Testing Forest Area as a random effect

R. Willem Vervoort[1], Eliana Nervi, Jimena Alonso

Abstract

This supplementary material file compares different linear and non-linear relationships for the impact of forest cover on the change in stream flow.

1. Introduction

This supplementary material is related to 'Generalising the impact of forest cover on streamflow from experimental data: it is not that simple. Vervoort et al.'

This document outlines a test suggested in the first review of the manuscript. The AE of the Journal of Hydrology suggested that we test if including the total forest area as a variable would make sense. Not only the % change in area is important, but what total area of forest was there in the first place.

This test focusses on the large catchment data (< 1000 km^2) for two reasons:

1. This was a manageable subset and many of the papers are fairly accessible, in contrast to the small catchments.
2. Many of the small catchments have 100% area covered in forest, so are not useful to identify if total area of forest has an impact on the change in flow.

1.1. set up the packages

1.2. read in the data

Read in the data and massage column names to make columns useable in the model.

```
Data_wF <- read_csv("../../data/LCdatawithFA.xlsx - LW_TotForest_Area.csv")
```

```
## Rows: 61 Columns: 25
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): Watershed name, Forest type, Hydrological regime, yearFarea, Comme...
## dbl (15): Watershed #, Area(km2), Pa(mm), Farea_km2, Perc_Farea_pre, DeltaF_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

*Corresponding author

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

[1]Corresponding Author

```r
# change some column names
names(Data_wF)[3:4] <- c("Area_km2", "Pa_mm")
names(Data_wF)[5:6] <- c("Forest_type", "Hydrological_regime")
names(Data_wF)[13:14] <- c("Precip_data_type", "Assessment_technique")
```

1.3. Some changes to the overall data

1. calculating the dryness

```r
# calculate dryness index
Data_wF <- Data_wF %>%
  mutate(Dryness = E0/Pa_mm)
```

2. remove watershed 1 (the Amazon) from the analysis

```r
Data_wF <- Data_wF %>%
  filter(`Watershed #` != 1)
```

3. include length as a variable

```r
All_data <- Data_wF %>%
  mutate(length = To - From,
         mid_year = From + (To - From)/2)
```

2. Results

Run the model simply as a GLM, transform Area using log10, as in the manuscript.

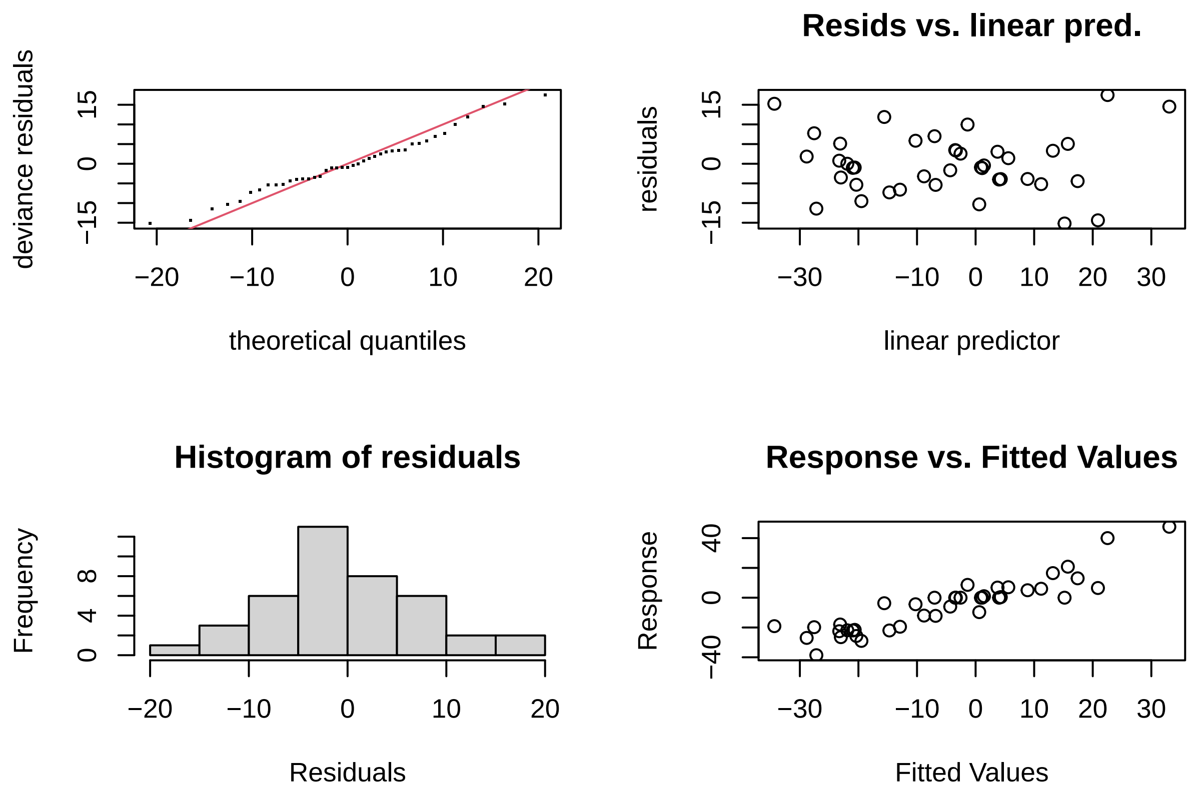
```r
Forest_model_all <- gam(DeltaQf_perc ~ DeltaF_perc +
                    log10(Area_km2) +
                    Dryness +
                    Perc_Farea_pre +
                    Precip_data_type +  Assessment_technique +
                    Forest_type +
                    Hydrological_regime
                    , data = All_data)
summary(Forest_model_all)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## DeltaQf_perc ~ DeltaF_perc + log10(Area_km2) + Dryness + Perc_Farea_pre +
```

