Model selection

Jake Cavaiani

1/28/2022

The purpose of this script is to model HI/FI against fixed and random effects and perform model selection techniques to see which model is best fit

Input: 2019\_2020 HI for each catchment that is will be from Output\_from\_analysis->07\_Models Step 1) import HI.dat file which is HI for individual storms in 2018-2021 across DoD sites Step 2) standardize and center the data Step 3) add fixed (storm total, intensity, week precip, month precip, 3month precip and doy) and random effects (catchments) Output: linear model plot

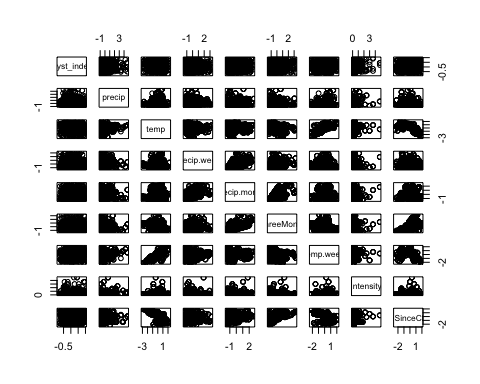
# Read in data #   
antecedent\_HI\_FI\_1\_0 <- read\_csv("~/Documents/Storms/Output\_from\_analysis/06\_HI\_fire\_permafrost\_script/antecedent\_HI\_FI\_1.0.csv")

## Warning: Missing column names filled in: 'X1' [1]

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## .default = col\_double(),  
## site.ID = col\_character(),  
## storm.ID = col\_character(),  
## response\_var = col\_character(),  
## burn = col\_character(),  
## pf = col\_character(),  
## date = col\_date(format = "")  
## )  
## ℹ Use `spec()` for the full column specifications.

is.na(antecedent\_HI\_FI\_1\_0) <- sapply(antecedent\_HI\_FI\_1\_0, is.na)  
# scaling data #   
#scaling #   
antecedent\_HI\_FI\_1\_0[c(20:25,32)] <- lapply(antecedent\_HI\_FI\_1\_0[c(20:25,32)], function(x) c(scale(x)))

## Assessing model assumptions

 Scatterplot matrix: It looks like precipitation week and precipitation month are correlated with each other

## Hyst\_index precip temp precip.week  
## Hyst\_index 1.000000000 -0.008387285 -0.0617517336 -0.0534602397  
## precip -0.008387285 1.000000000 0.0858749176 0.0954988234  
## temp -0.061751734 0.085874918 1.0000000000 0.0005875696  
## precip.week -0.053460240 0.095498823 0.0005875696 1.0000000000  
## precip.month -0.005587977 0.058721262 0.0933085239 0.5452518877  
## ThreeMonth 0.046766467 -0.085174810 -0.2631579693 0.2192151598  
## temp.week -0.023375851 0.172753379 0.8063858435 0.0650767967  
## Intensity NaN NaN NaN NaN  
## TimeSinceChena 0.086005838 -0.139963395 -0.7452130851 -0.0785204329  
## precip.month ThreeMonth temp.week Intensity TimeSinceChena  
## Hyst\_index -0.005587977 0.04676647 -0.02337585 NaN 0.08600584  
## precip 0.058721262 -0.08517481 0.17275338 NaN -0.13996339  
## temp 0.093308524 -0.26315797 0.80638584 NaN -0.74521309  
## precip.week 0.545251888 0.21921516 0.06507680 NaN -0.07852043  
## precip.month 1.000000000 0.60931531 0.09422631 NaN 0.13553049  
## ThreeMonth 0.609315307 1.00000000 -0.19328275 NaN 0.59988180  
## temp.week 0.094226308 -0.19328275 1.00000000 NaN -0.61507242  
## Intensity NaN NaN NaN 1 NaN  
## TimeSinceChena 0.135530486 0.59988180 -0.61507242 NaN 1.00000000

Correlation values >0.5 are 1) Temperature during the storm and temperature week leading up to a storm are heavily correlated (0.806) and since temperature during the storm is slightly more correlated with HI I will keep temperature during the storm in my model 2) Precipitation week leading up and Precipitation month = 0.545 and since precip week is more correlated with HI I will keep precip week in the model 3) Precip month is correlated with 3 month and since precip month is taken out I will also not include 3month

