# Time Series Analysis on London Mortality

A 618 Project

Egan McClave & Aijin Wang

April 23, 2019

# Table of contents

# Introduction

# Introduction

**Dataset**
Various time series from London from 2002 to 2006

- Ozone
- Relative Humidity
- Temperature
- Number of Deaths

**Research Question**
Investigation of the associations between environmental factors and human mortality

# Exploratory Data Analysis

# Overall Series

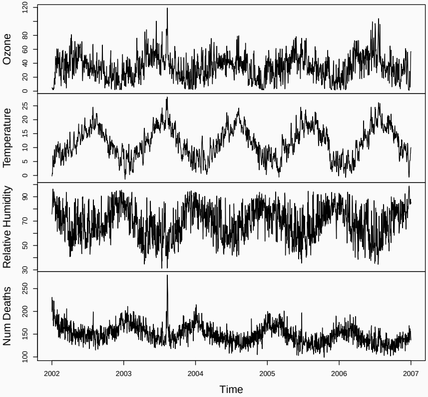



**Figure 1:** Individual Series for London (2002 - 2007)

what's that weird spike?

# Pairs Plot

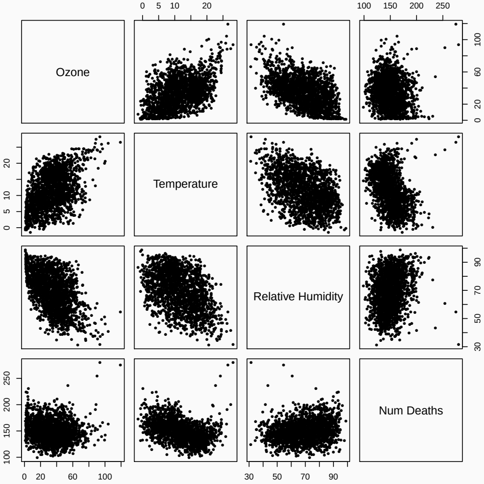



**Figure 2:** Pairs Plot for All Variables

# Decomposed Seasonality Components

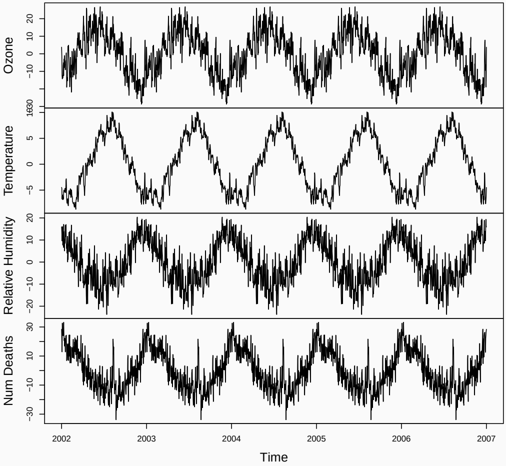



**Figure 3:** Decomposed Seasonality Components for London (2002-2007)

# ACF/CCF Plots

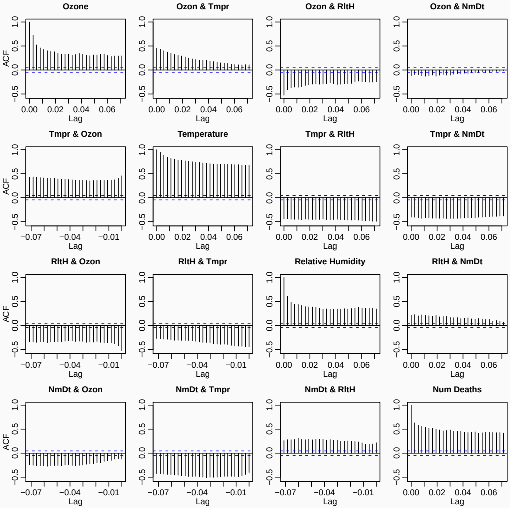



**Figure 4:** ACF/CCF Plots

# PACF/PCCF Plots

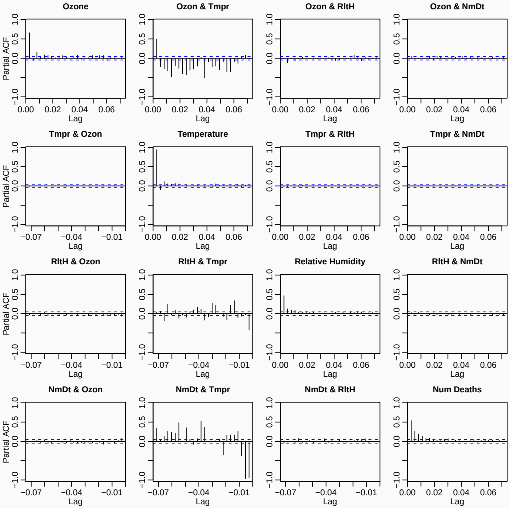



**Figure 5:** PACF/PCCF Plots

# Variables under Consideration

**Independent Variables**

- Time
- Mean Centered Temperature
- (Mean Centered Temperature)$^2$
- Ozone levels
- Relative Humidity

**Response Variable**

- Number of Deaths

# Model Analysis

# Overview

- Types of Models

# Overview

- Types of Models
  - Time series regression

# Overview

- Types of Models
  - Time series regression
  - VAR model

# Overview

- Types of Models
  - Time series regression
  - VAR model ← primary model for this presentation

# Overview

- Types of Models
  - Time series regression
  - VAR model
  - NNETAR model

# Overview

- Types of Models
    - Time series regression
    - VAR model
    - NNETAR model
- How do we evaluate them?

# Overview

- Types of Models
    - Time series regression
    - VAR model
    - NNETAR model
- How do we evaluate them?
    - Original data is 1826 observations long

# Overview

- Types of Models
    - Time series regression
    - VAR model
    - NNETAR model
- How do we evaluate them?
    - Original data is 1826 observations long
    - Training data is 1461 observations (2002 - 2005)

# Overview

- Types of Models
  - Time series regression
  - VAR model
  - NNETAR model
- How do we evaluate them?
  - Original data is 1826 observations long
  - Training data is 1461 observations (2002 - 2005)
  - Testing/validation data is 365 observations (all of 2006)

# Vector ARMA Model - Model Fitting

Parameter Selection:

- `VARselect()` $\rightarrow p = 4$
- intutition about series $\rightarrow$ `season = 365`