```
##     Precip_data_type + Assessment_technique + Forest_type + Hydrological_regime
##
## Parametric coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -16.04690   15.08196  -1.064  0.29676
## DeltaF_perc              -0.37440    0.11561  -3.239  0.00318 **
## log10(Area_km2)           1.55899    2.31470   0.674  0.50634
## Dryness                   4.40194    3.21852   1.368  0.18269
## Perc_Farea_pre            0.03722    0.07544   0.493  0.62572
## Precip_data_typeOB      -19.28048    6.55235  -2.943  0.00661 **
## Precip_data_typeSG      -12.00304    6.75886  -1.776  0.08702 .
## Assessment_techniqueEA, HM  1.36550    9.86348   0.138  0.89092
## Assessment_techniqueHM   15.44379    6.16998   2.503  0.01866 *
## Assessment_techniqueQPW   2.79556    9.90330   0.282  0.77988
## Assessment_techniqueSH   18.51105    7.64697   2.421  0.02249 *
## Forest_typeCF            -9.46812    6.16583  -1.536  0.13628
## Forest_typeMF             0.27749    4.64399   0.060  0.95279
## Hydrological_regimeSD    20.97769    6.66044   3.150  0.00397 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.735   Deviance explained = 82.1%
## GCV = 128.37  Scale est. = 84.534    n = 41
```

This suggest that forest area is not at all significant in the model. In fact, the hydrological modelling technique is much more important. Overall the model explains roughly 82% of the variation.

