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

R. Willem Vervoort[*,a,1], Eliana Nervi[b], Jimena Alonso[**,c]

[a]*School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia*
[b]*INIA, Uruguay*
[c]*Institute of Fluid Mechanics and Environmental Engineering, School of Engineering, Universidad de la República, 11200 Montevideo, Departamento de Montevideo, Uruguay*

Abstract

This is the abstract.

It consists of two paragraphs.

Introduction

Introduction

There has been an long and on-going discussion in the hydrologcal 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 stream flow 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 coded as count data and only a subset of 37 studies was analysed for actual water yield change.

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

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

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

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

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

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

As a result the objective of this paper is to 1) enhance the data set from Zhang et al. (2017) with further watersheds (such as from Filoso et al. (2017)) and spatial coordinates and 2) to analyse the possibility of non-linear, interactions and partial effects of the different factors and variables in the data using 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 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 km^2, the annual average precipitation (Pa) in mm, the forest type, hydrological regime, and climate type, the change in forest cover in % (ΔF%) and the change in streamflow in % (ΔQf(%), 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	elastictity 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 km^2) and large ($>= 1000$ 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 annual average potential evapotranspiration (E_{pot}) was extracted from the MODIS16 ET product, if a value of E0 was not available from the original papers. This involved downloading the MODIS product for PET at 500 m scale for the approximate catchment centroid using the package MODISTools via the MODIS/VIIRS subsets for the period 1 January 2000 - 31 December 2020. The average annual PET calculated from this series was used as the an approximation of average annual E0. 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).

Statistical modelling

```
## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion
```

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).

This methods section is no longer correct and needs rewriting

The first model applied in this analysis is based on the main hypothesis outlined above, can the change in streamflow be predicted from the change in forest cover, modulated by area, the length of the study and the climate.

$$\Delta\%Q \sim \ \Delta\%forest + Pa + Area + Latitude + Longitude + \varepsilon \tag{1}$$

However, the overall skewed distribution of the predictant ($\Delta\%Q$) is problematic, and this results in a skewed distribution of the GAM model residuals, which violates the linear model assumptions. As a result we transformed $\Delta\%Q$ back to fractions (0 - 1) and log transformed using $log10(x+1)$, where x is ΔQ. After transformation the model residuals approximate $\sim N(0,\sigma^2)$ and this results in the following equation:

$$\begin{aligned} log10(\frac{\Delta\%Q}{100}+1) \sim \ & \Delta\%forest\ cover + Pa + Area + Latitude + \\ & Longitude + \varepsilon \end{aligned} \tag{2}$$

A second model included all the variables in the analysis from Zhang et al. (2017) in one model:

$$\begin{aligned} log10(\frac{\Delta\%Q}{100}+1) \sim \ & \Delta\%forest\ cover + s(Pa, k=3) + s(Area, k=3) + \\ & forest\ type + climate\ type + assessment\ type + \\ & hydrologic\ regime + \varepsilon \end{aligned} \tag{3}$$

In this model, no direct interactions are assumed, and the assumption is that all continuous variables (such as Pa) can have a linear or non-linear relationship with $log10(\Delta Q)$. This means that a smooth function $s()$ is applied to the variable. To restrict the smoothness of the fit, the smoothness factor k is restricted to a value of 3 (Wood, 2006). This restriction was applied to smooth variables throughout this paper and we have dropped this from the notation in subsequent equations.

For the model in equation 3, we 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 analysis. Given that in Zhang et al. (2017), dryness ($\frac{E0}{Pa}$) is used to look at variations in the change in flow, we also fitted the following model:

$$\begin{aligned} log10(\frac{\Delta\%Q}{100}+1) \sim\ & \Delta\%forest\ cover + s(\frac{E0}{Pa}) + s(Area) + forest\ type + \\ & climate\ type + assessment\ type + \\ & hydrologic\ regime + \varepsilon \end{aligned} \tag{4}$$

Subsequently, using the full data set, including the additional watersheds and the additional variables the following two models were fitted:

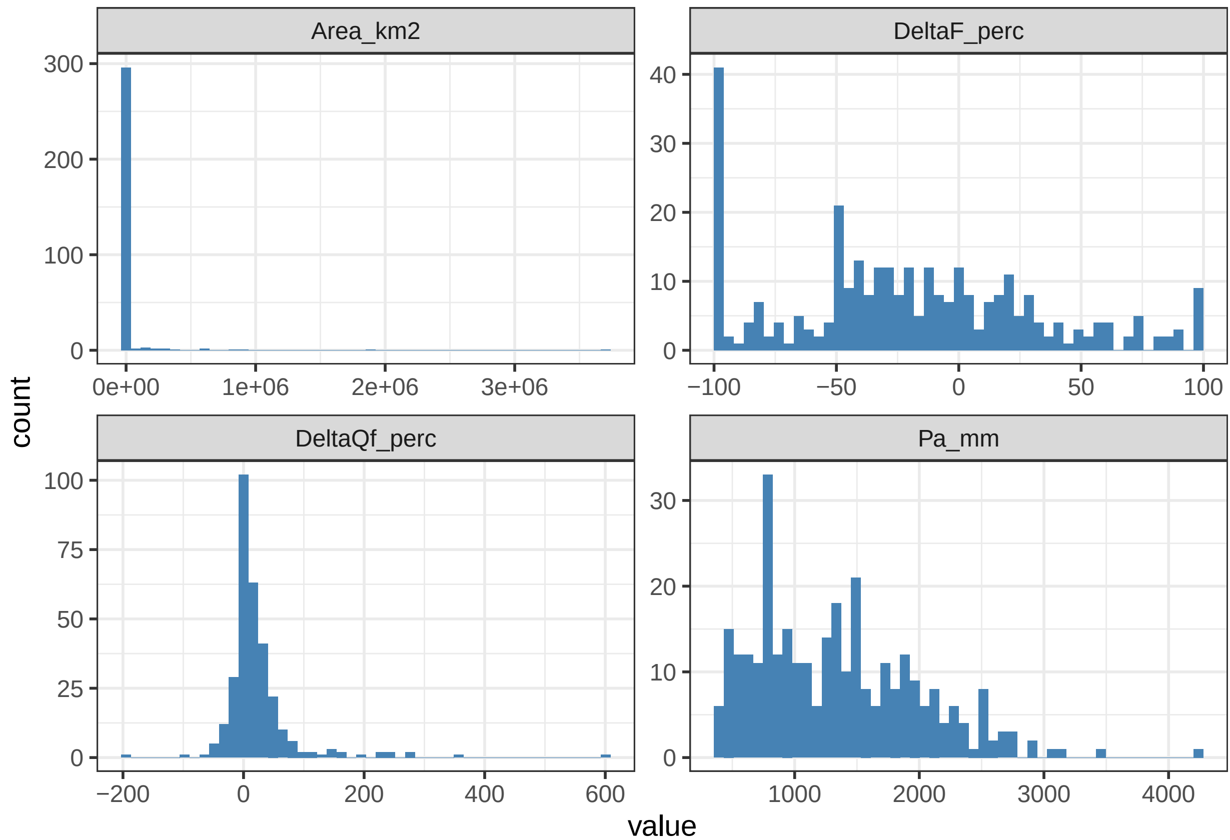
$$\begin{aligned} log10(\frac{\Delta\%Q}{100}+1) \sim\ & \Delta\%forest\ cover + s(Pa) + s(Area) + s(Latitude) + \\ & s(Longitude) + s(begin_{year}) + s(length_{study}) + \\ & forest\ type + climate\ type + assessment\ type + \\ & hydrologic\ regime + \varepsilon \end{aligned} \tag{5}$$

$$\begin{aligned} log10(\frac{\Delta\%Q}{100}+1) \sim\ & \Delta\%forest\ cover + s(\frac{E0}{Pa}) + s(Area) + \\ & s(Latitude) + s(Longitude) + s(begin_{year}) + \\ & s(length_{study}) + forest\ type + climate\ type + \\ & assessment\ type + hydrologic\ regime + \varepsilon \end{aligned} \tag{6}$$

