Air pollution in Japanese skies :

Kriging estimation of the time dependent distribution of airborne fine particles (PM2.5) above Okayama city, Japan - Electrochemical batteries analogy.

April 01st - September 01st, 2017

CHICHE Ariel
3A EPEE
2016/2017
Supervisor : Pr Junji Yamakawa
ufeq0285@cc.okayama-u.ac.jp

Friday September, 8th 2017

Acknowledgments

First of all I would like to thank Pr Yamakawa for his daily support and his very interesting directions for this internship. He was a really good supervisor and he taught me new methods and explained to me important details about Japanese culture. He created a nice atmosphere in the laboratory and trusted me so that I could work in the best conditions.

I would also like to thank Pr Suzuki for his help in the administrative works and his interest in my work. I also appreciated to be invited to some geology lessons, this field was very new for and I could discover it in a simple and entertaining way.

Dr Chenevier was also very helpful for me and I would like to thank him. He helped me a lot for all the procedures to come to Japan. He also helped me to write this master thesis by giving me some very precious advices.

I would like to thank my laboratory colleague, Miss Yu Kurozumi, for her kindness and for welcoming me very well in the laboratory.

Last but not least, I want to thank the Department administrator, the Faculty administrator and the Global partners of Okayama University for their acceptance of this internship and their very good help, especially with all the administrative works.

Abstract

English Particulate matters in the air above cities is a real issue for modern society, indeed a too long exposure to those particulates matters can have dramatic effects on the human respiratory system and can become a real general health care issue. This is why it is important to analyze the distribution of the particulate matters above the cities in order to highlight a periodicity of the distribution so it will be possible to foresee some peaks of pollution and take measures. Okayama prefecture has a network of observatories that evaluate the concentration of those particulate matters in the air every hour. Such studies requires geostatistical skills and analysis and this is why the Kriging method, that can make prediction of a value at several points using value of actual points thanks to their location, is important in this study. It allows to make PM2.5 concentration prediction on an interesting area and to make some statistical analysis in order to find the time dependent change of those particulate matters. Several weeks were analyzed and compared in this study and it was possible to see a 40 - 45 hours general periodicity of the PM2.5 distribution. This value is very important and the authorities can now take measures to reduce the risks of exposure using this periodicity. Finally, this surprising result can be related with the cycle of charge and unloading of electrochemical batteries. An analogy can be made between the concentration of PM2.5 and electrochemical batteries, this analogy shall be studied in the future.

French Les particules en suspension dans l'air sont un véritable problème pour les sociétés modernes, en effet, une trop longue exposition à ces particules peut avoir des conséquences sur le système respiratoire humain et devenir un véritable problème de santé publique. La préfécture d'Okayama (Sud - Ouest du Japon) possde un réseau de stations d'observation qui enregistrent la concentration de ces particules en suspension dans l'air chaque heure. De telles études nécessitent des compétences en géostatistiques car la maille du réseau de stations est grande. La méthode de Kriging, qui réalise l'interpolation spatiale d'une variable régionalisée par calcul de l'espérance mathématique d'une variable aléatoire en utilisant un variogramme expérimental, a donc été utilisée dans le cas des densités de particules dans l'air afin prédire des concentrations de PM2.5 sur une zone d'étude et de faire des analyses statistiques pour trouver la dépendance en temps de la distribution des particules en suspension dans lair. Les données collectées pendant plusieures semaines ont été analysées et comparées dans cette étude et il a été possible de mettre en lumière une périodicité générale de 40 - 45 heures de la distribution des PM2.5. Cette valeur est importante et peut aider les pouvoirs publics dans leur politique de prévention sanitaire, en particulier en adaptant les paramètres de régulation du traffic routier. Enfin, cette surprenante valeur de périodicité peut être comparée au temps de cycle d'une batterie éléctrochimique. Une analogie peut être faite entre le fonctionnement d'une batterie et l'évolution de la densité des PM2.5 dans l'air et des études afin de montrer cela devraient être menées dans le futur.

