

|  |
| --- |
| Coventry University  5011CEM Big Data Programming Project Specification Document  Low level O3 concentration study |

**Author: Guilherme Miguel**

**Document Version 2.1**

**18/04/2020**

**Table of Contents**

[1. Introduction 3](#_Toc30950202)

[2. Project Requirements 3](#_Toc30950203)

[2.1 Related documents 3](#_Toc30950204)

[2.2 Terms/Acronyms and Definitions 3](#_Toc30950205)

[3. Risks and Assumptions 3](#_Toc30950206)

[4. Out of Scope 3](#_Toc30950207)

[5. System/ Solution Overview 3](#_Toc30950208)

[5.1 Context Diagram/ Interface Diagram/ Data Flow Diagram, Application Screen Flow, Sitemap, Process Flow 3](#_Toc30950209)

[6. Project Management 4](#_Toc30950210)

[7. References 4](#_Toc30950211)

[8. Open Issues 4](#_Toc30950212)

[Appendix 4](#_Toc30950213)

# Introduction

Relating the concentration of O3 with pollution is in fact a smart choice, as the low-level of the gas concentration increases there are several health implications to take in mind. It might not affect the “Average Joe” since the implications it may cause are related with the difficulty to breathe since it can cause the muscles in the airways to constrict, trapping air in the alveoli (EPA, 2019). This can cause shortness of breath, aggravate existing lung diseases or bring new ones such as chronic obstructive pulmonary disease. As mentioned before, this will not affect everyone equally, it will affect the people with weaker lungs such as asthma patients or the older section of the population.

But how is it formed? This low-level ozone comes in fact from air pollution, it is created by a chemical reaction between oxides of nitrogen and volatile organic compounds which both are generated by the combustion of several substances encountered on petrol or other compounds that are used in factories, chemical plants and in our vehicles.

The objective of this project is to show how in fact the concentration changes throughout the day in Europe, deducting if in fact the pollution of the air by the industry can be a factor in the increase of low-level ozone. Visual tools will be created using different climate models from the ECMWF (European Centre for Medium Range Weather Forecasts) in hope of reaching the conclusion that the most developed countries in Europe could be in fact the ones who have higher concentrations during “working” hours.

# Project Requirements

The project is going to be developed with the intention of exploring a specific data set. The data set consists of different models each one able to produce > 250 chemical species and pollutants (Dr Hyde, R).

In theory the values deducted from the different models should be similar, therefore, if such doesn’t happen one or more models might lack accuracy.

Since the goal of the project Is to compare the models such work is going to be prioritized, a hourly plot of each model will be used which will help maintain the focus on the comparison itself.

Visualization of the results must be simple, so it doesn’t consume much project time, it will be the focus of the project therefore it must be well thought, and enough given time must be allocated to it.

The code to create the plots will be done on MATLAB exclusively, and will be divided in three parts, one to plot each model onto a map of Europe, another to calculate the mean concentration of O3 in Portugal per model and a last one where the input will be the mean concentration code and the output will be a chart with the value of the concentration per model. In terms of general outputs the outputs will always be saved as PNG, therefore we can have a picture being saved with any plot created, for the concentrations the output will be the floating point number with the value of the mean concentration of O3 in that specific model.

## Related documents

|  |  |  |
| --- | --- | --- |
| **Component** | **Name (with link to the document)** | **Description** |
|  |  |  |

## Terms/Acronyms and Definitions

|  |  |  |
| --- | --- | --- |
| **Term/Acronym** | **Definition** | **Description** |
|  |  |  |

# Risks and Assumptions

**Assumptions:**

System minimum requirements:

* + CPU – Dual Core CPU ~2.00GHz;
  + Memory – 4Gb (8Gb will improve significantly the speed of the plotting in MATLAB);
  + Operating System – Ubuntu 18.0 or higher, might have support for Mac OS and Windows 10 but the MATLAB installation process might be different
  + Available disk storage - ~5 Gb (accounting MATLAB needing installation)

No limitation on resolution since the outputs will be saved as pictures of the format PNG but it recommended that at least 1440x1080 so that the quality of the image is preserved and there is no stretching.

The input files are of the NC format and the outputs will be either PNG for plots or MATLAB files for the floating point numbers.

**Risks:**

For any file or data corruption GitHub version control will be used, so that the latest stable build is always backed up. Any algorithms not written by the author will have respective references to give credit to the original author, if the author doesn’t allow sharing of same work then it will not be used. Any bugs that are not related with the code but with the system or any standard libraries will be resolved by keeping the system updated throughout the development and usage of the project. The code will be entirely written by the developer so no issues with external code should appear, if such external code is used it will be referenced and therefore any troubleshooting manoeuvres will be addressed.

# Out of Scope

The goal of this project is a comparison of speed calculations of two different data analysis approaches, therefore it must be just that so the visualization of data must be simple but efficient, so that it doesn’t consume much development time. Any other approaches discovered during the development might be logged but not implemented so that the focus of the project isn’t deviated.

# System/ Solution Overview

# Context Diagram/ Interface Diagram/ Data Flow Diagram, Application Screen Flow, Sitemap, Process Flow

# Project Management

To manage the project efficiently it will be used some of the Agile Methodologies that are going to be managed using an online platform name Notion, since it’s a solo project Scrum will be partially off the table because it implies for meetings to be done which are not physically possible with just one individual. Therefore, to organize the work, tasks will be divided in User Stories and placed on a Kanban Board to facilitate visualization. Each week the board is updated so there won’t be any overload of tasks to fulfil in a short amount of time.

As said before, GitHub will the main storage option for the project, that is for two reasons: it does allow that backups are kept in a safe and tidy place where one can easily reach; it allows for work to be tracked so that when the project gets delivered there will be a log with all the changes done during development.

To prove for all of this there will be screenshots of the Notion pages since it can’t be shown as an external file, the pages might also be shared in the delivery process.

# References

United States Environmental Protection Agency, 2019, accessed on 20/02/2020 <<https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics#formation>>

# Open Issues

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Issue ID** | **Issue** | **Raised By** | **Raised On** | **Solution/ Decision** | **Resolved By** | **Resolved On** | **Status** |
|  |  |  |  |  |  |  |  |

# Appendices