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.

```
Zhang_all %>%
  pivot_longer(c(DeltaF_perc, DeltaQf_perc, Area_km2, Pa_mm),
               names_to = "variable", values_to = "value") %>%
  ggplot(aes(value)) + geom_histogram(fill = "steelblue", bins=50) + theme_bw() + facet_wrap
```

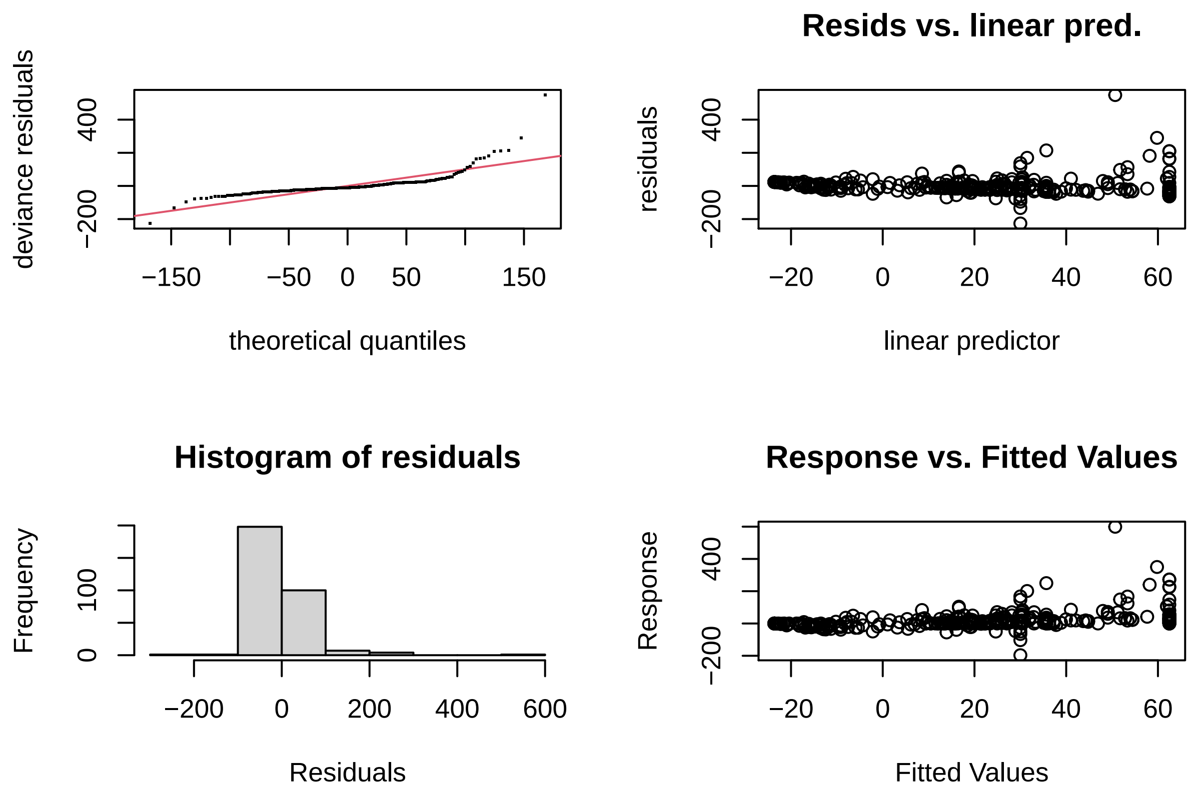


The changes in forest cover contain both positive (forestation) and negative values (deforestation). In (**???**) 2017, these changes were analysed jointly, which assumes that the effect on the change in flow is linear and non-hysteretic. However, it is a reasonable hypothesis that this is not the case, and that the impact of an increase in forest cover is different from the same fractional decrease in forest cover. To be able to analyse this difference, all the change in forest cover is converted to positive values, but an additional column is added that indicates whether it was a forest cover increase or decrease.

```
Zhang_all2 <- Zhang_all %>%
  mutate(Forest_Sign = ifelse(DeltaF_perc < 0,
                              "decrease", "increase"),
         DeltaF_perc_pos = ifelse(DeltaF_perc < 0,
                                  -1*DeltaF_perc,
                                  DeltaF_perc))
```

The initial relationship between change in forest cover and streamflow. We start of with a simple linear regression (i.e. following Zhang et al. (2017)) just looking at the change in flow as a result in the percent change forestry and including the direction of the change, increase in forest cover, or decrease in forest cover.

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## DeltaQf_perc ~ DeltaF_perc_pos + Forest_Sign
##
## Parametric coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           8.8466     6.5008   1.361    0.175
## DeltaF_perc_pos       0.5364     0.1017   5.274 2.52e-07 ***
## Forest_Signincrease -32.4760     6.9367  -4.682 4.26e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =   0.16   Deviance explained = 16.6%
## GCV = 3278.3  Scale est. = 3246.8    n = 312
```



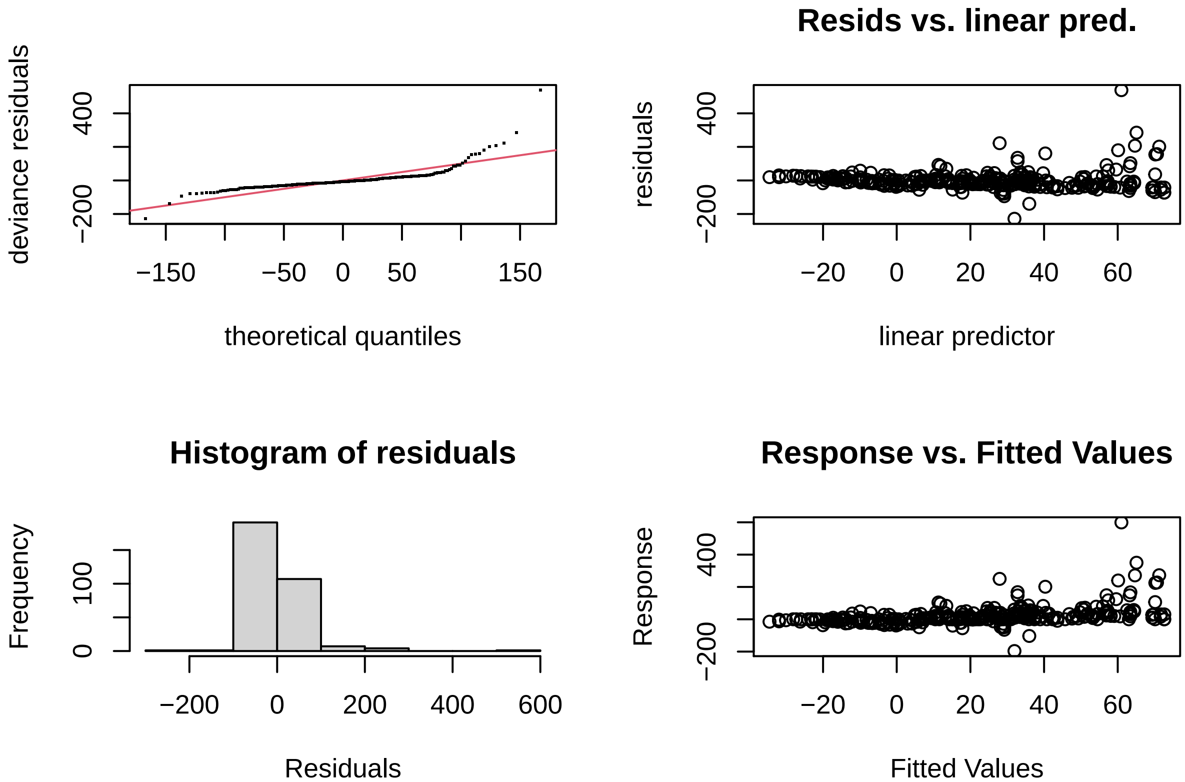

```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  3 / 3
```

While the overall variance explained in this model is not high at 0.16, it clearly indicates the hypothesised 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 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
```