Contents

List of Figures

Introduction

Particulate matter (PM) (also called particle pollution) is a mixture of airborne solid particles and liquid droplets found in the air. Some particles, such as dust, dirt, soot, or smoke, are large or dark enough to be seen with the naked eye. Particulate matter contains microscopic solids or liquid droplets that are so small that they can be inhaled and cause serious health respiratory problems. Particles smaller than 10 micrometers in diameter pose the greatest problems, because they can get deep into lungs, and some may even get into bloodstream. Most particles form in the atmosphere as a result of complex reactions of chemicals such as sulfur dioxide and nitrogen oxides, which are pollutants emitted from power plants, industries and automobiles. PM10 and PM2.5 include toxic particles that are small enough to penetrate the thoracic region of the respiratory system. At present, at the population level, there is not enough evidence to identify differences in the effects of particles with different chemical compositions or emanating from various sources.

For all those general health care reasons, PM10 and PM2.5 distribution above some region is very interesting and crucial to study, it is really important to find the periodicity of those particulate matters distribution in order to foresee peaks of pollution and take measures in order to reduce their severity. It is very crucial since it will help to reduce the risks of exposure for fragile people. Okayama (different from the larger Okayama prefecture area) is a middle size Japanese city (almost 1 000 000 inhabitants) and is famous for its dynamic University. That is why some researchers try to study the evolution of those particulate matters in a new way and one of them is Pr Yamakawa. There are several observation stations, they are scattered in such a way they evaluate particulate matter concentrations every hour all around Okayama prefecture, it gives a discontinuously spacial view of the particulate matter in the city and the main idea is to have a continuous one using geostatistical methods. This is why the use of the so-called "Kriging" method can a great help. The Kriging method (originates in the late 40's) allows to make predictions of a value at many locations using actual data. In the case of the air pollution of Okayama area, it is very interesting to use this method since that is allows to make concentration predictions of those particulate matters for many location around Okayama so that a clear view of those particulate matters distribution can been obtained.

This study is the opportunity to discover and use the Kriging method but also to highlight the periodicity of the distribution of some of those particulate matters. It is clear that the weather conditions and the traffic activity have an impact on those particulate matters and it is very interesting to observe it.

Glossary

PM : Particulate Matter
SK : Simple Kriging
UK : Universal Kriging
EPA : Environmental Protection Agency (USA)
KM : Kriging Method

Chapter 1

Presentation and subject

1.1 Okayama University and the Mineral Science Laboratory

This internship took place at Okayama University, a national university in Japan. The main campus is located in Tsushima-Naka, Okayama, Okayama Prefecture (location *Figure 2.8*). The school was founded in 1870 and it was established as a university in 1949. Okayama University, one of Japan's leading universities, is situated on an extensive campus and has eleven faculties and Master's and Doctoral courses embracing humanities, social, natural, environmental and life sciences, and pedagogic. Approximately 14,000 students including more than 600 international students are currently taking lessons in this university, and 3,000 staff are engaged in educational and research activities, and university management.

The Mineral Science Laboratory is located in the Graduated School of Natural Sciences and technology (Master's/Doctoral), this school was established in April 1987 to meet strong social and academic needs in the various fields of natural science. Now this school is a very important one in the University and student come to take classes of geology, environmental sciences or applied chemistry. Originally, the Mineral Science Laboratory is a laboratory where students study rocks sciences, mineralogy and crystallography but lately Pr Yamakawa has decided to use some geostatistic sciences in order to study the environment and the air quality of Okayama prefecture. That is why he offered to host this internship in this very laboratory, indeed, he wanted someone who would spend his all time here making geostatistic analysis to study the pollution of the air in the region.

1.2 Internship subject and materials

1.2.1 Particles Matters : PM2.5 and PM10

Particulate matter (PM) is the sum of all solid and liquid particles suspended in air many of which are hazardous. This complex mixture includes both organic and inorganic particles, such as dust, pollen, soot, smoke, and liquid droplets. Particle matters in the air are a real issue for all the urban areas since the human activity (industries, traffic ...) have a huge impact on those particles and their concentration in the air. [1] [2] [9]

