Review of Statistical approaches for identification of low flow drivers: temporal aspects by Anne Fangmann and Uwe Haberlandt

Martin Hanel, 22. 11. 2018

The authors addressed carefully most comments from the review. I still have few points that should be addressed beforee publication of the manuscript (see below).

Major comments

Extrapolation of statistical models

The predictive skill of statistical models under conditions deviating strongly from those the models were fitted to is questionable. It is evident in the case of modelling hydrological system, where e.g. the relation between precipitation and minimum flow would be very different depending on the presence of snow. In this perspective, the process-based models still have an advantage of following physical laws, which is not guaranteed in statistical models. I lack at least some discussion on this in the paper.

One may even think of some validation of considered statistical concepts in climate change affected conditions, which may be out of scope of the paper but on the other hand could bring interesting information on the extrapolation skill of statistical models. This could be done either by comparing the results of climate change impact assessment from statistical models to those based on hydrological model. Alternatively, one may perform a simulation study using hydrological model for generating present and future runoff, fitting the statistical model for present runoff and validate for future runoff.

Resampling procedure

I still have problems with the bootstrap procedure used by authors for detection of the stationarity of the regression. If I now understand correctly, authors create 1000 resamples of original data and for a tested combination of predictant and predictors they fit linear model. Then the fraction of significant regressions (from 1000 fitted models) is calculated and if the fraction is lower than 95 % they call the relationship non-stationary.

If I got it right the procedure can be demonstrated by following R code:

- generate variable `x`, 30 years and variable `y` that depend on `x`

```
x = rnorm(30)
y = x + rnorm(30, 0, 2)
```

- create a function to assess significance of the linear model `lm(y~x)` allowing for resampling `x` and `y` by an index `i`

```
sig = function(i){
  s = summary(lm(y[i]~x[i]))$fstatistic # get the F-statistics of the regression
  1 - pf(s[1], s[2], s[3])  # calculate p-value
}
```

- resample and calculate the fraction of significant regressions

```
e = replicate(1000, sig(sample(1:30, 30, replace = TRUE) ))
length(which(e <= 0.05))/1000
```

```
## [1] 0.911
```

If you run this code several times you can observe that the result strongly depends on the strength of the original relation - e.g. if you decrease the standard deviation of the random error term in `y` (e.g. `y = x + rnorm(30, 0, 1)`) then you get more significant results.

Note, that it is expected that when you estimate slopes on resamples you will get some spread. Under certain conditions this should match the estimate of the confidence interval for slope of the original regression and in the same way also the fraction of significant regressions should be related to the significance of the original regression.

This can be demonstrated by

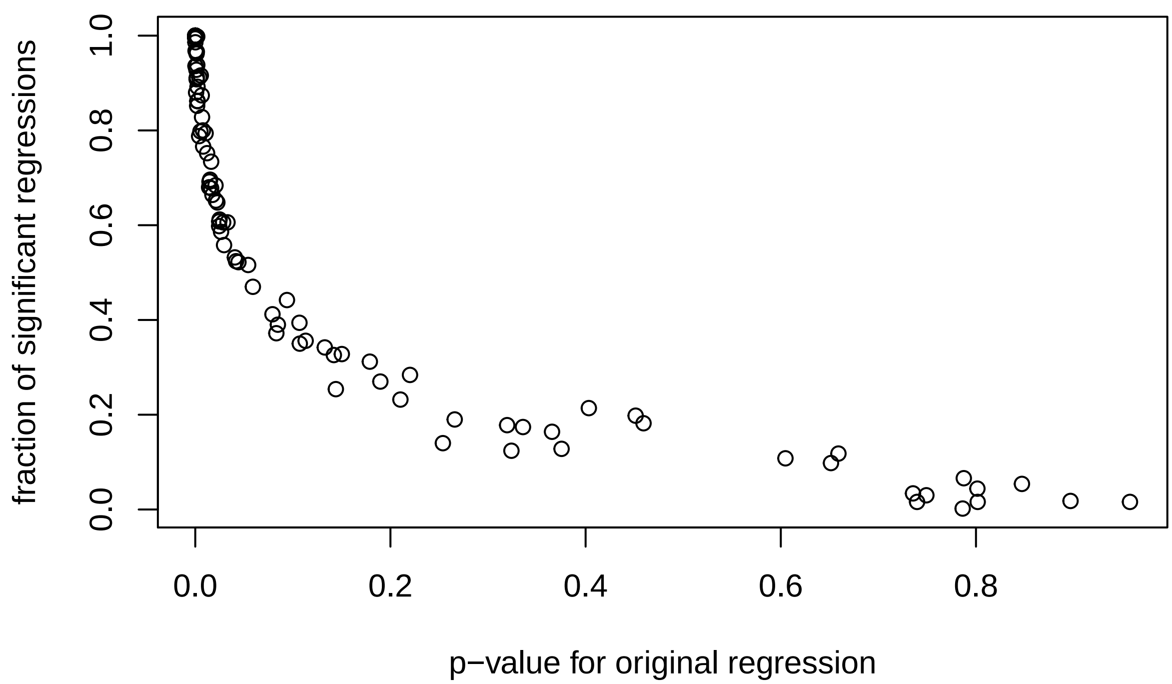
```
E = c()
O = c()

for (j in 1:100){

  x = rnorm(30)
  y = x + rnorm(30, 0, runif(1, 0.2, 5))
  O[j] = sig(1:30)
  e = replicate(500, sig(sample(1:30, 30, replace = TRUE) ))
  E[j] = length(which(e<=0.05))/length(e)

}

plot(O, E, xlab = 'p-value for original regression', ylab = 'fraction of significant regressions')
```



where we vary the standard deviation of the error term and collect the p-values for the original regression in `O` and the bootstraped fraction of significant regressions in `E`. It can be seen that in this case the fraction is strongly dependent on the original p-value. This behaviour can be different for more complex cases. But since the function generating `y` is linear and stationary, it shows, that the procedure cannot be used for detection of non-stationarity of the regression.

One may wonder, what exactly was intended to be tested here. How would the non-stationarity of the regression appear in data - by non-linearity? insignificance?

The part refering to the resampling procedure should be removed from the manuscript and the results section should be modified accordingly. I believe, that the main points of the paper are not affected by this.

Minor comments

- the 5th point from the conclusions on the differences between calibration and validation periods applies also to statistical models, while as it reads it seems that only to hydrological models