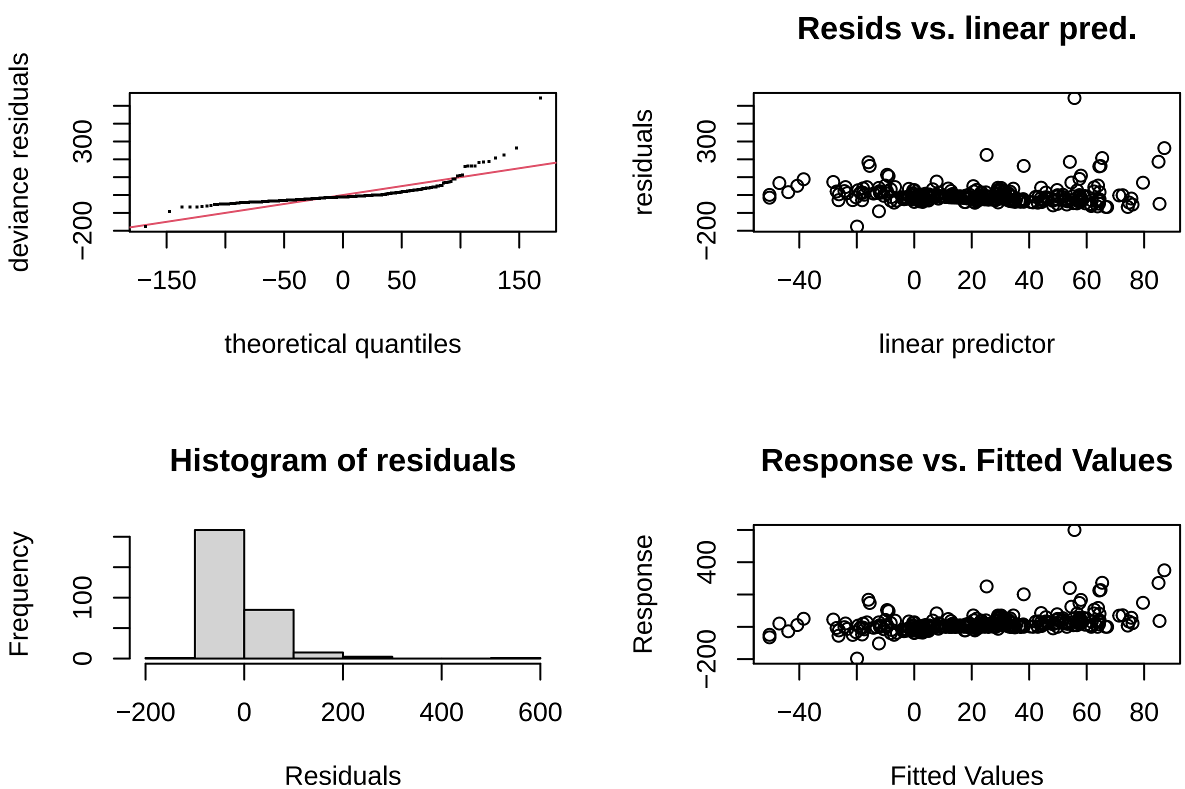

```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  5 / 5
```

Including area and annual precipitation does not really improve the overall explaining power of the model, in fact, annual precipitation appears to be only a very small confounding factor, representing only a -0.01/% partial effect in the change in streamflow, holding all other factors constant. The catchment area does not appear to have an effect at all in contrast to earlier reported studies (Filoso et al., 2017; Zhang et al., 2017). The main effects remain the change in forest cover and whether this is an increase or decrease

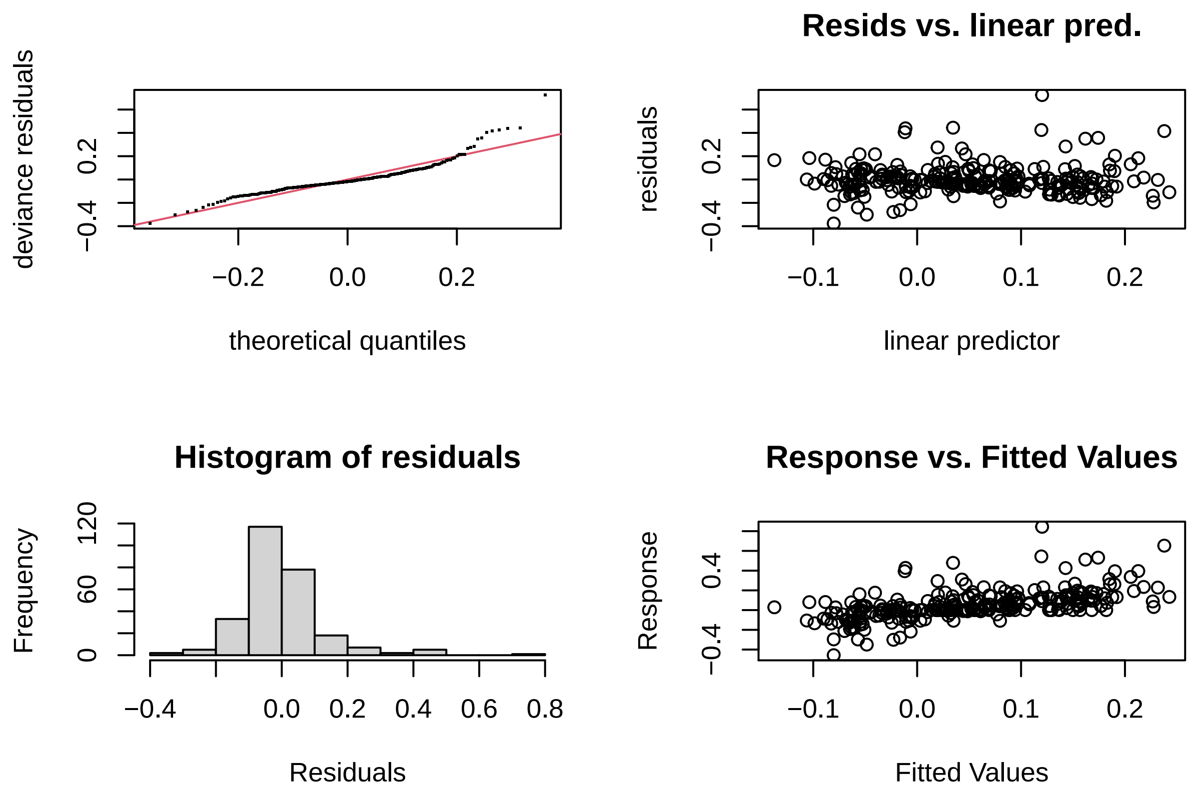
The effect of location on the globe.

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
```

```
## DeltaQf_perc ~ DeltaF_perc + Forest_Sign + Area_km2 + Pa_mm +
##     Latitude + Longitude
##
## Parametric coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.474e+01  1.055e+01   3.295  0.00110 **
## DeltaF_perc           -4.951e-01  1.064e-01  -4.652 4.95e-06 ***
## Forest_Signincrease -3.222e+00  1.160e+01  -0.278  0.78140
## Area_km2              -9.488e-06  1.339e-05  -0.709  0.47903
## Pa_mm                 -1.007e-02  5.309e-03  -1.897  0.05883 .
## Latitude              -3.806e-01  1.239e-01  -3.071  0.00233 **
## Longitude              1.074e-03  3.955e-02   0.027  0.97834
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.172   Deviance explained = 18.8%
## GCV = 3336.1  Scale est. = 3259.8    n = 306
```