The aerodynamic properties of particles determine how they are transported in air and how they can be removed from it. These properties also govern how far they get into the air passages of the respiratory system and then be a real threat for human health. Additionally, they provide information on the chemical composition and the sources of particles. Particles have irregular shapes and their aerodynamic behavior is expressed in terms of the diameter of an idealized sphere. The sampling and description of particles is based on this aerodynamic diameter, which is usually simply referred to as particle size. Particles having the same aerodynamic diameter may have different dimensions and shapes. Some airborne particles are over 10,000 times bigger than others in terms of aerodynamic diameter. Particulate matter is often divided into two main groups:

- The coarse fraction contains the larger particles with a size ranging from 2.5 to 10 micro-m : **PM10**.
- The fine fraction contains the smaller ones with a size up to 2.5 micro-m : **PM2.5**. The particles in the fine fraction which are smaller than 0.1 are called ultrafine particles.

Those particles can have a dramatic effect on the human health and even healthy people, may feel temporary symptoms if they are exposed to high levels of particle pollution. Numerous scientific studies connect particle pollution exposure to a variety of health issues, including: irritation of the eyes, nose and throat, irregular heartbeat, asthma attacks or heart attacks, that is why it is very important to study their concentration above cities and to try to find some properties and periodicity in their behavior. the United States Environmental Protection Agency (EPA) established National Ambient Air Quality Standards for PM2.5 in 1997 and revised them in 2006. National Ambient Air Standards are established to be protective of public health. The short-term standard (24-hour or daily average) is 35 $\mu g/m^3$ of air and the long-term standard (annual average) is 15 $\mu g/m^3$.

1.2.2 Internship subject

Okayama Prefecture has several observatories that analyze the quality of the air. Many of those observatories are used to estimate the PM10 concentration every hour and some are used for the PM2.5 concentration, those data can be collect and analyze by anyone who wants to. So, thanks to those observatories, is it possible to have the concentration on PM10 and PM2.5 at some locations of Okayama Prefecture, it is in fact a discrete view of the PM10 and PM2.5 distribution above this region. The main goal of this internship is to obtain a continuous view of PM2.5 and PM10 concentration, distribution and accumulated weight above Okayama city thanks to those observatories data by using geostatistical methods called Kriging.

Furthermore, by analyzing those data and making spatial prediction for every hour, the goal is to highlight a clear periodicity for PM2.5 distribution and to prove the effect of several parameters on it such as humidity or atmospheric pressure. This also the opportunity to discover the Kriging Method which is not very famous that is why this internship's name is : **Kriging estimation of the time dependent distribution of PM2.5 above Okayama city, Japan**.

Therefore, before the beginning of this internship, Pr Yamakawa has already made some kriging predictions but the procedure was complicated and took long time so it will be explained in this report how the prediction method has evolved and what are the main results.

Finally, the weather conditions in Japan can change very quickly during the summer season that is why it will be very interesting to study the effect of dramatic weather change (typhoon, yellow sand ...) on PM2.5 distribution and to compare those results with more stable period in order to highlight the effect of weather change on those particles.

1.2.3 Materials

In order to do all the geostatistical analysis, it was very important to have several powerful softwares and website, those were very crucial to collect some PM, weather or traffic data and make analysis. Here is a exhaustive list of those materials and their purpose :

- GNU Octave : a high-level interpreted language, primarily intended for numerical computations. It is a free version of Matlab. In order to make all those statistical analysis two libraries were used, **STK** (Simple tool for Kriging, this library has the functions that make the kriging prediction) and **Signal** (this one was full of statistical function that were very useful during this internship)
- gstat.exe : this software was used at the beginning of the internship to make kriging prediction.

- QGIS : a cross-platform free and open-source desktop geographic information system (GIS) application that supports viewing, editing, and analysis of geospatial data. This software was used to visualize the results of the kriging predictions and highlight the very concentrated areas.
- Google Earth : an other geographic information system. This one was very useful to see the human activity in the prediction areas.
- http://pref-okayamataiki.blue.coocan.jp/top.html : this website was the one used to collect the PM2.5 and PM10 concentration hour by hour everyday.
- Microsoft Power BI : a new Microsoft software used to collect data from the internet, in this case the PM concentration and to merge those data with other important data such as the location of the observatories, that was very important for the kriging prediction.
- http://www.data.jma.go.jp, this website gives all the data about the atmospheric pressure conditions and the humidity.
- http://www.mlit.go.jp/road/census/h27/index.html, this website gives some informations about the car traffic in Okayama.

