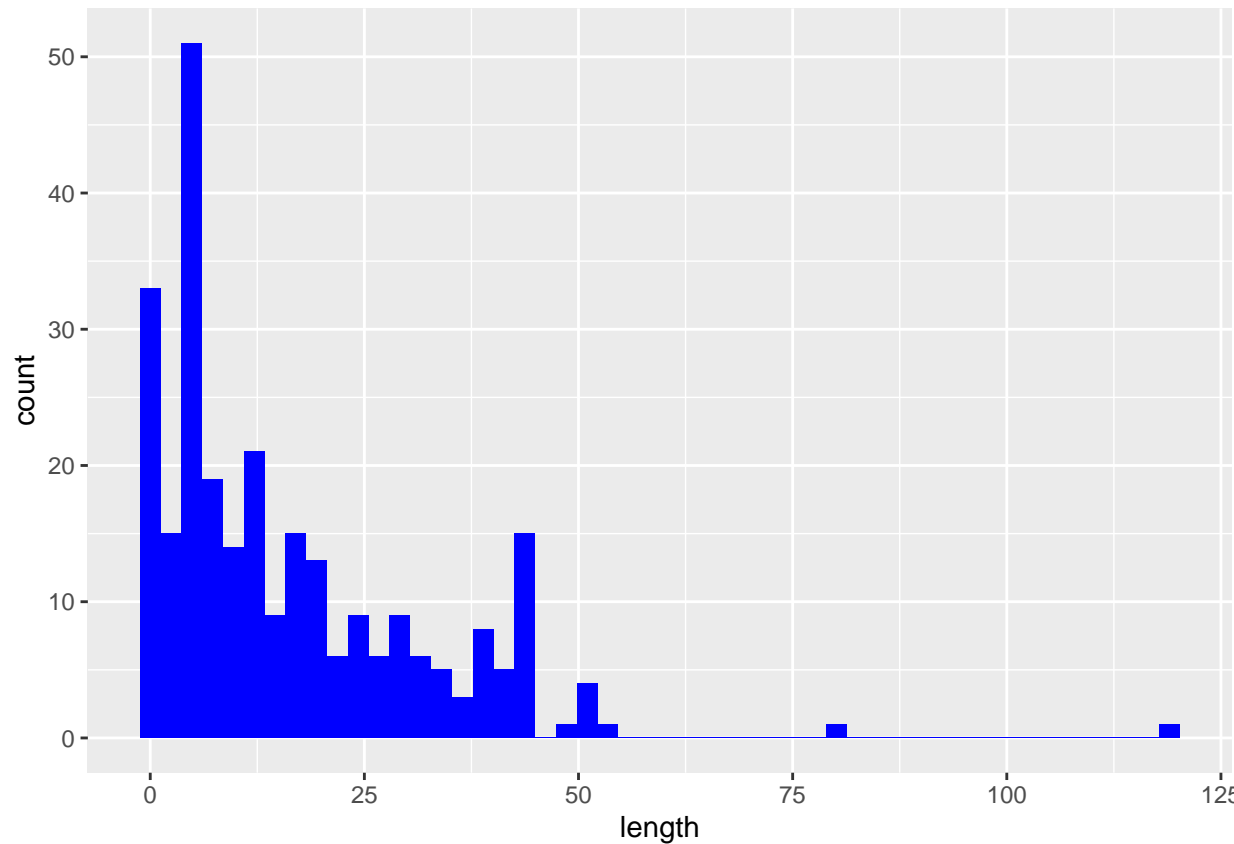
491

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

492 ## [1] 0.6570513

```
Zhang_all2 %>%
  ggplot(aes(length)) + geom_histogram(fill="blue", bins =50)
```

493 ## Warning: Removed 42 rows containing non-finite values (stat_bin).



494

495 Essentially, the analysis shows at the moment that there is no evidence that the
 496 size of the catchment influences the

497 References

- 498 Andréassian, V., 2004. Waters and forests: From historical controversy to
 499 scientific debate. *Journal of Hydrology* 291, 1–27. doi:<https://doi.org/10.1016/j.jhydrol.2003.12.015>
- 500 Borg, H., Bell, R.W., Loh, I.C., 1988. Streamflow and stream salinity in
 501 a small water supply catchment in southwest western australia after reforesta-
 502 tion. *Journal of Hydrology* 103, 323–333. doi:[https://doi.org/10.1016/0022-](https://doi.org/10.1016/0022-1694(88)90141-2)
 503 [1694\(88\)90141-2](https://doi.org/10.1016/0022-1694(88)90141-2)
- 504 Bosch, J.M., Hewlett, J.D., 1982. A review of catchment experiments to de-
 505 termine the effect of vegetation changes on water yield and evapotranspiration.
 506 *Journal of Hydrology* 55, 3–23.
- 507 Brown, A.E., Western, A.W., McMahon, T.A., Zhang, L., 2013. Impact of
 508 forest cover changes on annual streamflow and flow duration curves. *Journal of*
 509 *Hydrology* 483, 39–50. doi:<http://dx.doi.org/10.1016/j.jhydrol.2012.12.031>

