

# NO<sub>2</sub> prediction in Madrid, Spain

Meng Lu

October 2017

## 1 Introduction

The number of air quality measurement stations is often limited to quantify air pollution continuously in space and time. Other data sources such as physical and chemical model simulations and spatiotemporal correlations may play an important role in improving air pollution interpolation. In this study, we investigated different Spatiotemporal and non-spatiotemporal statistical models and integrated historical records and predictions from a MACC physi-chemical model simulations to station measurements to predict air quality pollution.

We compared the results of air prediction from linear and non-linear regression models, geostatistical models, and random forest models. The linear and non-linear regression models assume no spatial correlation and purely predict using the MACC model simulations. We consider an ordinary kriging and an universal Kriging model to integrate spatial correlation and predict air pollution. For each linear/nonlinear regression, geostatistical, and random forest models, we attempted different methods and attempted to achieve the best results from these models.

The objectives are:

- Can spatial correlation between a limited number of air quality measurement stations be used to improve air quality prediction?
- Can the MACC model simulation improve air quality prediction?
- How can we construct the random forest model to predict air quality with station measurements, MACC simulation, and temporal variables? Does the inclusion of time series characters, such as trend and harmonic terms improve model prediction?

## 2 Method

1. Interpolate MACC simulation
2. Linear regression model with MACC as an independent variable.

3. Polynomial and GAM (general additive model) with different orders of MACC as independent variables.
4. Universal Kriging with MACC simulation.
5. Random forest prediction with MACC and different temporal variables.

### **3 Data**

Three years historical record and prediction of MACC simulation grid. Three years 24 station measurements (to train model or validate model)