All those softwares and website were very important to make all the analysis and to make the Kriging procedure way more efficient and fast. Some of them like QGIS, Power BI or gstat.exe have never been used before so a period of learning was necessary but thanks to Pr Yamakawa it was easy to understand those new tools and to use them for the study.

1.3 General theory of Kriging

1.3.1 Definition

Kriging is a method of spatial interpolation that originated in the field of mining geology as is named after South African mining engineer Danie Krige. As said before the idea is to make prediction and obtain continuous value out of discrete data. In this study the value is the concentration of PM2.5 and PM10. Kriging also generates estimates of the uncertainty surrounding each interpolated value.
The Kriging weights are calculated such that points nearby to the location of interest are given more weight than those farther away. Clustering of points is also taken into account, so that clusters of points are weighted less heavily (in effect, they contain less information than single points). This helps to reduce bias in the predictions. [4] [3] [5] [6]

$Z(x_0)$ is the prediction value at location x_0.
m is the average value of the samples.

Therefore, the theory of Kriging was not the real subject of this internship and has not really been study, that is why this part is a general presentation of the Kriging, its procedure and its most important parameters.

The Kriging predictor is an optimal linear predictor and an exact interpolator, it means that each interpolated value is calculated to minimize the prediction error for that point. The value that is generated from the Kriging process for any sampled location will be equal to the observed value at this point, and all the interpolated values will be the Best Linear Unbiased Predictors (BLUPs).

There are two main assumptions to make Kriging prediction and to provide best linear unbiased prediction are:

- Stationarity : the joint probability distribution does not vary across the study space. Therefore, parameters (such as the overall mean of the values, and the range and sill of the variogram) do not vary across the study space.
- Isotropy : uniformity in all directions

In this study, since data are collected every hour it is possible to say that this is a stationary case and it is clear that there is no favorite directions for PM2.5 and PM10 so it is possible to use the Kriging procedure to make predictions.

Finally, one of the most important value for the Kriging is the **Weight**, λ. It can be compare to the weight of a point for a barycenter, indeed, the weight corresponds to the importance of each sample location for each prediction. The weights are determined from the variogram based on the spatial structure of the data, and are applied to the sampled points.

1.3.2 Variogram

A **Variogram** or semi-variogram is a function describing the degree of spatial dependence of a spatial random field Z(x) Z(y). A variogram is used to display the variability between data points as a function of distance. For each pair of points in the sampled data, the gamma-value which is a measure of the half mean-squared difference between their values is plotted against the distance, or lag, between them.

Here is the basic equation of the variogram (2γ(x,y)) :

$$2\gamma(x, y) = var[Z(x) - Z(y)] \tag{1.1}$$

Variogram models are drawn from a limited number of functions, including linear, spherical, exponential, and power models as it is shown behind. Making a variogram is plotting a theoretical one and add the experimental one to see which model fits the best and because the Kriging algorithm requires a positive definite model of spatial variability, the experimental variogram can not be used directly. Instead, a model must be fitted to the data to approximately describe the spatial continuity of the data. Certain models that are known to be positive definite are used in the modeling step.