$\Rightarrow$ `fit <- VAR(# Deaths, p=4, season=365, type='none')`

# Vector ARMA Model - Evaluating Fit 1

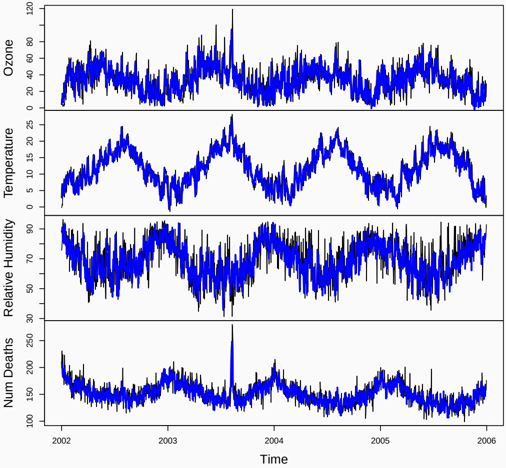



**Figure 6:** Fitted Values vs Original Series

# Vector AR Model - Evaluating Fit 2

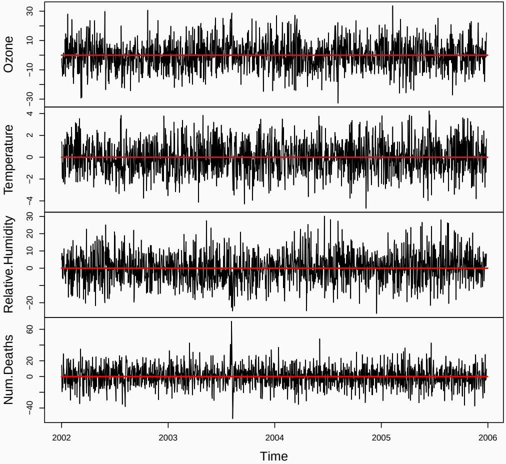



**Figure 7:** Residuals for Individual Series

# Vector AR Model - Model Forecasting 1

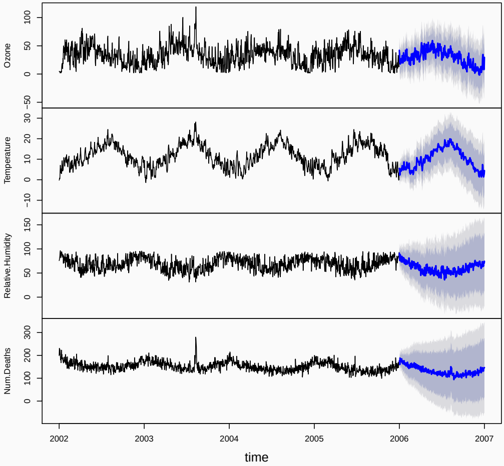



**Figure 8:** Forecasting Individual Series

# Vector AR Model - Model Forecasting 2

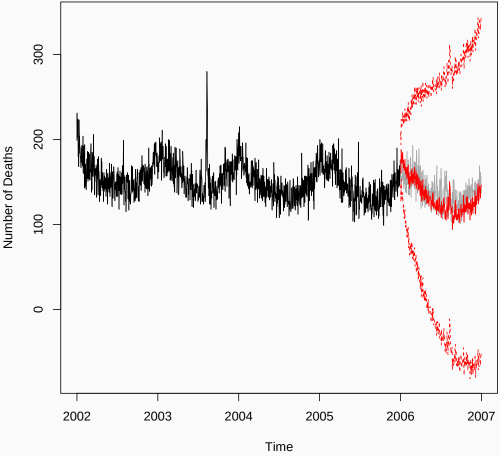



**Figure 9:** Forecasting Response Series

# Vector AR Model - Model Forecasting 3

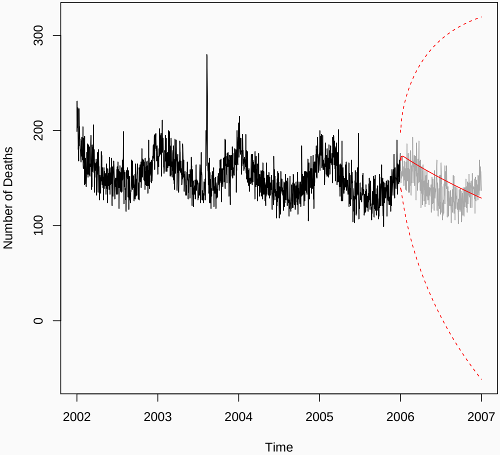



**Figure 10:** Forecasting Response Series (`season=NULL`)

# Vector AR Model - Model Inference

| | Num Deaths | | | |
|---|---|---|---|---|
| Coefficient | Estimate | Std. Error | t value | Pr(>\|t\|) |
| Ozone lag 1 | 0.045 | 0.040 | 1.122 | 0.262 |
| **Temperature lag 1** | 0.516 | 0.245 | 2.108 | 0.035 |
| Relative Humidity lag 1 | 0.048 | 0.042 | 1.152 | 0.250 |
| **Num Deaths lag 1** | 0.375 | 0.028 | 13.222 | 0.000 |
| Ozone lag 2 | 0.061 | 0.049 | 1.244 | 0.214 |
| **Temperature lag 2** | -0.345 | 0.340 | -1.015 | 0.310 |
| Relative Humidity lag 2 | 0.068 | 0.045 | 1.489 | 0.137 |
| Num Deaths lag 2 | 0.182 | 0.030 | 6.014 | 0.000 |
| Ozone lag 3 | -0.039 | 0.049 | -0.790 | 0.430 |
| **Temperature lag 3** | -0.791 | 0.338 | -2.340 | 0.019 |
| Relative Humidity lag 3 | -0.026 | 0.046 | -0.577 | 0.564 |
| Num Deaths lag 3 | 0.158 | 0.030 | 5.237 | 0.000 |
| Ozone lag 4 | 0.054 | 0.040 | 1.353 | 0.176 |
| **Temperature lag 4** | 0.482 | 0.246 | 1.958 | 0.050 |
| Relative Humidity lag 4 | 0.165 | 0.042 | 3.967 | 0.000 |
| Num Deaths lag 4 | 0.151 | 0.028 | 5.304 | 0.000 |