```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  7 / 7
## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced
##
## Family: gaussian
## Link function: identity
##
## Formula:
## log10(DeltaQf_perc/100 + 1) ~ DeltaF_perc + log10(Area_km2) +
##     Pa_mm + Latitude + Longitude + From + length
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -6.962e-01  1.110e+00  -0.627 0.531002
## DeltaF_perc     -1.253e-03  1.612e-04  -7.772  1.8e-13 ***
## log10(Area_km2) -1.886e-02  5.186e-03  -3.637 0.000332 ***
## Pa_mm           -7.362e-06  1.257e-05  -0.586 0.558457
## Latitude        -1.070e-03  2.823e-04  -3.791 0.000186 ***
## Longitude        6.832e-05  9.635e-05   0.709 0.478879
## From             3.951e-04  5.604e-04   0.705 0.481473
## length          -3.424e-04  7.026e-04  -0.487 0.626447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.303   Deviance explained = 32.1%
## GCV = 0.016004  Scale est. = 0.015526  n = 268
```

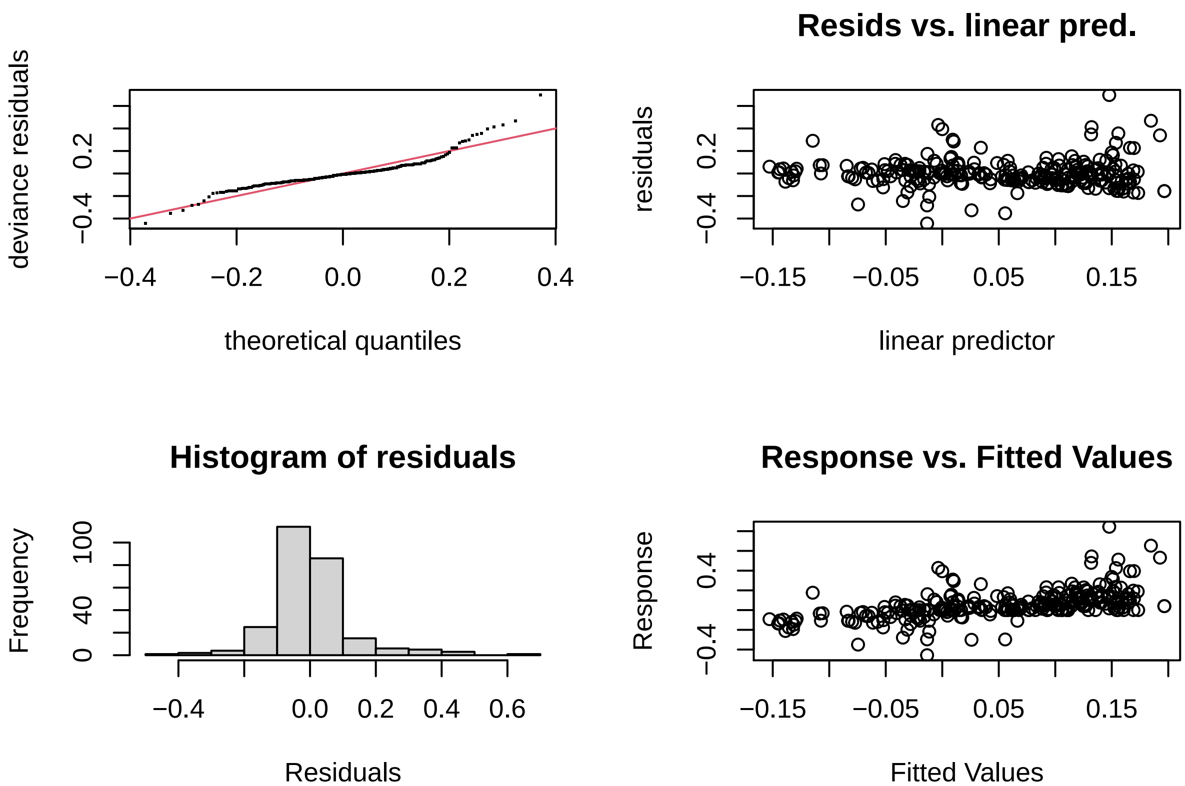



```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  8 / 8
```

```
## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced
```

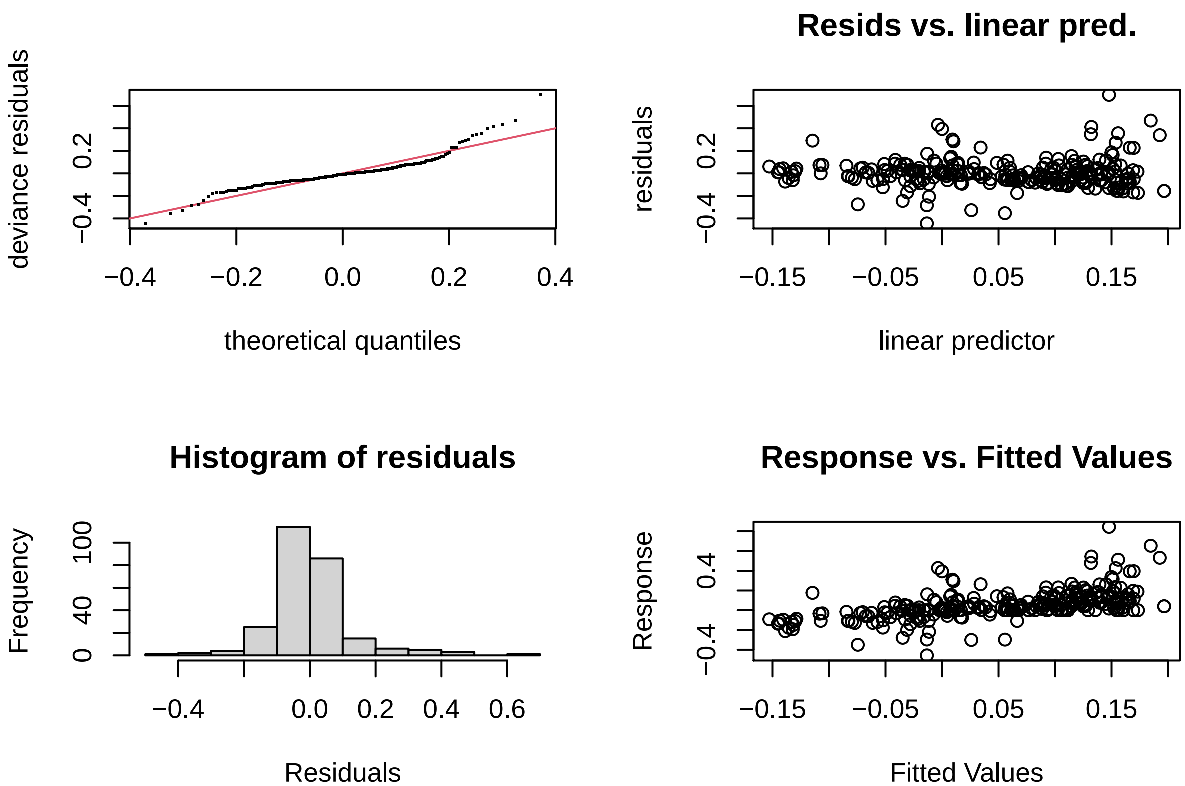
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## log10(DeltaQf_perc/100 + 1) ~ DeltaF_perc + s(Area_km2, k = 3) +
##     s(Pa_mm, k = 3) + s(From, k = 3) + s(length, k = 3) + Precip_data_type +
##     Assessment_technique + Forest_type + Hydrological_regime
##
## Parametric coefficients:
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -0.0962315  0.0558563  -1.723   0.0862 .
## DeltaF_perc                  -0.0008839  0.0001711  -5.165 4.98e-07 ***
## Precip_data_typeOB           -0.0329926  0.0418383  -0.789   0.4311
## Precip_data_typeSG            0.0595846  0.0474089   1.257   0.2100
## Assessment_techniqueEA, HM    0.0143199  0.1329271   0.108   0.9143
## Assessment_techniqueHM        0.0910165  0.0445991   2.041   0.0423 *
## Assessment_techniquePWE       0.2041286  0.0457201   4.465 1.22e-05 ***
## Assessment_techniquePWE, HM   0.0977846  0.1370941   0.713   0.4764
## Assessment_techniqueQPW       0.0850446  0.0731186   1.163   0.2459
## Assessment_techniqueSH        0.1066145  0.0473168   2.253   0.0251 *
## Forest_typeCF                -0.0029430  0.0268072  -0.110   0.9127
## Forest_typeMF                -0.0463517  0.0265852  -1.744   0.0825 .
## Hydrological_regimeSD         0.0111796  0.0319040   0.350   0.7263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df     F p-value
## s(Area_km2) 1.00  1.000 0.301  0.5839
## s(Pa_mm)    1.25  1.437 0.640  0.3574
## s(From)     1.00  1.000 3.194  0.0751 .
## s(length)   1.00  1.000 0.540  0.4631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.274   Deviance explained = 31.9%
## GCV = 0.017635  Scale est. = 0.016474  n = 262
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 8 iterations.
## The RMS GCV score gradient at convergence was 2.042593e-08 .
## The Hessian was positive definite.
## Model rank =  21 / 21
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                 k'  edf k-index p-value
## s(Area_km2) 2.00 1.00    0.97   0.265
## s(Pa_mm)    2.00 1.25    0.78  <2e-16 ***
## s(From)     2.00 1.00    0.85   0.005 **
## s(length)   2.00 1.00    0.90   0.060 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

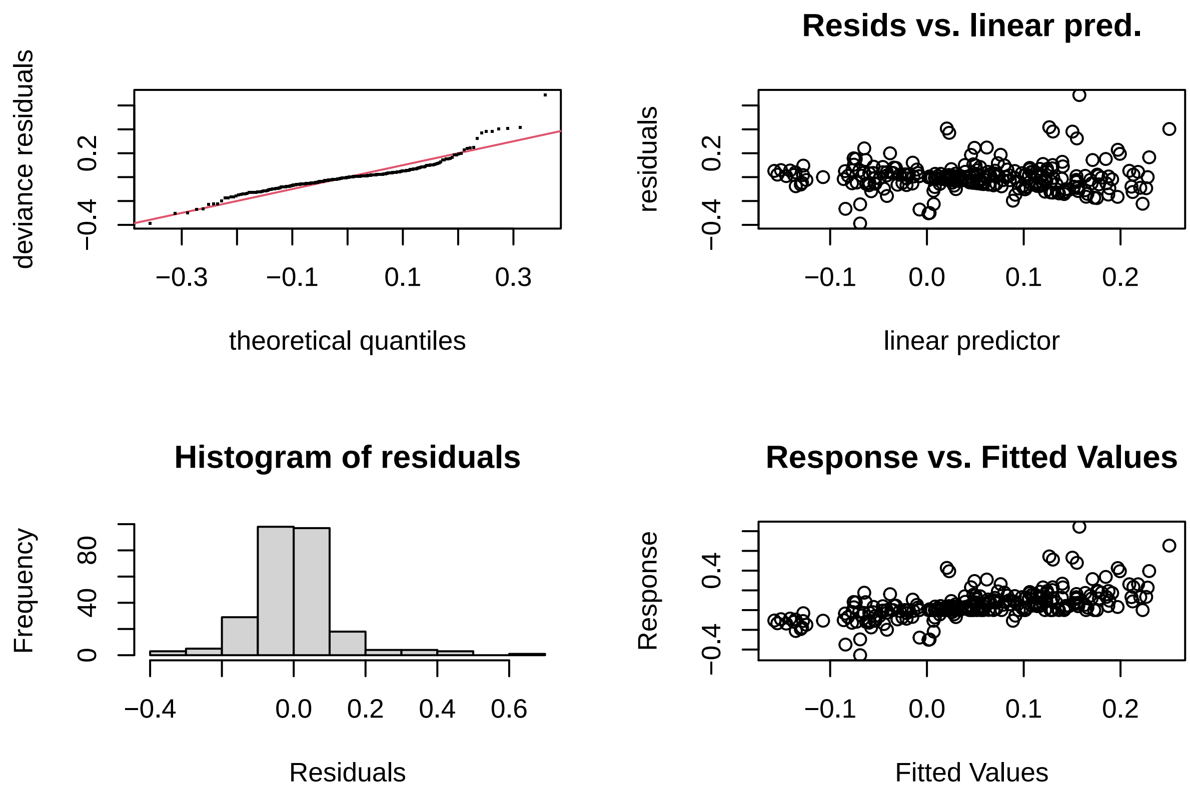
```
## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced

