NOAA Buoy Data

December 2017

Introduction

The National Oceanic and Atmospheric Administration (NOAA) is the American federal agency in charge of collecting information and making decisions related to the oceans and the atmosphere. Throughout North America, they supply weather stations which are located both along the coast as well as in the middle of the ocean (on buoys). Among other variables, the weather stations collect information on wind, humidity, temperature, visibility, and atmospheric pressure. The data is all publicly available on NOAA's website, http://www.ndbc.noaa.gov/.

Data information & loading data

All the buoys are listed at http://www.ndbc.noaa.gov/to_station.shtml. The Santa Monica buoy information is at http://www.ndbc.noaa.gov/station_page.php?station=46025. The historical data is given at http://www.ndbc.noaa.gov/station_history.php?station=46025.

Always a good idea to look at the data! One thing to notice is that there are some variables coded as 99/999/9999. From user experience, we surmise that those values should be NA. Additionally, if we want to consider only the 2014 data, we should remove any previous data.

```
summary(buoy_data)
```

```
##       #YY            MM                  DD                  hh
##  Min.   :2013   Length:8611        Length:8611        Length:8611
##  1st Qu.:2014   Class :character   Class :character   Class :character
##  Median :2014   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2014
##  3rd Qu.:2014
##  Max.   :2014
##        mm          WDIR            WSPD             GST
##  Min.   :50   Min.   :  1.0   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:50   1st Qu.:165.0   1st Qu.: 2.000   1st Qu.: 2.800
##  Median :50   Median :266.0   Median : 3.200   Median : 4.100
##  Mean   :50   Mean   :231.3   Mean   : 3.589   Mean   : 4.545
##  3rd Qu.:50   3rd Qu.:298.0   3rd Qu.: 4.700   3rd Qu.: 5.800
##  Max.   :50   Max.   :360.0   Max.   :14.700   Max.   :18.100
##       WVHT             DPD             APD              MWD
##  Min.   : 0.360   Min.   : 2.74   Min.   : 3.620   Min.   :  1.0
##  1st Qu.: 0.810   1st Qu.:10.00   1st Qu.: 5.420   1st Qu.:204.0
##  Median : 0.990   Median :12.90   Median : 6.090   Median :253.0
##  Mean   : 1.273   Mean   :12.25   Mean   : 6.612   Mean   :236.5
##  3rd Qu.: 1.280   3rd Qu.:14.81   3rd Qu.: 7.220   3rd Qu.:268.0
##  Max.   :99.000   Max.   :99.00   Max.   :99.000   Max.   :999.0
##       PRES           ATMP            WTMP            DEWP
##  Min.   :1003   Min.   :10.10   Min.   :12.90   Min.   :-8.70
##  1st Qu.:1012   1st Qu.:15.20   1st Qu.:15.80   1st Qu.:11.50
##  Median :1014   Median :17.30   Median :18.60   Median :13.50
##  Mean   :1016   Mean   :17.25   Mean   :18.52   Mean   :13.22
```

```
##  3rd Qu.:1017   3rd Qu.:19.30   3rd Qu.:21.00   3rd Qu.:15.70
##  Max.   :9999   Max.   :24.50   Max.   :24.50   Max.   :99.00
##       VIS          TIDE
##  Min.   :99   Min.   :99
##  1st Qu.:99   1st Qu.:99
##  Median :99   Median :99
##  Mean   :99   Mean   :99
##  3rd Qu.:99   3rd Qu.:99
##  Max.   :99   Max.   :99
```

```
buoy_data <- buoy_data %>%
  mutate(WVHT = ifelse(WVHT==99, NA, WVHT)) %>%
  mutate(DPD = ifelse(DPD==99, NA, DPD)) %>%
  mutate(APD = ifelse(APD==99, NA, APD)) %>%
  mutate(MWD = ifelse(MWD==999, NA, MWD)) %>%
  mutate(PRES = ifelse(PRES==9999, NA, PRES)) %>%
  mutate(DEWP = ifelse(DEWP==99, NA, DEWP)) %>%
  select(-VIS, -TIDE) %>% filter(`#YY`==2014)

dim(buoy_data)
```

```
## [1] 8610   16
```

```
summary(buoy_data)
```

```
##       #YY             MM                  DD                  hh
##  Min.   :2014   Length:8610        Length:8610        Length:8610
##  1st Qu.:2014   Class :character   Class :character   Class :character
##  Median :2014   Mode  :character   Mode  :character   Mode  :character
##  Mean   :2014
##  3rd Qu.:2014
##  Max.   :2014
##
##        mm          WDIR            WSPD             GST
##  Min.   :50   Min.   :  1.0   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:50   1st Qu.:165.0   1st Qu.: 2.000   1st Qu.: 2.800
##  Median :50   Median :266.0   Median : 3.200   Median : 4.100
##  Mean   :50   Mean   :231.3   Mean   : 3.589   Mean   : 4.545
##  3rd Qu.:50   3rd Qu.:298.0   3rd Qu.: 4.700   3rd Qu.: 5.800
##  Max.   :50   Max.   :360.0   Max.   :14.700   Max.   :18.100
##
##       WVHT             DPD              APD              MWD
##  Min.   :0.360   Min.   : 2.74   Min.   : 3.62   Min.   :  1.0
##  1st Qu.:0.810   1st Qu.:10.00   1st Qu.: 5.42   1st Qu.:204.0
##  Median :0.990   Median :12.90   Median : 6.09   Median :253.0
##  Mean   :1.091   Mean   :12.09   Mean   : 6.44   Mean   :235.1
##  3rd Qu.:1.280   3rd Qu.:14.81   3rd Qu.: 7.21   3rd Qu.:268.0
##  Max.   :4.800   Max.   :23.53   Max.   :13.52   Max.   :359.0
##  NA's   :16      NA's   :16      NA's   :16      NA's   :16
##       PRES           ATMP            WTMP             DEWP
##  Min.   :1003   Min.   :10.10   Min.   :12.90   Min.   :-8.7
##  1st Qu.:1012   1st Qu.:15.20   1st Qu.:15.80   1st Qu.:11.5
##  Median :1014   Median :17.30   Median :18.60   Median :13.5
##  Mean   :1015   Mean   :17.25   Mean   :18.52   Mean   :13.2
##  3rd Qu.:1017   3rd Qu.:19.30   3rd Qu.:21.00   3rd Qu.:15.7
```