**Table 1:** VAR(4) Summary for Number of Deaths

# Summary

# Summary

- Temperature and Number of Deaths move in phase
- There is a non-linear relationship between Temperature and Number of Deaths
- Generalizability of the fitted model is good
  - Captures seasonality
  - Captures downward linear trend
  - Captures daily volatility

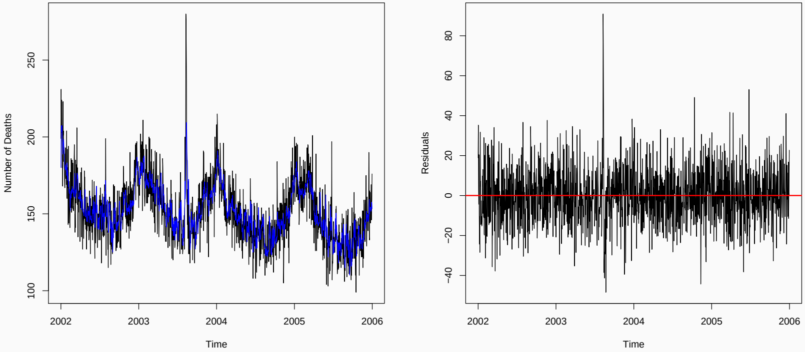
**Questions?**

# Back up Slides

# Time Series Regression Model - Model Fitting

```
fit <- forecast::auto.arima(Num Deaths, xreg=...,
seasonal=T)
```

# Time Series Regression Model - Evaluating Fit 1

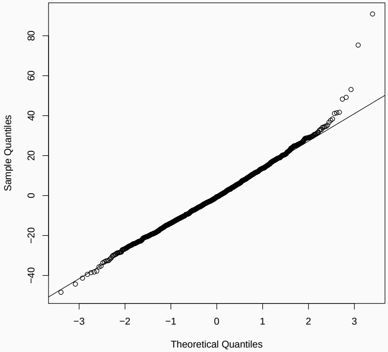



**(a)** Fitted Values and Original Series

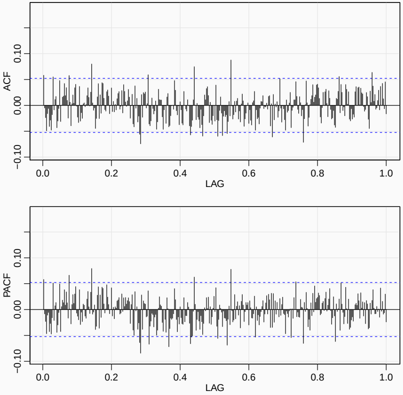
**(b)** Residuals over Time

**Figure 11:** Visualizing Model Fit

# Time Series Regression Model - Evaluating Fit 2




**(a)** QQ Plot of Residuals




**(b)** ACF/PACF of Residuals

**Figure 12:** Visualizing Model Residuals

# Time Series Regression Model - Model Forecasting

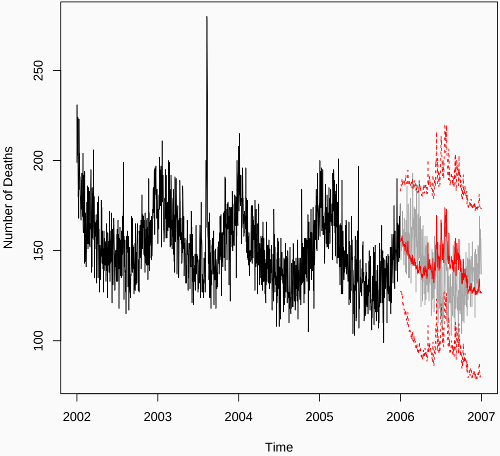