##
## Family: gaussian
## Link function: identity
##
## Formula:
## log10(DeltaQf_perc/100 + 1) ~ DeltaF_perc + s(Area_km2, k = 3) +
##     s(Pa_mm, k = 3) + s(From, k = 3) + s(length, k = 3) + Precip_data_type +
##     Assessment_technique + Forest_type + Hydrological_regime
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -0.0962315  0.0558563  -1.723   0.0862 .
## DeltaF_perc                 -0.0008839  0.0001711  -5.165 4.98e-07 ***
## Precip_data_typeOB          -0.0329926  0.0418383  -0.789   0.4311
## Precip_data_typeSG           0.0595846  0.0474089   1.257   0.2100
## Assessment_techniqueEA, HM   0.0143199  0.1329271   0.108   0.9143
## Assessment_techniqueHM       0.0910165  0.0445991   2.041   0.0423 *
## Assessment_techniquePWE      0.2041286  0.0457201   4.465 1.22e-05 ***
## Assessment_techniquePWE, HM  0.0977846  0.1370941   0.713   0.4764
## Assessment_techniqueQPW      0.0850446  0.0731186   1.163   0.2459
## Assessment_techniqueSH       0.1066145  0.0473168   2.253   0.0251 *
## Forest_typeCF               -0.0029430  0.0268072  -0.110   0.9127
## Forest_typeMF               -0.0463517  0.0265852  -1.744   0.0825 .
## Hydrological_regimeSD        0.0111796  0.0319040   0.350   0.7263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df     F p-value
## s(Area_km2) 1.00  1.000 0.301  0.5839
## s(Pa_mm)    1.25  1.437 0.640  0.3574
## s(From)     1.00  1.000 3.194  0.0751 .
## s(length)   1.00  1.000 0.540  0.4631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.274   Deviance explained = 31.9%
## GCV = 0.017635  Scale est. = 0.016474  n = 262
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 8 iterations.
## The RMS GCV score gradient at convergence was 2.042593e-08 .
## The Hessian was positive definite.
## Model rank =  21 / 21
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                k'  edf k-index p-value
## s(Area_km2) 2.00 1.00    0.97   0.280
## s(Pa_mm)    2.00 1.25    0.78  <2e-16 ***
## s(From)     2.00 1.00    0.85   0.015 *
## s(length)   2.00 1.00    0.90   0.040 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