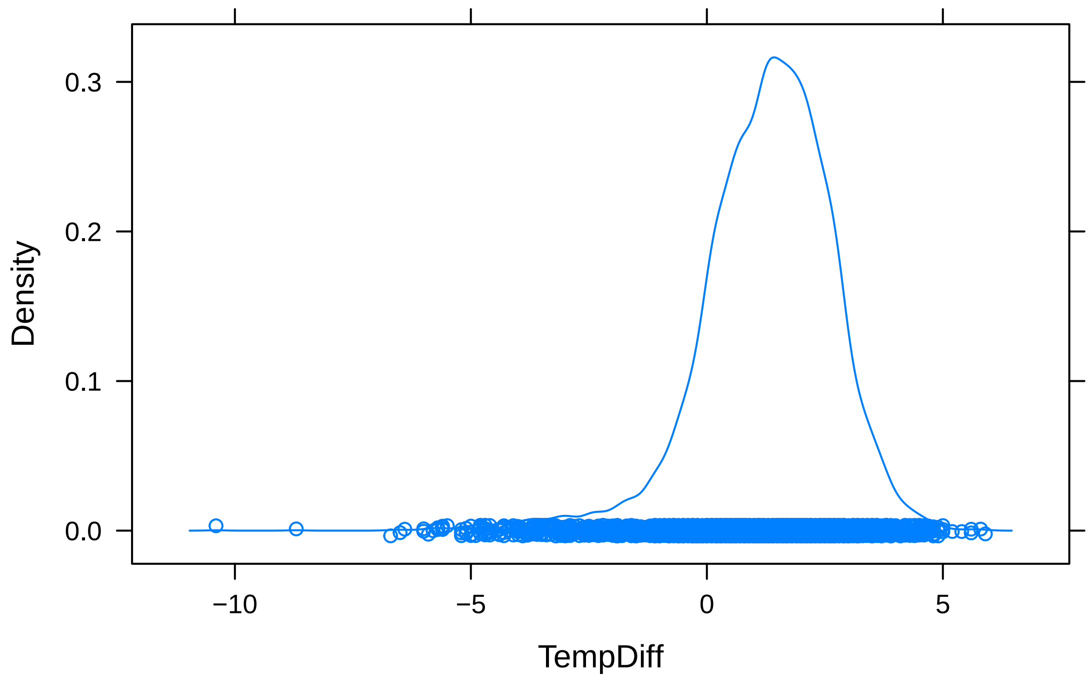

Figure 1: Density of the temperature difference between wind and air.

```
##  Max.   :1029   Max.   :24.50   Max.   :24.50   Max.   :21.0
##  NA's   :1                                      NA's   :2
```