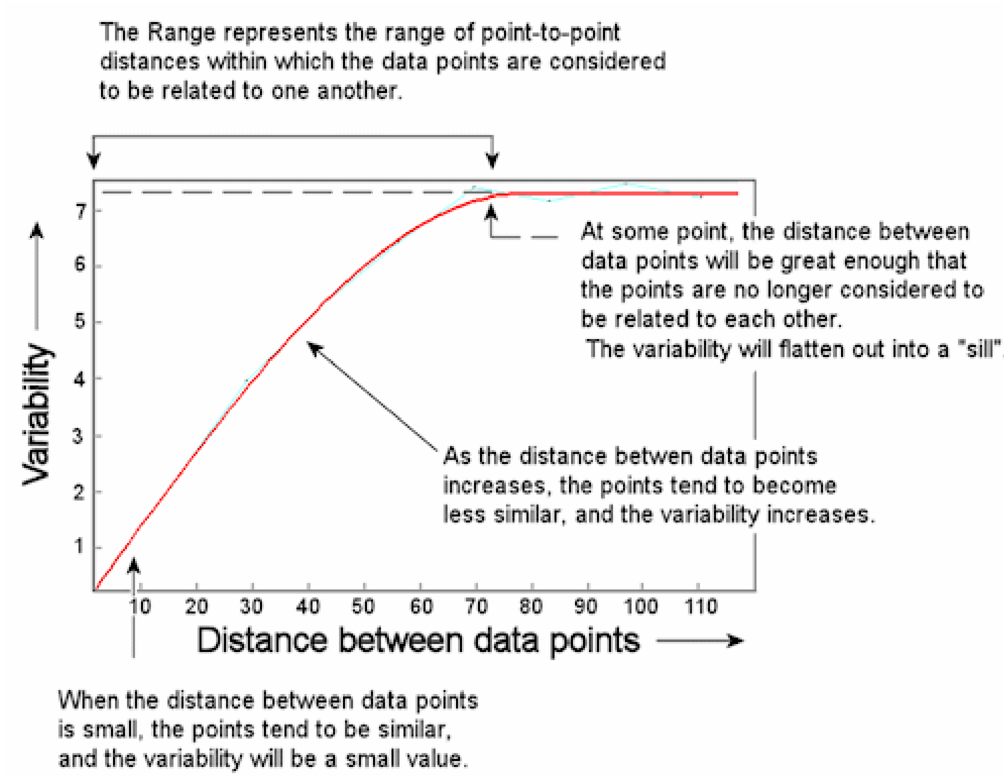
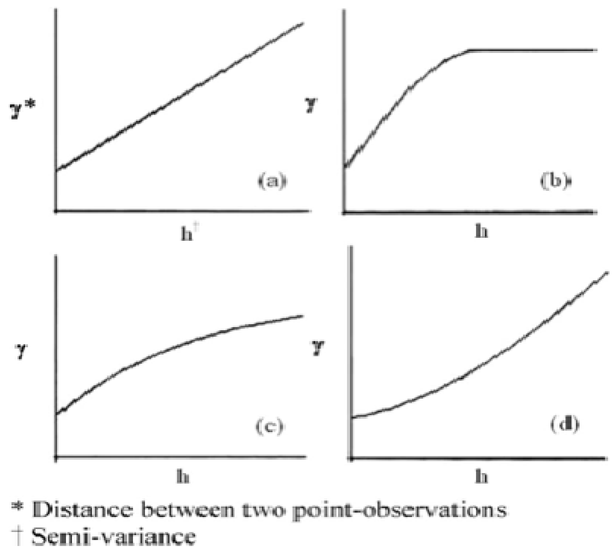


Figure 1.1: Example of an explained variogram

The *Figure 1.1* (Source : http://help.rockware.com) shows the evolution of a typical Variogram and how to analyze it, indeed, it shows that at some point the two locations are two far to have a strong correlation between each other, that is why the relative of those points will be very low for the prediction.

In fact, the variogram is very important because it is a thanks to it that it is possible to know which model to use for the Kriging Predictions so it is crucial to plot some variograms before making the predictions.

The *Figure 1.2* (Source : https://www.mailman.columbia.edu) shows the different types of variogram models that are often used in such analysis. The variogram remains of the most important tool in order to prepare the Kriging prediction procedure. As said before, it is very important to plot a variogram, that will also help us to choose which type of Kriging should be used for the predictions and the analysis.



(a) Linear semi-variogram; (b) spherical semi-variogram; (c) exponential semi-variogram; and (d) power semi-variogram.

