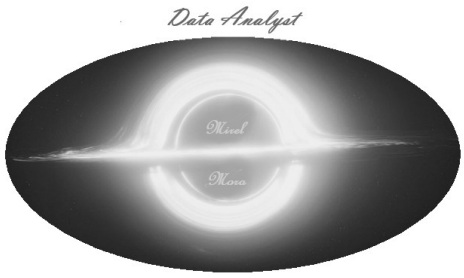
IBM DIGITAL CREDENTIAL

FORECASTING AIR POLLUTION PARTICULATE MATTER (PM2.5)

MIREL MORA



CERTIFICATE APPLIED DATA SCIENCE CAPSTONE

APRIL 14, 2021

Abstract

Air pollution in London is a mixture of emission created locally, and those from background concentrations, particles of pollution can remain suspended for weeks and so can be transported long distances. This has a negative effect on a number of different aspects of human health. In London, 9,400 premature deaths are attributed to poor air quality and a cost of between £1.4 and £3.7 billion a year to the health service[1]. Air pollution also contributed to other environmental issues, such as global warming, acid rain and ozone depletion. However, human well is the primary importance to study about air pollution. Then, Particulate matter (these are usually split into 2 sizes: PM2.5 and PM10) and NO_2. These are usually seen as dangerous form of air pollution due to their high concentrations and the negative health impacts they create. In this report predictive model for forecasting particulate matter concentration in London

[1]London Councils - https://www.londoncouncils.gov.uk/

1 | Introduction

1.1 Air pollution in London

It is often assumed that air pollution in London is a recent phenomenon, however, legislation attempting to control air pollution was enacted as early as 1306. Coal smoke and its associated problems remained a matter of concern in London up until the late 20th century with the famous smogs of the 1950s and 60s.

In recent years, the pollutants in the capital's air have altered considerably. This is primarily because of the decline in the use of coal in industry and domestic heating, which has led to large reductions in the emissions of sulphur dioxide and particles of soot over the past 40 years. At the same time the increased number of motor vehicles is producing considerable amounts of nitrogen dioxide and small particles.

1.2 Problem statement

There is mounting evidence of health effects from everyday exposure to air pollution. The modern-day small particles are our main problem for air pollution health effects, whilst other pollutants such as nitrogen dioxide and ozone are also a major cause for concern.

There are different effects depending on the length and intensity of exposure. For example, short term exposure (a few hours) to high levels of NO_2 can irritate the airways and cause severe coughing and exacerbate existing respiratory illnesses, which is uncomfortable at best, and dangerous at worst for vulnerable people (sick and older or younger people for example).

Long term exposure can contribute to someone developing a number of illnesses, such as asthma, pulmonary disease and lung cancer. It has also been shown to stunt the growth of children's lungs. This is particularly worrying, as around one-third of London's schools have been found to be close to busy roads that suffer illegal levels of NO2 pollution.

1.2.1 Important pollutant

The pollutants most widely referred to in the literature are:

- Particulate matter (these are usually split into 2 sizes: PM2.5 and PM10)
- Nitrogen dioxide (NO_2)
- Sulphur dioxide (SO_2)
- Ozone (O_3)
- And occasionally, Carbon Monoxide (CO)

PM and NO2 are commonly seen as the most dangerous forms of air pollution due to their high concentrations and the negative health impacts they create.

Particulate Matter 2.5 and 10 (PM)

PM is made up of a wide range of materials and arises from a variety of sources. Concentrations of PM comprise primary particles emitted directly into the atmosphere from combustion sources and secondary particles formed by chemical reactions in the air. It consists of a complex mixture of solid and liquid particles of human-made (such as diesel soot) and natural substances suspended in the air (such as sea spray and Saharan dust). In the UK the biggest human-made sources are stationary fuel combustion (power generators) and transport.

NO_2

Nitrogen dioxide is human made, with the major sources of emissions of NO_2 being combustion processes (heating, power generation, and engines in vehicles and ships)1. Nitrogen is released during the combustion of fuel and then combines with oxygen atoms to create

nitric oxide (NO). This further combines with oxygen to create nitrogen dioxide (NO2). Nitric oxide is not considered to be hazardous to health at typical ambient concentrations, but nitrogen dioxide is. Nitrogen dioxide and nitric oxide are referred to together as oxides of nitrogen (NO_x). NO_x gases can also react to form smog and contribute to acid rain. NO_x is also central to the formation of fine particles or particulate matter (PM) and ground level ozone (O_3), both of which are associated with adverse health effects.

1.3 Air pollution forecast

Air pollution forecasting is a worthwhile investment on multiple levels - individual, community, national and global. Accurate forecasting helps people plan ahead, decreasing the effects on health and the costs associated.

If people are aware of variations in the quality of the air they breathe, the effect of pollutants on health as well as concentrations likely to cause adverse effects and actions to curtail pollution. Furthermore, there is a greater likelihood of motivating changes in both individual behaviour and public policy [1], as people want air quality information.

[1] Kelly F. (2012). Monitoring air pollution: Use of early warning systems for public health./

2 | Study Area and Methodology

2.1 Data set description

London is the capital and largest city of United Kingdom that is situated between the latitude of 52.3555° North and 1.1743° West. The data has been collected from London Air via the Openair project tools.

Openair is an R package developed for the purpose of analysing air quality data — or more generally atmospheric composition data. The package is extensively used in academia, the public and private sectors. The project was initially funded by the UK Natural Environment Research Council (NERC), with additional funds from Defra. The most up to date information on openair can be found in the package itself and at the book website here.

Openair tool provide a extensive access to UK air quality data. The networks includes:

- For importing data from the UK national network called Automatic Urban and Rural Network. This is the main UK network. (importAURN)]
- For accessing data from Air Quality Scotland network. (importSAQN)
- For accessing data from the Air Quality Wales network. (importWAQN)
- For accessing data from the Air Quality England network of sites. (importAQE)
- For accessing data from the Northern Ireland network of sites. (importNI)
- A simplified version of a function to give basic access to hourly European data based on Stuart Grange's package see.
- For accessing data from the sites operated by King's College London, primarily includ-

ing the The London Air Quality Network. (importKCL)

In this case, the networks used were UK national, Air Quality England and The London Air Quality.

2.2 Data cleaning

Data downloaded from three different network from Openair are combined into two data frames. The process is divide in the follow steps:

1. Extract data from Openair with R (Github - Jupyter)
2. Create a general data set with three networks (UK national, England and The London Air Quality).
3. Choosing stations (Stations in Greater London).
4. Filling missing values using Spatial interpolation.
5. Filling missing values using Panda method.

2.2.1 Extract data from Openair with R

Openair Tool works with programming language R. Openair has documents where explains how it works the package. It has been collected data from all stations in the United Kingdom for the years 2019 and 2020. It is generated two data set of 1495 stations around the United Kingdom. Data sets are divide into two groups that are particular and chronological data.

a) Particular data in the United Kingdom, England and London

b) Chronological data in the United Kingdom, England and London

2.2.2 Create a general data set with three networks

Generated one data frame with Particular data in the United Kingdom, England and London and another data frame with Chronological data in the United Kingdom, England and London using Pandas Library. Chronological data refers to that information, which quickly