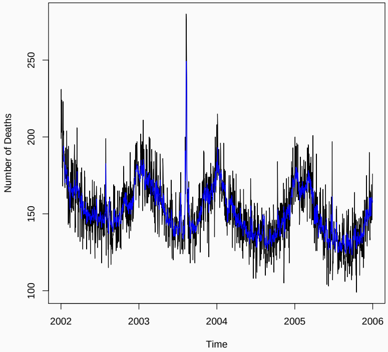



**Figure 13:** Visualizing Forecasted Series

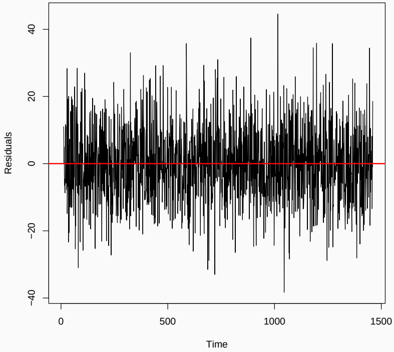
# Neural Network Time Series Model - Model Fitting

```
fit <- forecast::nnetar(Num Deaths, xreg=..., seasonal=T)
```

# Neural Network Time Series Model - Evaluating Fit 1



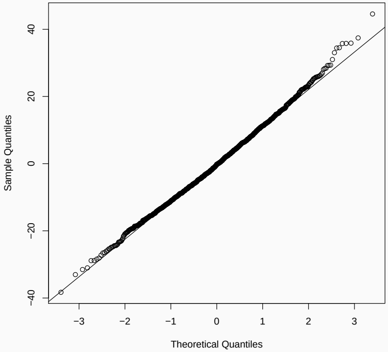

**(a)** Fitted Values and Original Series



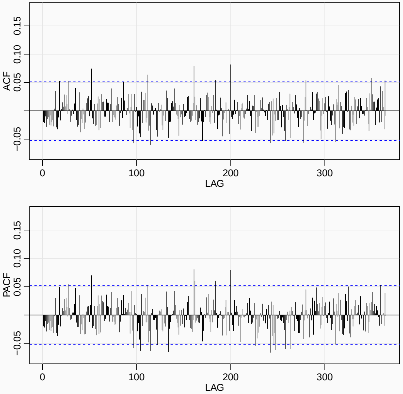

**(b)** Residuals over Time

**Figure 14:** Visualizing Model Fit

# Neural Network Time Series Model - Evaluating Fit 2




**(a)** QQ Plot of Residuals




**(b)** ACF/PACF of Residuals

**Figure 15:** Visualizing Model Residuals

# Neural Network Time Series Model - Model Forecasting

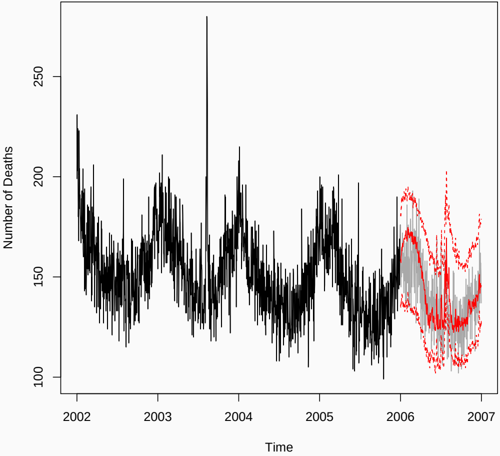



**Figure 16:** Visualizing Forecasted Series