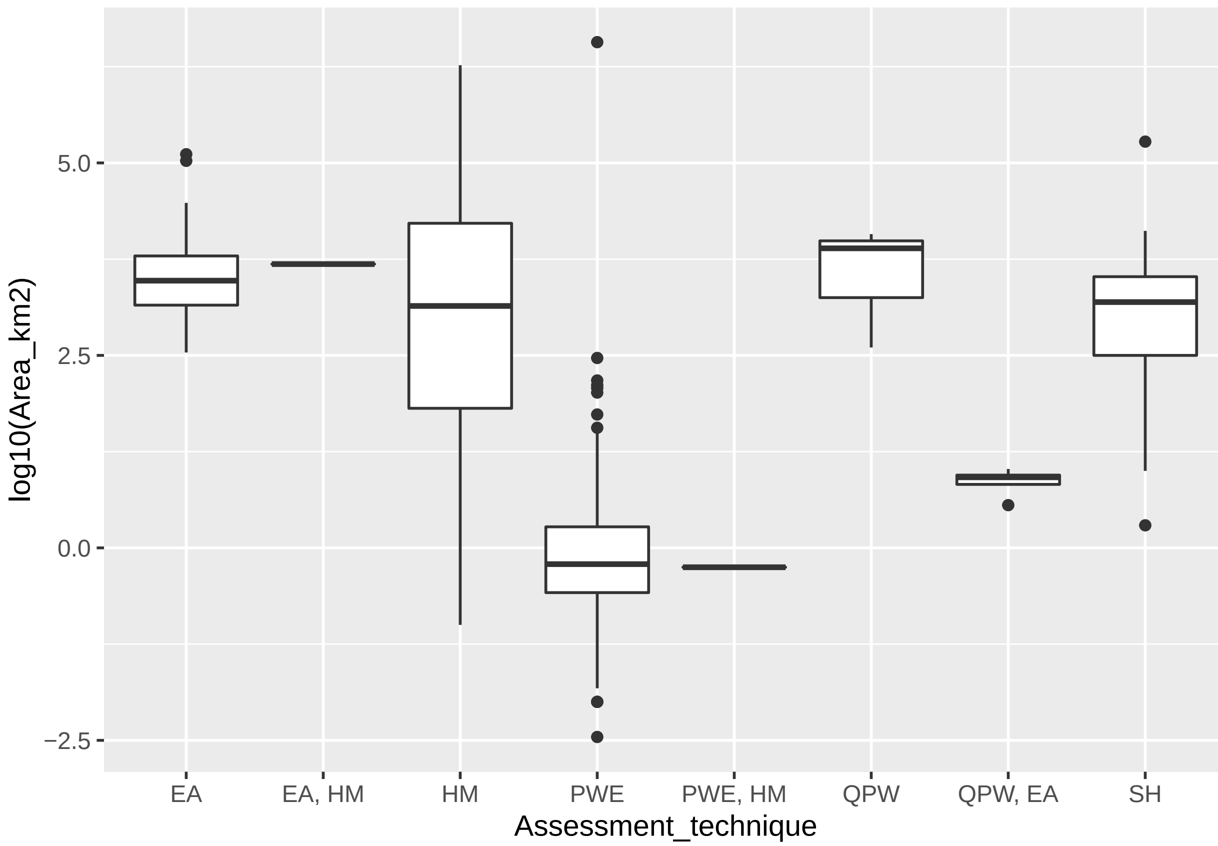
No evidence of effect of area

```
## Warning in eval(predvars, data, env): NaNs produced

## Warning in eval(predvars, data, env): NaNs produced

##
## Family: gaussian
## Link function: identity
##
## Formula:
## log10(DeltaQf_perc/100 + 1) ~ DeltaF_perc + log10(Area_km2) +
##     s(Pa_mm, k = 3) + From + length + Precip_data_type + Assessment_technique +
##     Forest_type + Hydrological_regime + Latitude + Longitude
##
## Parametric coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -1.2633792  1.1731739  -1.077  0.28260
## DeltaF_perc                 -0.0010709  0.0001693  -6.324 1.21e-09 ***
## log10(Area_km2)             -0.0106973  0.0079411  -1.347  0.17921
## From                         0.0006320  0.0005902   1.071  0.28528
## length                       0.0002020  0.0007979   0.253  0.80032
## Precip_data_typeOB          -0.0496200  0.0405500  -1.224  0.22226
## Precip_data_typeSG           0.0379293  0.0474175   0.800  0.42455
## Assessment_techniqueEA, HM   0.0190552  0.1279391   0.149  0.88173
## Assessment_techniqueHM       0.0665563  0.0447830   1.486  0.13852
## Assessment_techniquePWE      0.1335480  0.0529257   2.523  0.01226 *
## Assessment_techniquePWE, HM  0.0724841  0.1341116   0.540  0.58936
## Assessment_techniqueQPW      0.0585720  0.0713099   0.821  0.41224
## Assessment_techniqueSH       0.0649925  0.0482180   1.348  0.17895
## Forest_typeCF                0.0129954  0.0267350   0.486  0.62735
## Forest_typeMF               -0.0171200  0.0262238  -0.653  0.51448
## Hydrological_regimeSD        0.0425574  0.0316727   1.344  0.18031
## Latitude                    -0.0010038  0.0003307  -3.035  0.00267 **
## Longitude                    0.0001756  0.0001025   1.713  0.08803 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf Ref.df     F p-value
## s(Pa_mm) 1.004  1.008 1.285   0.256
##
## R-sq.(adj) =  0.328   Deviance explained = 37.5%
## GCV = 0.016441  Scale est. = 0.015249  n = 262
```

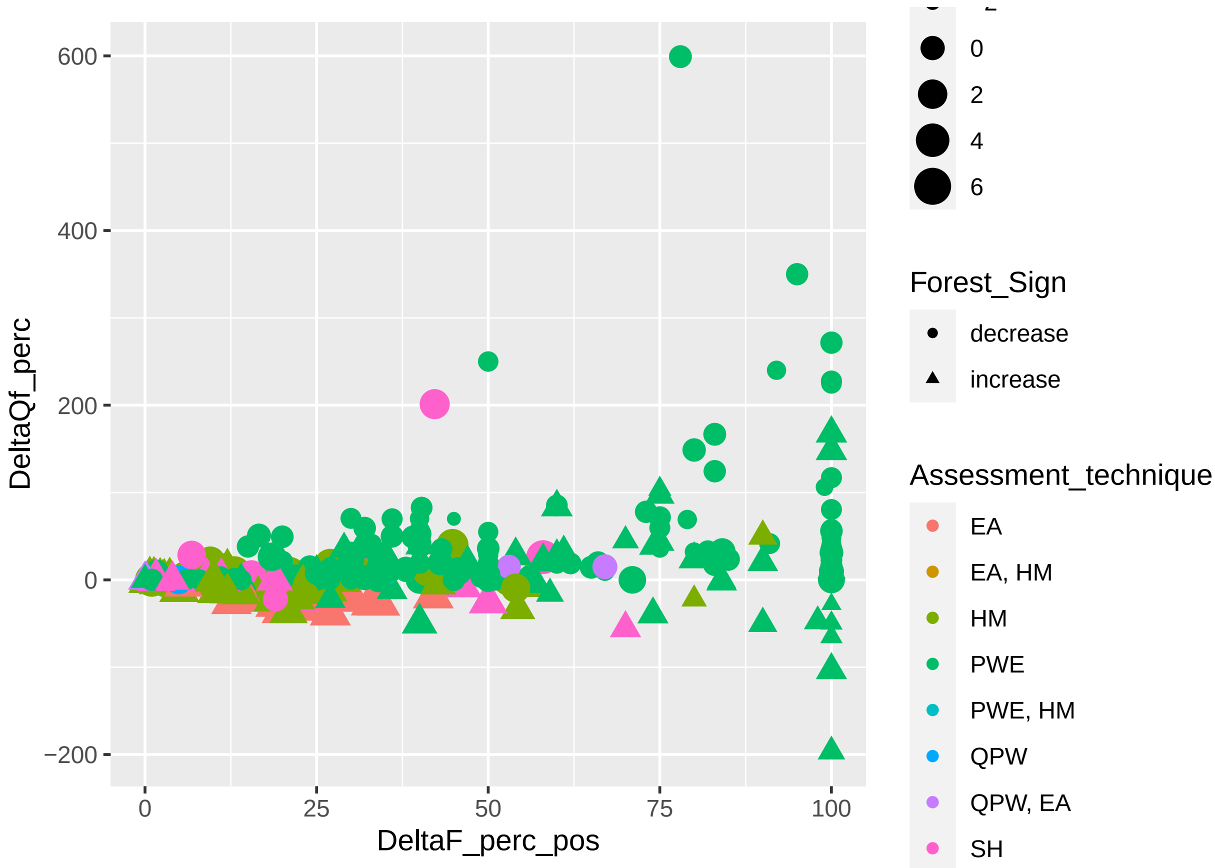



```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 5 iterations.
## The RMS GCV score gradient at convergence was 2.648312e-07 .
## The Hessian was positive definite.
## Model rank =  20 / 20
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##            k' edf k-index p-value
## s(Pa_mm)  2   1     0.8   0.005 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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