```
gam.check(Forest_model_all)
```




```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  14 / 14
```

```
#plot(Forest_model_all)
```

The fit of the model is also well-behaved, as would be expected. The data are mostly from hydrological modelling, stabilising the response of the change in forest to change in streamflow.

2.1. run the model on forest area rather than %

Possibly, using the total % forest is not taking into account the differences in total area of forest. So we can also test with total ha. Note that this can interact with total area: larger catchments will have larger total areas of forest.

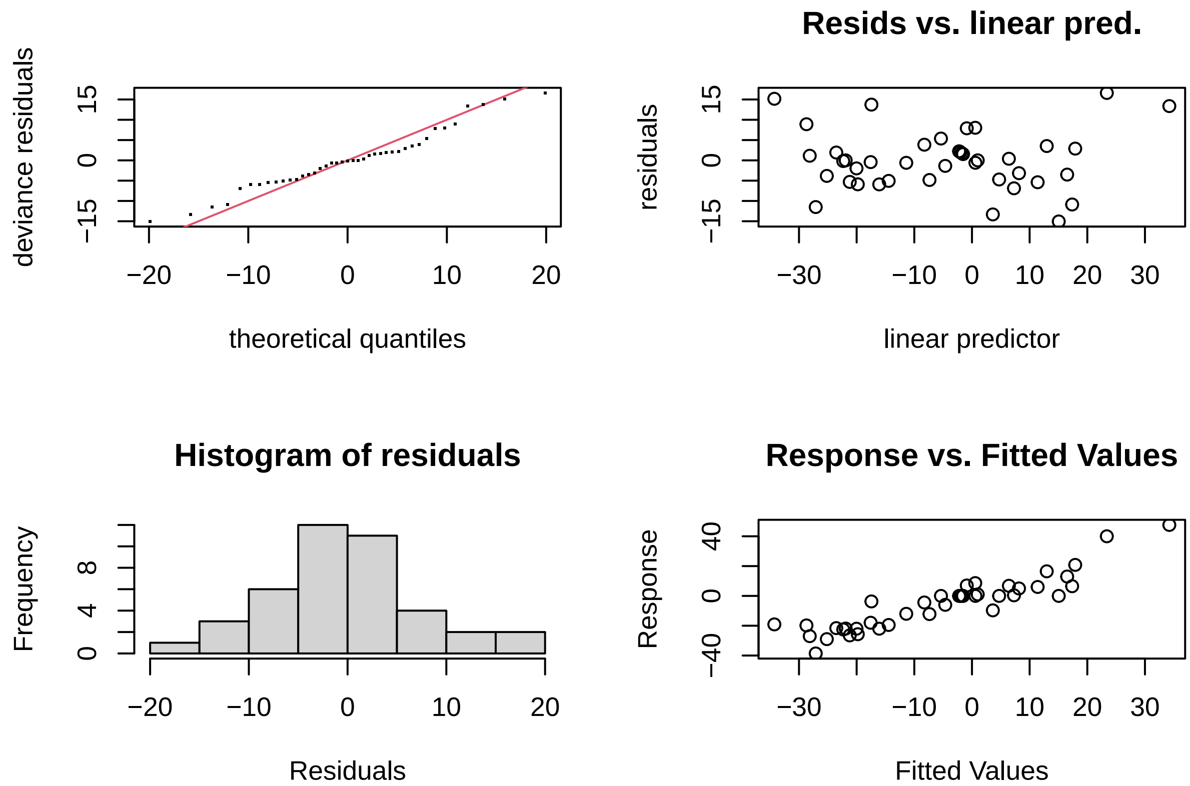
We appear to have less data here as % is reported more, but we can calculate total ha by multiplying % by total area.

```r
All_data <- Data_wF %>%
  mutate(Farea_km2 = ifelse(is.na(Farea_km2)==T,Area_km2*Perc_Farea_pre/100,Farea_km2))

Forest_model_area <- gam(DeltaQf_perc ~ DeltaF_perc +
                    log10(Area_km2) +
                    Dryness +
                    Farea_km2 +
                    Precip_data_type +  Assessment_technique +
                    Forest_type +
                    Hydrological_regime
                    , data = All_data)
summary(Forest_model_area)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## DeltaQf_perc ~ DeltaF_perc + log10(Area_km2) + Dryness + Farea_km2 +
##     Precip_data_type + Assessment_technique + Forest_type + Hydrological_regime
##
## Parametric coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -3.021e+01  1.641e+01  -1.841 0.076621 .
## DeltaF_perc                  -3.718e-01  9.294e-02  -4.000 0.000442 ***
## log10(Area_km2)               6.084e+00  3.346e+00   1.818 0.080140 .
## Dryness                       3.255e+00  3.156e+00   1.031 0.311481
## Farea_km2                    -6.197e-05  3.922e-05  -1.580 0.125688
## Precip_data_typeOB           -1.740e+01  6.224e+00  -2.795 0.009431 **
## Precip_data_typeSG           -8.667e+00  6.680e+00  -1.297 0.205467
## Assessment_techniqueEA, HM  4.465e-01  9.411e+00   0.047 0.962513
## Assessment_techniqueHM       1.436e+01  5.973e+00   2.404 0.023345 *
## Assessment_techniqueQPW     -6.118e-01  9.487e+00  -0.064 0.949055
## Assessment_techniqueSH       1.741e+01  7.243e+00   2.404 0.023362 *
## Forest_typeCF                -8.111e+00  5.993e+00  -1.353 0.187195
## Forest_typeMF                -4.063e-01  4.433e+00  -0.092 0.927641
## Hydrological_regimeSD         2.346e+01  6.605e+00   3.552 0.001428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## R-sq.(adj) =  0.756   Deviance explained = 83.5%
## GCV = 118.56  Scale est. = 78.075    n = 41
```

```
gam.check(Forest_model_area)
```




```
##
## Method: GCV   Optimizer: magic
## Model required no smoothing parameter selectionModel rank =  14 / 14
```

```
#plot(Forest_model_all)
```

Again, not significant and the model is well-behaved. Note that the Hydrological modelling assessment technique (HM) is again highly significant.

3. Conclusion

There is no effect of forest area in the relationship between change in flow and percent change in forest cover.