

Dataframe used for lmm and prediction

Data preprocessing (same as EDA)

```
origin_df <- read.csv("hurrican703.csv")

hurricane_df <- origin_df %>%
  mutate(
    Month = factor(Month, levels = month.name[-c(2:3)]), # April-January (January ref, may choose another)
    Nature = as.factor(Nature), # TS,ET,DS,SS,NR (DS ref, may choose another)
    # note: one hurricane can have multiple natures throughout its life
    time = gsub("[()]", "", time),
    time = paste0(ifelse(substr(time, 1, 2) > 23, "19", "20"), time),
    time = as.POSIXct(time, format = "%Y-%m-%d %H:%M:%S"),
    hour = substr(time, 12, 19)
  ) %>%
  # remove data not at six-hour time intervals. (613 observations)
  filter(hour %in% c("00:00:00", "06:00:00", "12:00:00", "18:00:00")) %>%
  dplyr::select(-hour)

# remove hurricanes that has only 2 (<3) observations (change the threshold if you wish)
few_id <- hurricane_df %>%
  group_by(ID) %>%
  summarize(obs = n()) %>%
  filter(obs < 3) %>%
  . $ID
hurricane_df <- hurricane_df %>% filter(!(ID %in% few_id)) # remove 3 hurricanes

# manually correct hurricanes that have same names but are actually different
hurricane_df <-
  hurricane_df %>%
  mutate(
    # 2 hurricanes with the name ALICE.1954
    ID = ifelse(ID == "ALICE.1954" & Month == "June", "ALICE.1954(1)", ID),
    ID = ifelse(ID == "ALICE.1954", "ALICE.1954(2)", ID),
    # 4 hurricanes with the name SUBTROP:UNNAMED.1974
    ID = ifelse(ID == "SUBTROP:UNNAMED.1974" & Month == "June", "SUBTROP:UNNAMED.1974(1)", ID),
    ID = ifelse(ID == "SUBTROP:UNNAMED.1974" & Month == "July", "SUBTROP:UNNAMED.1974(2)", ID),
    ID = ifelse(ID == "SUBTROP:UNNAMED.1974" & Month == "August", "SUBTROP:UNNAMED.1974(3)", ID),
    ID = ifelse(ID == "SUBTROP:UNNAMED.1974", "SUBTROP:UNNAMED.1974(4)", ID),
    # 2 hurricanes with the name SUBTROP:UNNAMED.1976
    ID = ifelse(ID == "SUBTROP:UNNAMED.1976" & Month == "May", "SUBTROP:UNNAMED.1976(1)", ID),
    ID = ifelse(ID == "SUBTROP:UNNAMED.1976", "SUBTROP:UNNAMED.1976(2)", ID)
  )

df <- hurricane_df %>%
  group_by(ID) %>%
  mutate(
```

```

Season = first(Season),
Month = first(Month),
Nature = first(Nature),
wind_pre = lag(Wind.kt),
lat_diff = lag(Latitude) - lag(Latitude, 2),
long_diff = lag(Longitude) - lag(Longitude, 2),
wind_diff = lag(Wind.kt) - lag(Wind.kt, 2),
) %>%
drop_na %>%
dplyr::select(ID, Wind.kt, Season, Month, Nature, wind_pre, lat_diff, long_diff, wind_diff)

lmm <- lmer(Wind.kt ~ Season + Month + Nature + wind_pre + lat_diff + long_diff + wind_diff +
            (1 + wind_pre + lat_diff + long_diff + wind_diff | ID),
            data = df)

## boundary (singular) fit: see help('isSingular')

summary(lmm)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Wind.kt ~ Season + Month + Nature + wind_pre + lat_diff + long_diff +
##          wind_diff + (1 + wind_pre + lat_diff + long_diff + wind_diff | ID)
## Data: df
##
## REML criterion at convergence: 126058.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.7128  -0.3834  -0.0441   0.4696  10.0294
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   ID      (Intercept)  0.000000  0.00000
##           wind_pre     0.000193  0.01389   NaN
##           lat_diff     0.088349  0.29724   NaN -0.91
##           long_diff    0.072451  0.26917   NaN -0.87  0.63
##           wind_diff    0.016787  0.12957   NaN  0.31 -0.40  0.10
## Residual                28.370286  5.32638
## Number of obs: 20283, groups: ID, 704
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  23.927855   5.327205   4.492
## Season      -0.011179   0.002550  -4.385
## MonthApril    0.940675   1.719999   0.547
## MonthMay      0.523716   1.544654   0.339
## MonthJune     0.560594   1.525785   0.367
## MonthJuly     0.840846   1.518658   0.554
## MonthAugust   1.082921   1.513325   0.716
## MonthSeptember 1.268582   1.513119   0.838
## MonthOctober  1.190459   1.514973   0.786
## MonthNovember 1.242881   1.525055   0.815
## MonthDecember 0.682839   1.583124   0.431
## NatureET      0.068206   0.338901   0.201
## NatureNR      0.134254   0.597706   0.225

```

```

## NatureSS      0.060741  0.236561  0.257
## NatureTS      0.021720  0.177322  0.122
## wind_pre      0.941290  0.001890 498.090
## lat_diff      0.006555  0.065402  0.100
## long_diff     -0.243833  0.034082 -7.154
## wind_diff      0.468805  0.008776 53.417

##
## Correlation matrix not shown by default, as p = 19 > 12.
## Use print(x, correlation=TRUE) or
##      vcov(x)          if you need it

## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

```

```
fixed.effects(lmm)
```

##	(Intercept)	Season	MonthApril	MonthMay	MonthJune
##	23.927854961	-0.011178824	0.940675124	0.523716172	0.560593889
##	MonthJuly	MonthAugust	MonthSeptember	MonthOctober	MonthNovember
##	0.840845512	1.082920930	1.268581778	1.190458537	1.242880726
##	MonthDecember	NatureET	NatureNR	NatureSS	NatureTS
##	0.682838839	0.068205879	0.134254057	0.060741369	0.021719929
##	wind_pre	lat_diff	long_diff	wind_diff	
##	0.941289644	0.006555392	-0.243833138	0.468804969	