```
Zhang_all2 %>%
  ggplot(aes(DeltaF_perc_pos, DeltaQf_perc, colour = Assessment_technique,size = log10(Area_
```

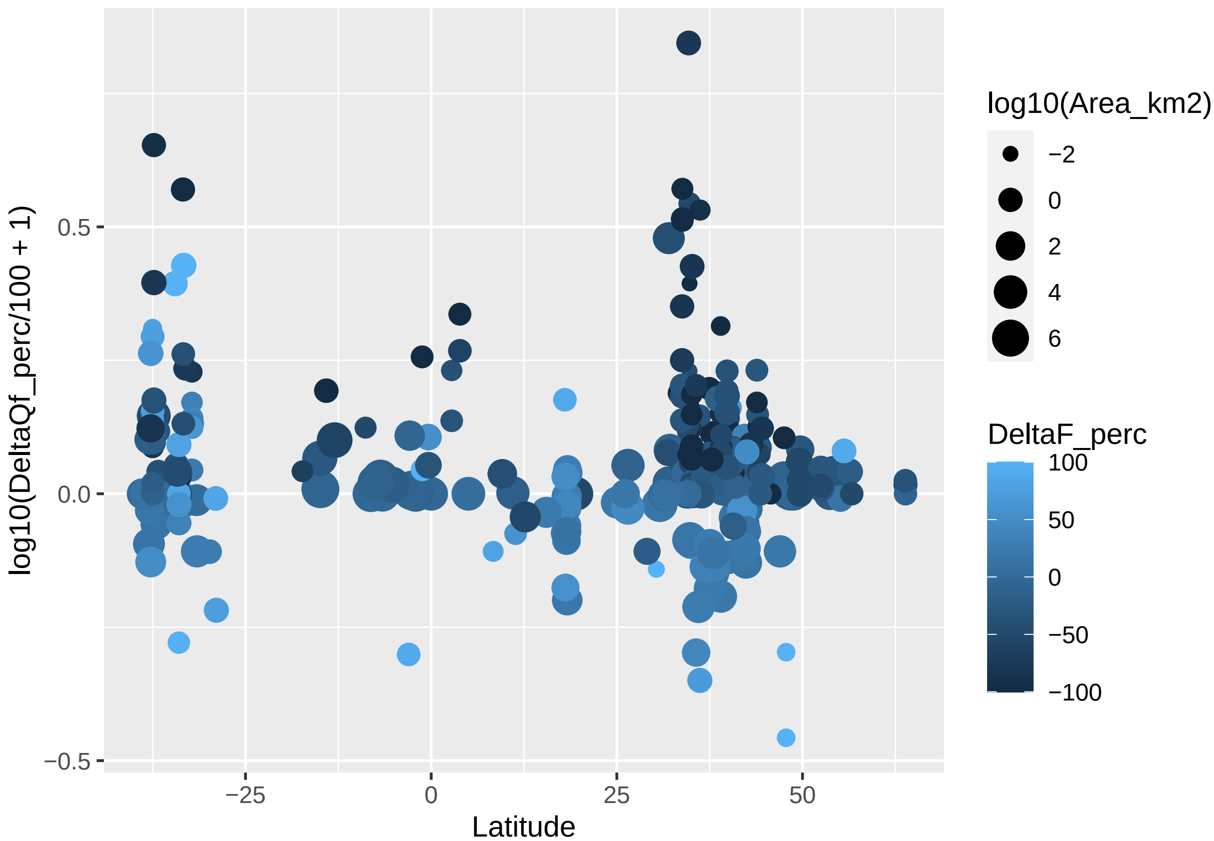



```
tiff("flow_forest_byArea.tiff", width = 2500, height = 1800, res = 300)
Zhang_all2 %>%
  ggplot(aes(DeltaF_perc_pos, DeltaQf_perc, colour = Assessment_technique,size = log10(Area_
             shape = Forest_Sign)) + geom_point(alpha = 0.5) +
  theme_bw() + ylab("% change in flow") +
  theme(axis.title = 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()
```

```
## pdf
##   2
```

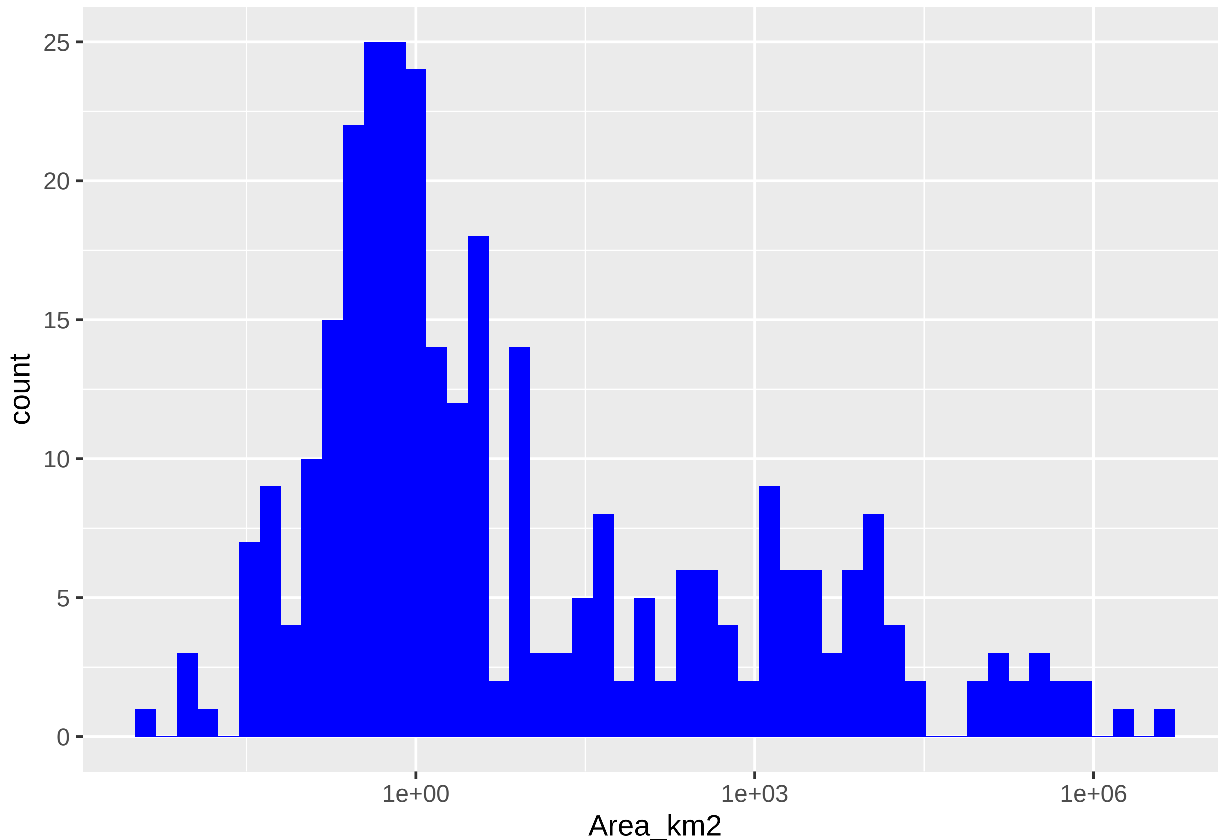
```
Zhang_all %>%
  ggplot(aes(Latitude, log10(DeltaQf_perc/100 + 1), colour = DeltaF_perc,size = log10(Area_k
```

