Xinjing Zhu
January 2012
NEAS Times Series Student Project

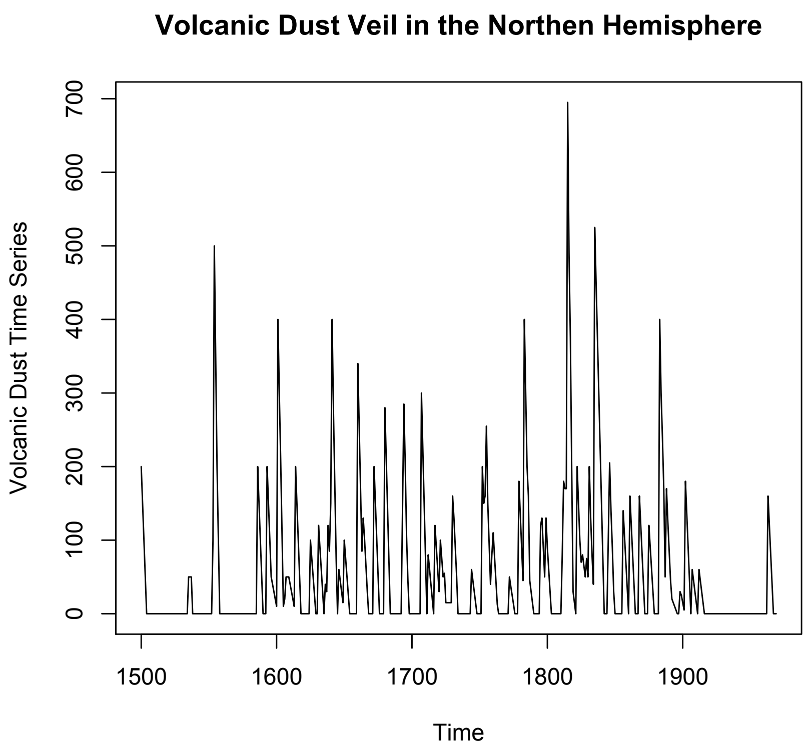
Introduction

For my student project for NEAS Times Series course, I looked at data on the volcanic dust veil index in the northern hemisphere from 1500 to 1969. This index is a measure of the impact from volcanic eruptions' release of dust into the environment. Statistical package R is used to support my analysis.

Data

The source of my data is the following website:
<http://robjhyndman.com/tsdldata/annual/dvi.dat>

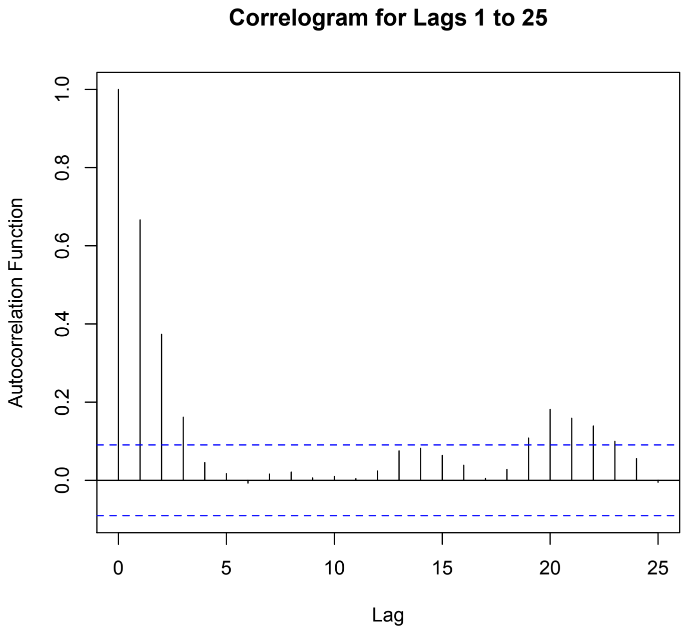
Following is a graph of the data used for this project:



Model Specification

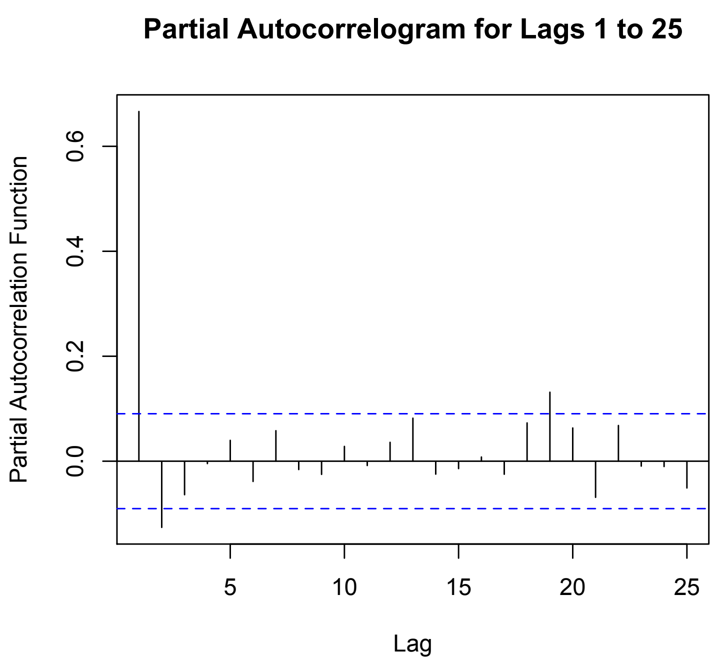
Since the data is on an annual basis, there is no seasonality. Furthermore, the time series appears to be stationary in mean and variance, as its level and fluctuation seems to be roughly constant over time. Therefore, we do not need to difference this time series in order to fit an ARIMA model. In this case, d=0.