510 Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W., Vertessy, R.A.,
511 2005. A review of paired catchment studies for determining changes in water
512 yield resulting from alterations in vegetation. *Journal of Hydrology* 310, 28–61.

513 Cosandey, C., Andréassian, V., Martin, C., Didon-Lescot, J.F., Lavabre,
514 J., Folton, N., Mathys, N., Richard, D., 2005. The hydrological impact of the
515 mediterranean forest: A review of french research. *Journal of Hydrology* 301,
516 235–249. doi:<https://doi.org/10.1016/j.jhydrol.2004.06.040>

517 Filoso, S., Bezerra, M.O., Weiss, K.C.B., Palmer, M.A., 2017. Impacts
518 of forest restoration on water yield: A systematic review. *PLOS ONE* 12,
519 e0183210. doi:[10.1371/journal.pone.0183210](https://doi.org/10.1371/journal.pone.0183210)

520 Jackson, R.B., Jobbagy, E.G., Avissar, R., Roy, S.B., Barrett, D.J., Cook,
521 C.W., Farley, K.A., Maitre, D.C. le, McCarl, B.A., Murray, B.C., 2005. Trading
522 water for carbon with biological carbon sequestration. *Science* 310, 1944–1947.
523 doi:[10.1126/science.1119282](https://doi.org/10.1126/science.1119282)

524 Peña-Arancibia, J.L., Dijk, A.I.J.M. van, Guerschman, J.P., Mulligan, M.,
525 Bruijnzeel, L.A., McVicar, T.R., 2012. Detecting changes in streamflow after
526 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>

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

530 Ruprecht, J.K., Schofield, N.J., Crombie, D.S., Vertessy, R.A., Stoneman,
531 G.L., 1991. Early hydrological response to intense forest thinning in southwest-
532 ern australia. *Journal of Hydrology* 127, 261–277. doi:[https://doi.org/10.1016/0022-](https://doi.org/10.1016/0022-1694(91)90118-2)
533 [1694\(91\)90118-2](https://doi.org/10.1016/0022-1694(91)90118-2)

534 Thornton, C.M., Cowie, B.A., Freebairn, D.M., Playford, C.L., 2007. The
535 brigalow catchment study: II*. Clearing brigalow (acacia harpophylla) for crop-
536 ping or pasture increases runoff. *Australian Journal of Soil Research* 45, 496–
537 511. doi:[doi:10.1071/SR07064](https://doi.org/10.1071/SR07064)

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

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

543 Zhang, L., Zhao, F., Chen, Y., Dixon, R.N.M., 2011. Estimating effects of
544 plantation expansion and climate variability on streamflow for catchments in
545 australia. *Water Resources Research* 47, W12539. doi:[10.1029/2011wr010711](https://doi.org/10.1029/2011wr010711)

546 Zhang, M., Liu, N., Harper, R., Li, Q., Liu, K., Wei, X., Ning, D., Hou, Y.,
547 Liu, S., 2017. A global review on hydrological responses to forest change across
548 multiple spatial scales: Importance of scale, climate, forest type and hydrological
549 regime. *Journal of Hydrology* 546, 44–59. doi:<https://doi.org/10.1016/j.jhydrol.2016.12.040>

550 Zhao, F., Zhang, L., Xu, Z., Scott, D.F., 2010. Evaluation of methods for
551 estimating the effects of vegetation change and climate variability on streamflow.
552 *Water Resources Research* 46, W03505. doi:[10.1029/2009wr007702](https://doi.org/10.1029/2009wr007702)

553 Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., Sun, G., Scott,
554 D.F., Zhou, S., Han, L., Su, Y., 2015. Global pattern for the effect of climate and

556 land cover on water yield. Nature Communications 6, 5918. doi:10.1038/ncomms6918
557 Zhou, G., Wei, X., Luo, Y., Zhang, M., Li, Y., Qiao, Y., Liu, H., Wang,
558 C., 2010. Forest recovery and river discharge at the regional scale of guangdong
559 province, china. Water Resources Research 46. doi:https://doi.org/10.1029/2009WR008829