

SEB113 Collaborative Scientific Article (CSA)

Yun Kai Siow, 9598138

Sunday, June 12, 2016

Contents

1. Aim	1
2. Methods	2
3. Data	3
4. Analysis	4
5. Interpret	6
References	9

1. Aim

Question

What influence do meteorological measurements such as wind speed and wind direction have on the quality of air, particularly concentrations of PM2.5?

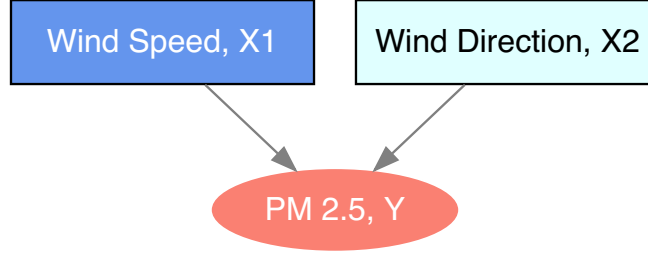


Figure 1: Global average across Ocean Health Index goals.

2. Methods

The scientific conceptual model

Diagram

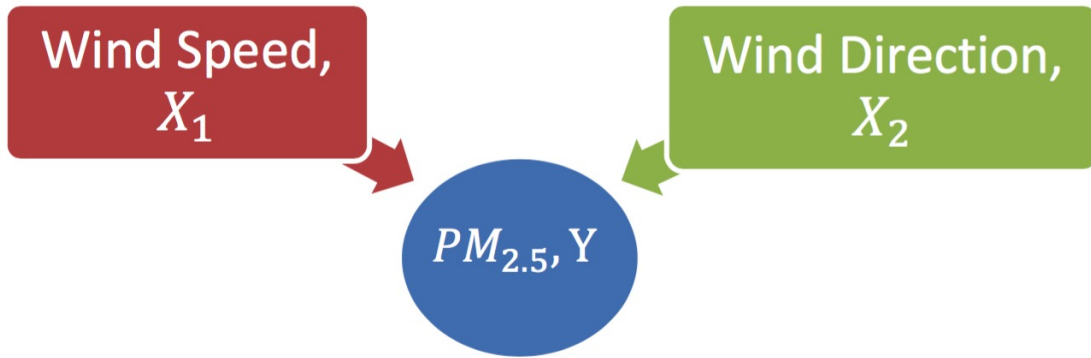


Figure 1: Visual conceptual model of how PM2.5 concentration in the air varies according to wind speed and wind direction.

As with other meteorological conditions, the explanatory variables, wind speed and wind direction are believed to have play an important role on direct or indirect correlation with the dispersion of air pollutant (e.g. PM2.5) concentration in the air ((???), (???)) (Dawson et al., 2007; Elminir, 2005).

For instance, if there is a forest fire happening at the south west of our current location, a gust of south western wind with the right speed will certainly bring the pollutant, therefore increase the pollutant concentration in the air. Vice versa, wind from other direction with certain speed could also carry away and disperse pollutants in the air.

There is no specified functional form from a scientific law to describe the influence of wind speed

and wind direction affect the concentration of PM2.5 in the air, so linear terms is used in this model.

The quatitative model

$$\log PM2.5_i = \sum_{j=1}^J \beta_j \cdot I(WD_i = j) + \sum_{k=1}^K \beta_k \cdot WS_i \cdot I(WD_i = k) + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Variables:

- $\log PM2.5_i$: i th observation of log PM2.5
- WS_i : Wind speed value for observation i
- WD_i : Wind direction value for observation i
- J & K : Total number of wind direction, 8 (e.g. N, NE, E, SE, S, SW, W, NW)
- $I(\cdot)$: an indicator variable that tell us whether or not the statement inside (that Wind Direction has a particular value) is true.

Formulate a hypothesis:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

3. Data

Preparation

Dataset

Data dictionary

Data dictionary - variables

Table 1: Analysis of Deviance Table

Abbreviation	Variable	Description	Units
ws	Wind speed	blabla	blabla
wd	Wind Direction	blabla	blabla
wd.label	Content	blabla	blabla
pm2.5	Content	blabla	blabla

Data dictionary - experimental units measured

Variable	Description	Values	Population	Sampled
ws	Wind speed	blabla	blabla	blabla
wd	Wind Direction	blabla	blabla	blabla
wd.label	Content	blabla	blabla	blabla
pm2.5	Content	blabla	blabla	blabla

The data set comprises time series of: wind speed and direction; temperature; relative humidity; rainfall; barometric pressure; oxides of nitrogen; sulfur dioxide; PM10 / PM2.5 readings; and visibility-reducing particles. All updated hourly over the period from 1st January to 31st December 2015, recorded at Clinton, Gladstone Queensland (Latitude: -23.8701; Longitude: 151.2216).

The dataset is released under a Creative Commons Attribution 3.0 Australia (CC BY) licence.