Using dynamic data within a typical classroom

Although the data do not constitute a random sample, they are very likely to be quite representative with respect to the difference in wind and air temperature at that location over the year. We used a paired analysis (i.e., subtract the two variables and treat them as a single variable) to find a confidence interval for the true difference in temperature between wind and air. Also, we find a prediction interval for the difference in temperatures across individual measurements.

One might be interested in the difference between the wind temperature and the air temperature. Generally, the air temperature is cooler than the wind temperature, but confidence intervals and prediction intervals allow us to quantify the difference. Note that the data lend themselves nicely to ideas of paired observations acting as a univariate sample. As expected, a 95% confidence interval for the true difference in temperatures gives us a value of between 1.25 and 1.31 degrees. However, 95% of the individual observations have a difference in wind and air temperature between -1.5 degrees (air is warmer) and 4.06 degrees (wind is warmer).

```
buoy_data$TempDiff <- buoy_data$WTMP - buoy_data$ATMP
densityplot(~TempDiff, data=buoy_data)
```

```
tempdiff.mod <- lm(TempDiff ~ 1, data=buoy_data)
tempdiff.func <- makeFun(tempdiff.mod)
tempdiff.func()
```

```
##        1
## 1.276249
```

```
tempdiff.func( interval="prediction")
```

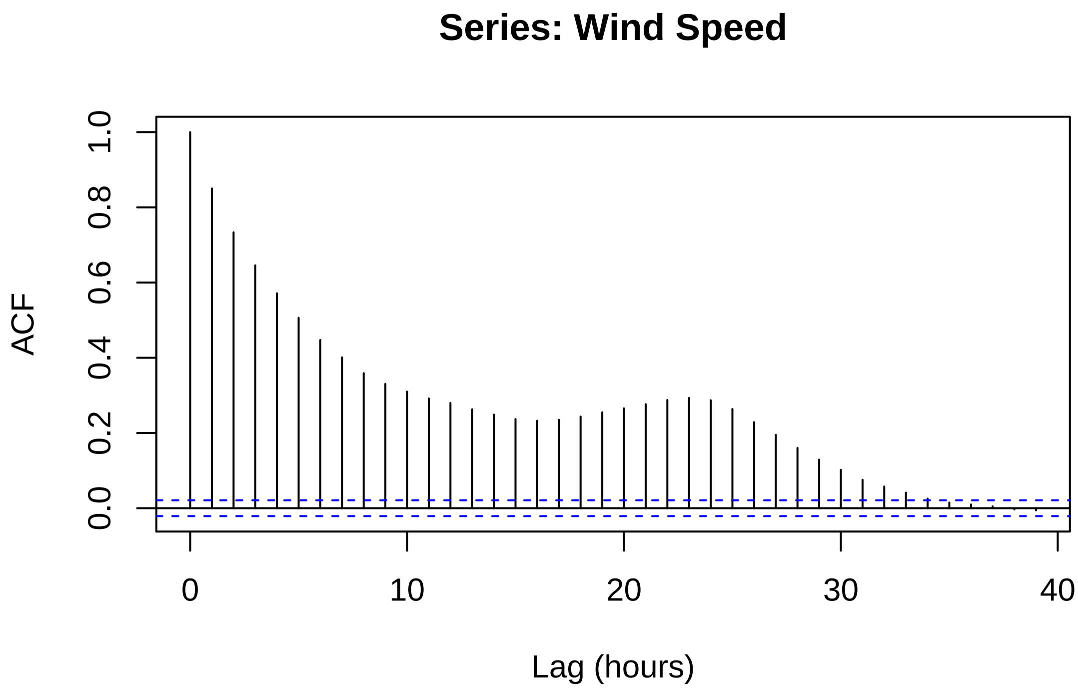

Figure 2: Autocorrelation function shows strong correlations at a few hours and at 24 hours.

```
##        fit      lwr      upr
## 1 1.276249 -1.50516 4.057657
```

```
tempdiff.func( interval="confidence")
```

```
##        fit      lwr      upr
## 1 1.276249 1.246275 1.306222
```

Thinking outside the box

Although a full analysis of the data would warrant multiple years of data (so as to understand yearly trends), we can estimate the spectral density of the time series using a smoothed periodogram (the data below represent measurements every hour for all of 2014). In the smoothed periodogram (see Figure 3) the x-axis is the frequency (one over the period) and y-axis represents the correlation (normalized) between the cosine wave at that frequency and the time series. We can see that wind speed has strong correlation at period 12 hours and period 24 hours. A more sophisticated analysis or longer project could include collecting data from multiple buoys, extended years, and/or additional information on storms https://www.ncdc.noaa.gov/stormevents/.