The autocorrelation function plot for the original volcanic dust time series in next page:



We can see from the correlogram that the autocorrelations for lags 1, 2 and 3 exceeds the significance bounds. However, the autocorrelations decrease to 0 after lag 3 and remain around 0 and within significance bounds after lag 3 other than a few random fluctuations (lag 19 to 23).

In addition, the partial autocorrelation function plot for the original volcanic time series is shown in the next page.



We can see from the partial autocorrelation plot that the autocorrelation for lag 1 and 2 both exceed the significance bounds. However, it decreases to 0 after lag 2 and remain around 0 and within significance bounds after lag 2 other than one fluctuation at lag 19.

Model Fitting

Based on my analysis to this point, the following ARMA models are possible for the volcanic dust time series:

- An AR(2) model, since the partial autocorrelation decreases to 0 after lag 2.
- A MA(3) model, since the autocorrelation decreases to 0 after lag 3.

<u>Model 1: AR(2) model on volcanic dust data</u>

Following is the key regression output for an AR(2) model on the original time series:

```
Coefficients:
        ar1      ar2  intercept
     0.7533  -0.1268    57.5274
s.e. 0.0457   0.0458     8.5958
```

AR(2) Equation: $Y_t = 0.7533Y_{t-1} - 0.1268Y_{t-2} + 57.5274$
The stationarity conditions for an AR(2) model are met for this model:
$\Phi_1 + \Phi_2 < 1$ $(\Phi_1 + \Phi_2 = 0.7533-0.1268 = 0.6265)$
$\Phi_2 - \Phi_1 < 1$ $(\Phi_2 - \Phi_1 = -0.1268-0.7533 = -0.8801)$
$| \Phi_2 | < 1$ $(| \Phi_2 | = 0.1268)$

<u>Model 2: MA(3) model on volcanic dust data</u>

Following is the key regression output for an MA(3) model on the original time series:

```
Coefficients:
        ma1     ma2     ma3  intercept
     0.7438  0.4513  0.1916    57.4559
s.e. 0.0455  0.0502  0.0442     7.6534
```

MA(3) Equation: $Y_t = 57.4559 + e_t + 0.7438e_{t-1} +0.4513e_{t-2} + 0.1916e_{t-3}$

A MA(q) is stationary for every sequence of coefficients.

<u>Comparison of Models</u>

The AR(2) model has 2 parameters; the MA(3) model has 3 parameters. Therefore, using the principle of parsimony, the AR(2) model should be selected.

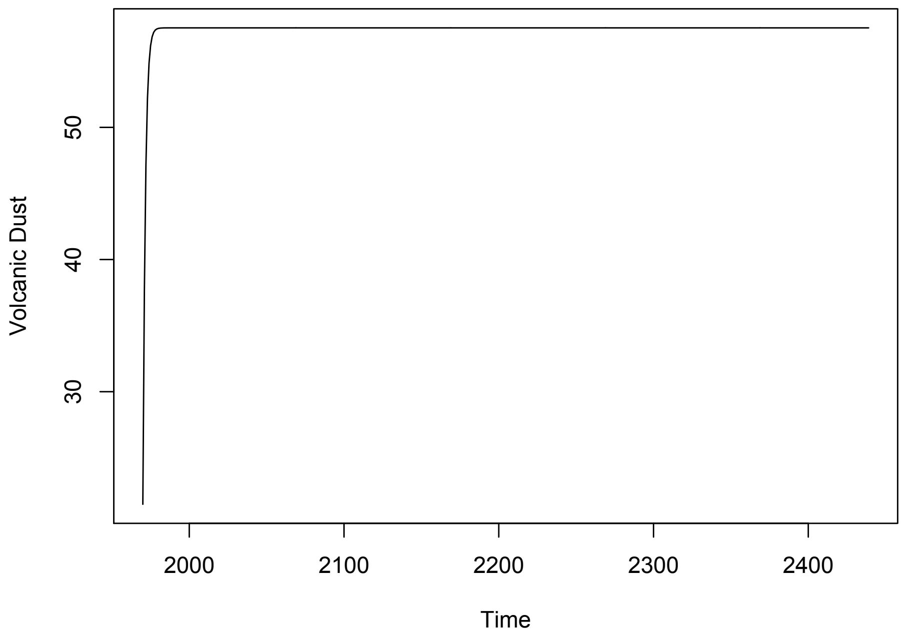
An AR (autoregressive) model is usually used to model a time series that shows long-term dependencies between successive observations. It also makes sense intuitively as volcanic dust in one year is expected to have an affect on those in later years, since the dust are unlikely to disappear quickly.

Therefore I will use AR(2) model based on the original volcanic dust times series for forecasting.

Forecasting

Since the original volcanic dust data includes 470 years from 1500 to 1969. Now I will use the AR(2) model to make predictions for year 1970 to 2439, another 470 years.

Forecast from AR(2) Model for Volcanic Dust in Next 470 Years



We can see that the predictions for the next 470 years differ drastically from the original data of 470 years. The forecast shows that the volcanic dust increases from 1970 to pre-2000 and then remains the same for the next 400 years or so.