```
## Warning in FUN(X[[i]], ...): NaNs produced

## Warning in FUN(X[[i]], ...): NaNs produced

## Warning: Removed 8 rows containing missing values (geom_point).
```



Check the size distribition of the catchments

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

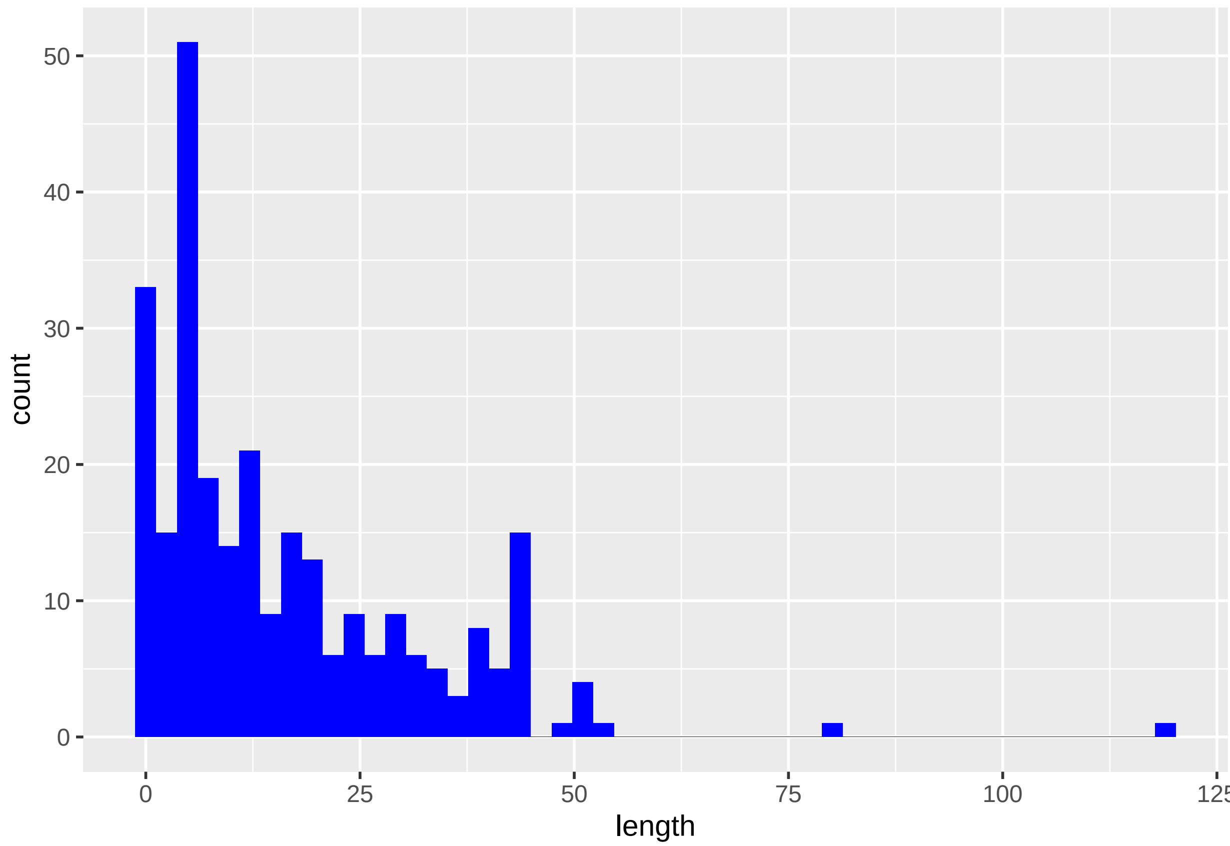



```
total <- nrow(Zhang_all)
length(Zhang_all$Area_km2[Zhang_all$Area_km2<10])/total
```

```
## [1] 0.6570513
```

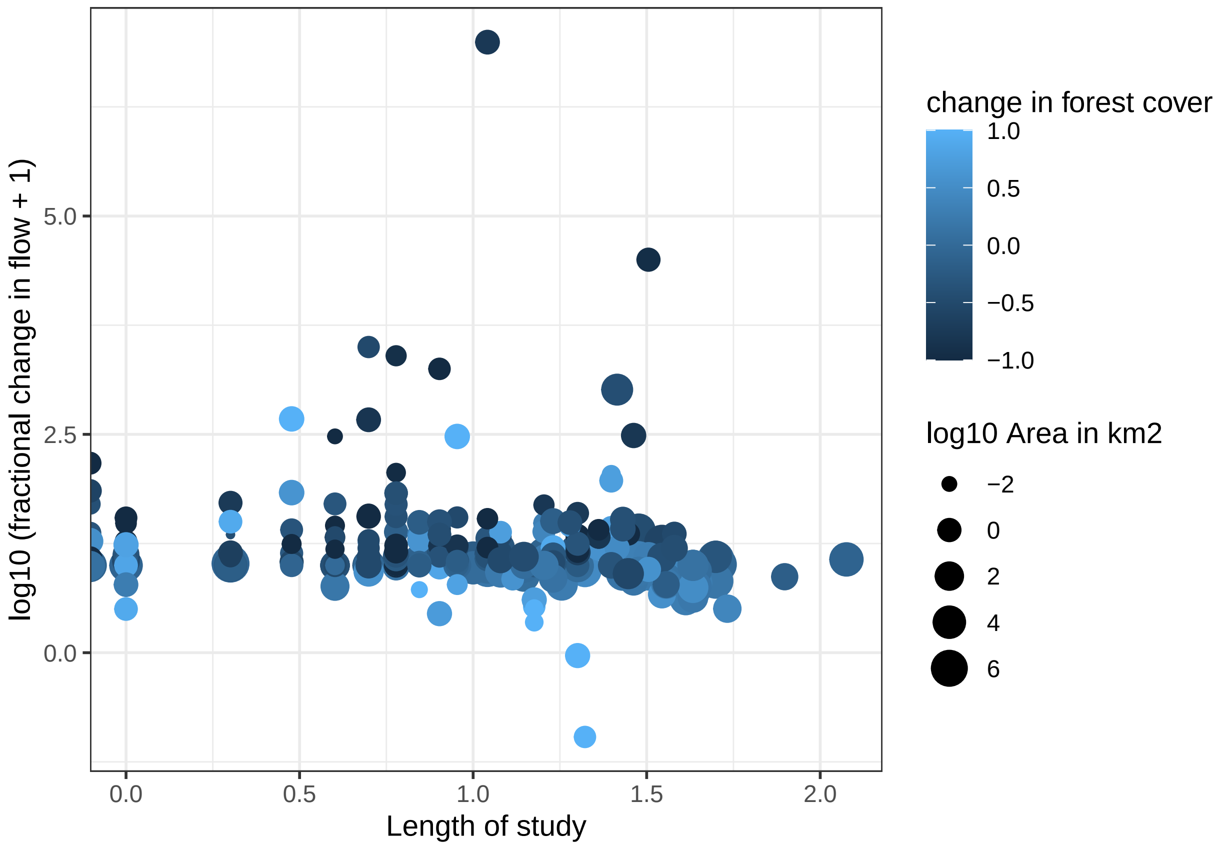
```
Zhang_all %>%
  ggplot(aes(length)) + geom_histogram(fill="blue", bins =50)
```

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




```r
#windows()
Zhang_all %>%
  ggplot(aes(log10(length), (DeltaQf_perc/100 + 1), colour = DeltaF_perc/100, size = log10(A
  theme_bw() + ylab("log10 (fractional change in flow + 1)") +
  xlab("Length of study") + scale_size_continuous(name = "log10 Area in km2") +
  scale_colour_continuous(name = "change in forest cover")
```

```
## Warning: Removed 42 rows containing missing values (geom_point).
```



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

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

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

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

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

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

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

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

Zhang, L., Zhao, F., Chen, Y., Dixon, R.N.M., 2011. Estimating effects of plantation expansion and climate variability on streamflow for catchments in australia. Water Resources Research 47, W12539. doi:10.1029/2011wr010711

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

Zhao, F., Zhang, L., Xu, Z., Scott, D.F., 2010. Evaluation of methods for estimating the effects of vegetation change and climate variability on streamflow. Water Resources Research 46, W03505. doi:10.1029/2009wr007702

Zhou, G., Wei, X., Chen, X., Zhou, P., Liu, X., Xiao, Y., Sun, G., Scott, D.F., Zhou, S., Han, L., Su, Y., 2015. Global pattern for the effect of climate and land cover on water yield. Nature Communications 6, 5918. doi:10.1038/ncomms6918

Zhou, G., Wei, X., Luo, Y., Zhang, M., Li, Y., Qiao, Y., Liu, H., Wang, C., 2010. Forest recovery and river discharge at the regional scale of guangdong province, china. Water Resources Research 46. doi:https://doi.org/10.1029/2009WR008829