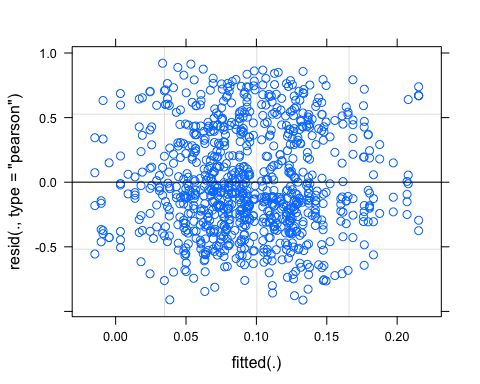
First model selection will include HI ~ precip + temp + precip.week + TimeSinceChena

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Hyst\_index ~ precip + temp + precip.week + TimeSinceChena + (1 |   
## site.ID)  
## Data: antecedent\_HI\_FI\_1\_0  
##   
## REML criterion at convergence: 969.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.2466 -0.7253 -0.1919 0.8336 2.2663   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## site.ID (Intercept) 0.00122 0.03493   
## Residual 0.16504 0.40625   
## Number of obs: 905, groups: site.ID, 6  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 0.0973749 0.0197698 4.925  
## precip 0.0004196 0.0137768 0.030  
## temp -0.0005279 0.0191250 -0.028  
## precip.week -0.0199353 0.0136668 -1.459  
## TimeSinceChena 0.0287052 0.0193336 1.485  
##   
## Correlation of Fixed Effects:  
## (Intr) precip temp prcp.w  
## precip -0.004   
## temp 0.005 -0.017   
## precip.week -0.001 -0.092 0.011   
## TimeSincChn -0.001 0.072 0.696 0.037

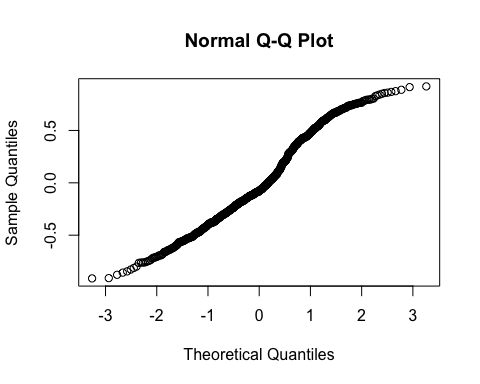
1. Ideally the median would be 0 and the quartiles would be similar on each side of 0 and the max and min would be similar as well.

2)Site taking up a small chunk of the total variance as a random effect

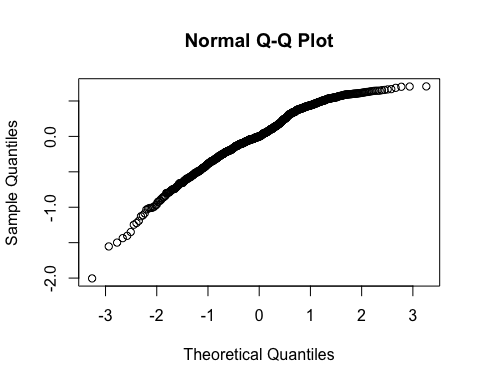
1. t values for each predictor reveal timesincechena would be most significant

 1) Ideally the residual plot will show no fitted pattern. The red line should be horizontal at 0.

This residual are spread equally along the ranges of predictors. It is good to see a horizontal line with equal spread points

 QQ plot of residuals can be used to visually check the normality assumptions… should be a straight line

Let me try log transforming the QQ-plot

 This does not improve the linearity of this reall so I am going to stick with the untransformed data

## Learn more about sjPlot with 'browseVignettes("sjPlot")'.

##   
## Attaching package: 'sjlabelled'

## The following object is masked from 'package:forcats':  
##   
## as\_factor

## The following object is masked from 'package:dplyr':  
##   
## as\_label

## The following object is masked from 'package:ggplot2':  
##   
## as\_label

##   
## Attaching package: 'sjmisc'

## The following object is masked from 'package:purrr':  
##   
## is\_empty

## The following object is masked from 'package:tidyr':  
##   
## replace\_na

## The following object is masked from 'package:tibble':  
##   
## add\_case

## Argument 'df\_method' is deprecated. Please use 'ci\_method' instead.