Metadata

4. Analysis

Exploratory data analysis

Table 3: Title of the table

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Quantitative analysis

Estimate model parameters

```
head(air.quality.clinton)
```

```
##
## Have a look on first few rows of data
## =====
##      date      time month day_of_week pm2.5  ws   wd  wd.label log.pm2.5
## -----
## 1 2015-01-01 01:00:00 01:00   Jan      Thurs    3.400 2.600 58      NE      1.224
## 2 2015-01-01 02:00:00 02:00   Jan      Thurs    2.100  3    63      NE      0.742
## 3 2015-01-01 04:00:00 04:00   Jan      Thurs    1.200 1.500 82      E       0.182
## 4 2015-01-01 05:00:00 05:00   Jan      Thurs     6    1   128     SE      1.792
## 5 2015-01-01 06:00:00 06:00   Jan      Thurs     5   1.600 120     SE      1.609
## 6 2015-01-01 07:00:00 07:00   Jan      Thurs    4.700 2.500 96      E       1.548
## -----
```

Assess model fit

Model checking

5. Interpret

Model interpretation

Link back

Compare

Formulate a ANOVA hypothesis:

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

Interpretation of ANOVA for regression models

The p value from this hypothesis test therefore represents the probability of obtaining an F statistic at least as big as what was seen if the restricted model explained the same amount of variation as the full model.

Rejecting the null hypothesis leads us to conclude that the inclusion of the extra terms in the full model explains more variation than if we had not included these terms.

The sum of square of the errors (residual sum of squares) for model 5?

```
sum(residuals(lm.pm_wswd.forR2)^2)
```

```
## [1] 5232.343
```

There are “ residual degrees of freedom for the model 5?

The sum of square of the errors (residual sum of squares) for the model involving wind direction.

```
sum(residuals(lm.pm_wd_wswd.forR2)^2)

## [1] 5179.542

anova(lm.pm_wswd.forR2, lm.pm_wd_wswd.forR2)

## Analysis of Variance Table
##
## Model 1: log.pm2.5 ~ ws:wd.label
## Model 2: log.pm2.5 ~ wd.label + ws:wd.label
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1     7880 5232.3
## 2     7873 5179.5   7    52.801 11.465 1.459e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

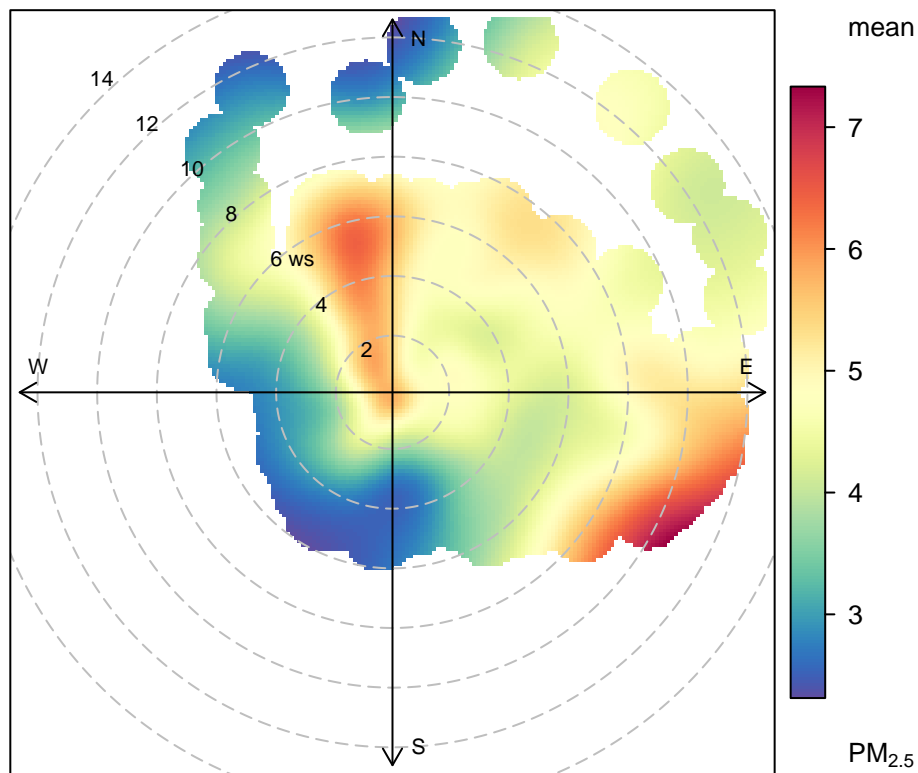


Figure 2: Global average across Ocean Health Index goals.

All analyses were conducted using the statistical software program, R (R Core Team, 2016).

D. Kahle and H. Wickham. (Kahle & Wickham, 2013): Spatial Visualization with ggplot2

Carslaw D and Ropkins K (2016). (Carslaw & Ropkins, 2012): Open-source tools for the analysis of air pollution data.

References

- Carslaw, D. C., & Ropkins, K. (2012). Openair — an r package for air quality data analysis. *Environmental Modelling & Software*, 27–28(0), 52–61. <http://doi.org/10.1016/j.envsoft.2011.09.008>
- Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1), 144–161. Retrieved from <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>