Figure 1.2: Different types of Variograms

1.3.3 Different types of Kriging

There are several sub-types of kriging, including:

- **Simple Kriging** (SK), for which the assumption of stationarity must be assumed. This is one of the simplest forms of kriging, but the stationarity assumption is not often met in applications relevant to environmental health, such as air pollution distributions. This is the most used and simple Kriging method. Here is the equation of Simple Kriging :

$$Z(x_0) = m + \sum_{i=1}^{k} \lambda_i Z(x_i) \tag{1.2}$$

- Ordinary Kriging, it is assuming that the mean is constant over the entire domain and that it is constant in the local neighborhood of each estimation point. Here is the equation of Ordinary Kriging :

$$Z(x_0) = m(1 - \sum_{i=1}^{k} \lambda_i) + \sum_{i=1}^{k} \lambda_i Z(x_i) \tag{1.3}$$

in this case,

$$\sum_{i=1}^{k} \lambda_i = 1 \tag{1.4}$$

- Universal kriging (UK), this method is more accurate, the main idea is to use the results of Simple Kriring of some value that should be very close to the wanted predictions : for example, since there are more PM10 observatories than PM2.5's it is possible to use the SK results of PM10 prediction as a model for PM2.5 predictions. This method will be discussed later in this report. It also relaxes the assumption of stationarity. Here is the equation for Universal Kriging :

$$Z(x_0) = \sum_{i=1}^{k} \lambda_i Z(x_i) + \sum_{i=1}^{k} \lambda_i X_1(x_i) \tag{1.5}$$

  Where X1(xi) is the value of the other results.

- Block kriging, which estimates averaged values over gridded blocks rather than single points. These blocks often have smaller prediction errors than are seen for individual points.

- Poisson kriging, for incidence counts and disease rates.

- And more

In this study only Simple Kriging and Universal Kriging have been used to make predictions because those are the most famous ones and easy to make.

Therefore, Kriging has some limitations and weaknesses, indeed, since it really depends on the variogram the predictions are very sensitive to mis-specification of the variogram model and moreover the accuracy of interpolation by Kriging will be limited if the number of sampled observations is small, the data is limited in spatial scope, or the data are in fact not amply spatially correlated.

1.4 Cross-Correlation and Auto-correlation

Statistical correlation is a statistical technique which tells if two variables are related. The main result of a correlation is called the correlation coefficient (or "r"). It ranges from -1 to +1. The closer r is to +1 or -1, the more closely the two variables are related. If r is close to 0, it means there is no relationship between the variables. If r is positive, it means that as one variable gets larger the other gets larger. If r is negative it means that as one gets larger, the other gets smaller (often called an "inverse" correlation). In statistics, the bias (or bias function) of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated. An estimator or decision rule with zero bias is called unbiased.

Cross-correlation is the study of the correlation between two different parameters whereas **auto-correlation** is the correlation between a parameter and itself that is why auto-correlation will be used to highlight the time dependent change of PM2.5 distribution. Cross-correlation will be used to show the effect of weather conditions and car traffic on PM2.5 distribution, it will also be used to compare different PM2.5 predictions (typhoon, yellow sand ...).

The population correlation coefficient corr(X,Y) between two random variables X and Y with expected values μX and μY and standard deviations σX and σY is :

$$corr(X,Y) = \frac{E[(X - \mu X)(Y - \mu Y)]}{\sigma X \sigma Y} \tag{1.6}$$

The *Figure 1.3* (Source : http://work.thaslwanter.at) shows an example of typical cross-correlation between two parameters. It shows that there are peaks of correlation when the two parameters are both increasing or decreasing. Cross-correlation and auto-correlation will be very important in this study.