The data are nicely set up to think about analyses is the time domain. Indeed, looking at the autocorrelation function shows clear 24-hour trends for the wind speed variable.

```
acf(buoy_data$WSPD, main="Series: Wind Speed", xlab="Lag (hours)")
```

Although a full analysis of the data would warrant multiple years of data (so as to understand yearly trends), we can estimate the spectral density of the time series using a smoothed periodogram.

```
spec.pgram(buoy_data$WSPD, spans=c(50), xlab="Frequency = 1/period",
           main="Wind Speed, Smoothed Periodogram")
abline(v=c(1/12,1/24), lty=2)
```

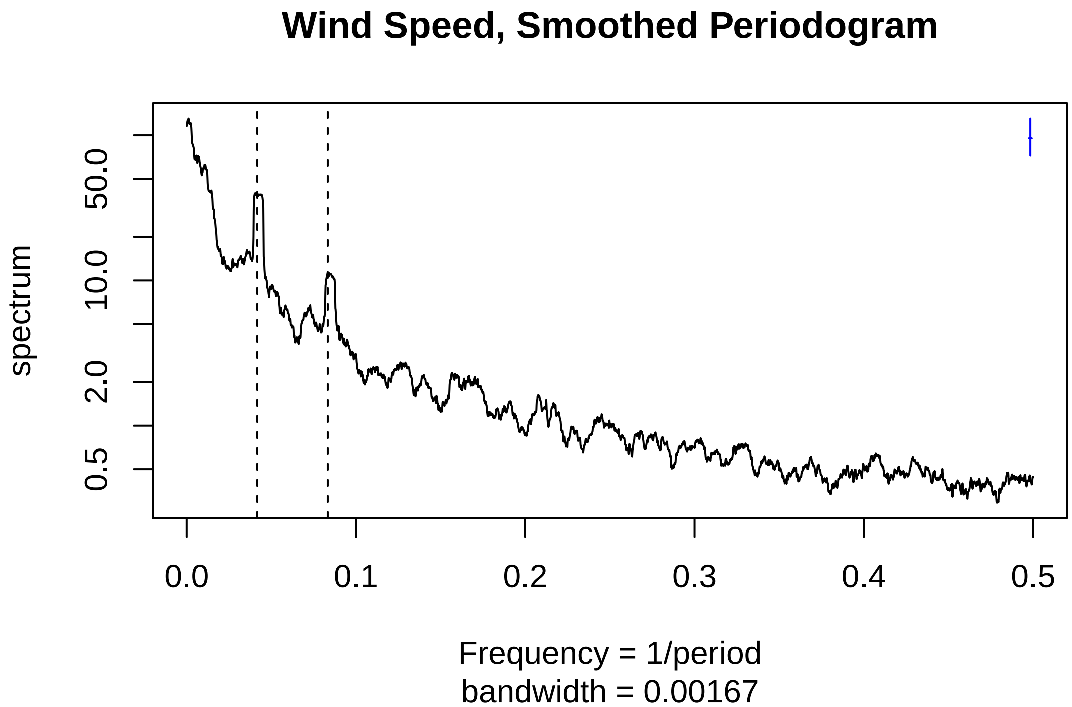

Figure 3: A smoothed periodogram of Wind Speed for buoy 46025 off the coast of Santa Monica. The dashed lines are at 1/12 and 1/24, representing 12-hour and 24-hour periodicity.

In the smoothed periodogram, the x-axis is the frequency (one over the period) and y-axis represents the correlation (normalized) between the cosine wave at that frequency and the time series. We can see that wind speed has strong correlation at period 12 hours and period 24 hours.

Additional ideas for analysis:

A more sophisticated analysis or longer project could include collecting data from multiple buoys, extended years, and/or additional information on storms https://www.ncdc.noaa.gov/stormevents/.