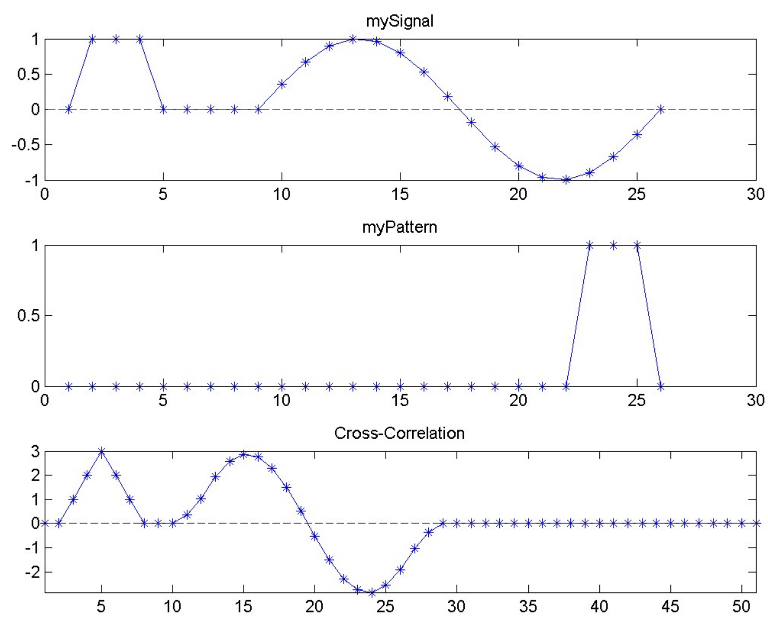


Figure 1.3: Example of a cross-correlation

Chapter 2

Method

Now that the subject and the theoretical points are settled, it will be interesting to explain the method and some achievement of this study. Since hour by hour concentration data are available for free, the main idea is to acquire all those data thanks to Power BI by Microsoft and to make a prediction for every hour for several weeks.

2.1 The Kriging script

Before the beginning of this internship, the Kriging was made thanks to the software gstat.exe in this laboratory. It required one specific script for each hour and it was necessary to plot a variogram for every hour before making the prediction so it took a really long time and it was not really convenient to study lots of data. Furthermore, it gave only 3-dimensional or 2-dimensional results but it was very complicated to estimate some average value of concentration to see the hourly and daily evolution of PM2.5 concentration and to make proper analysis for over 5 days long. Moreover, as it will be explained in the examples part, the 3-dimensional result was not very precise since it was just the plotting of a Matrix and then the location data were removed so it was not very easy to understand and use after. Finally, this method does not allow to make statistical analysis easily.
Therefore, even if this very first way of making Kriging was nos convenient and had lot of problems, it was a good opportunity to discover the Kriging procedure, understand this importance of the variogram and obtain some results. It was really useful at the beginning since it was the only method known so far in the laboratory to make Kriging prediction.

That is why after having being trained to use this complicated and long method, the main task of this internship was to create a more simple and efficient script that would able the users to make Kriging predictions for several continuous hours. Then it has been decided to use the software Octave using the STK library and the Signal one. This software is very close to Matlab so it was easy to start using it since the working environment was already familiar.

It took several weeks to make a clear script and one of the biggest advantage of this STK library is that it allows the user to get ride of the variogram as long as the types of variogram has been specified by the user, and because it is the same for all those prediction there is no need to plot a variogram every time.

Eventually, after several weeks, the final Octave script was completed and it allows to make the Kriging predictions and the statistical analysis at the same time for 5 continuous days (120 hours) of PM2.5 and PM10 data. It takes between 6 and 7 minutes to make all those analysis and the results are :

- Kriging Predictions for each hour (2D).
- Mean of the concentration in function of the time for the big area and each divided area.
- Accumulated weight of PM in function of time for the big area and each divided area.
- Auto-correlation of the PM accumulated weight for the big area and each divided area.
- Cross-correlation of PM accumulated weight or concentration between 2 different weeks for the big area and each divided area.
- Cross-correlation between the atmospheric pressure and the PM accumulated weight in the period.
- Cross-correlation between the humidity and the PM accumulated weight in the period.
- Cross-correlation between the basic traffic activity and the PM accumulated weight of the period.

Here are some explained key parts of this script :

```
DIM = 2;                    % Dimension of our study : 2 (because @ cordonates : X and Y)

% Create the prediction Area
BOX = [[-60000;-20000], [-170000; -140000]];      %%% Those are the cordonates of our Prediction Area
BOX = stk_hrect (BOX, {'Coordonate X', 'Coordonate Y'});
NT = 2500;    % Number of points that will be evaluated ==> NT^1/DIM has to be an integer (2500 or 10000)
XP = stk_sampling_regulargrid (NT, DIM,BOX) ;    % Generate the points
CONTOUR_LINES = 100; % number of levels in contour plots
DOT_STYLE = {'ro', 'MarkerFaceColor', 'r', 'MarkerSize', 4};

load XYZ.txt
XPP = [XYZ(:,1) XYZ(:,2)];
```

Figure 2.1: Octave Script for the contour of the prediction area

This part is very important because it sets the prediction area and it is the very first step for the Kriging. But the most important one is the following one, indeed, it is the Kriging itself.

```
NII = length(PM25_24_28_July(:,1));
model = stk_model ('stk_gausscov_iso', DIM);
% We just need to change the value of 'order' in the model
XII = [PM25_24_28_July(:,1) PM25_24_28_July(:,2)];
tic
for i = 3:122

% Re-estimate the parameters of the covariance
[param0, model.lognoisevariance] = stk_param_init (model, XII, PM25_24_28_July(:,i), BOX);
model.param = stk_param_estim (model, XII, PM25_24_28_July(:,i), param0);

% Carry out kriging prediction
zp = stk_predict (model, XII, PM25_24_28_July(:,i), XPP);
D = i-2;

for j=1:2500
if (zp(j,1) < 0)
  zp(j,1) = 0;
  end

  if (XYZ(j,3) > 250)
      zp(j,1) = 0;
end
end


%figure(i-2)
%contour (XP, zp(:,1), CONTOUR_LINES);
%tsc = sprintf ('Simple Kriging PM2.5 approximation for  %d h', D);  hold on;
%plot (XII(:, 1), XII(:, 2), DOT_STYLE{:});
%hold off;  axis (BOX(:));  stk_title (tsc);
% Plot the result
```

Figure 2.2: Octave Script for Kriging Prediction

Finally, the part of the script that makes the evaluation for the accumulated weight is very important : The value of 250 m will be explained later in this

```
Delta_X = 816;
Delta_Y = 612;

Accumulated_Weight = mean((250-XYZ(:,3))) * Delta_X * Delta_Y * sum(zpp(1:2401,:));
```

Figure 2.3: Octave Script for the accumulated weight

report. This part of the graph makes the calculation for all the prediction area.

2.2 First examples and results

In this part, some examples of Kriging will be shown and explained.

2.2.1 With gstat.exe

As said before, the firsts predictions were made thanks to gstat.exe and even if this was not the most efficient way to make those predictions it was a good beginning to have some results. Therefore, gstat.exe was very useful to plot some variogram. The *Figure 2.4* shows one the first plotted variogram in this study.

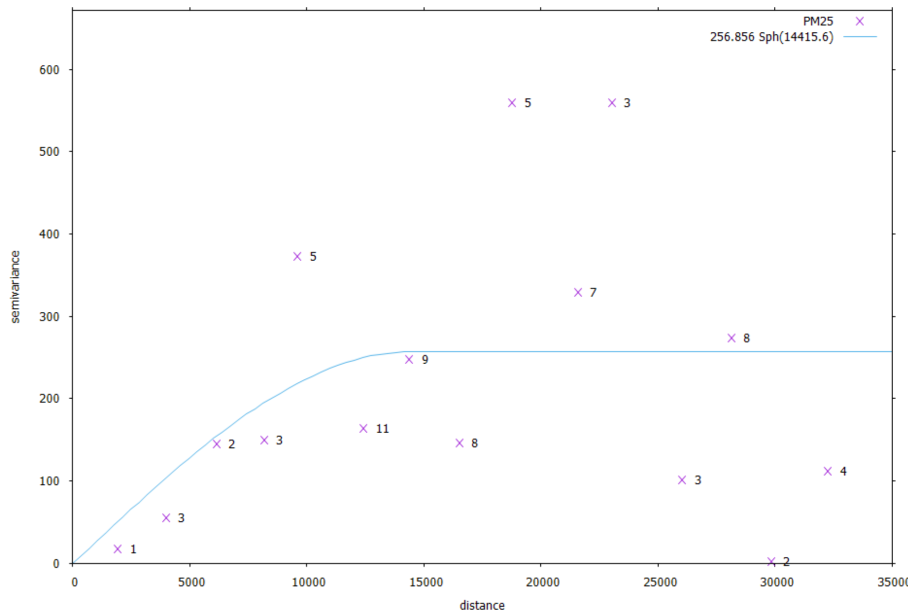


Figure 2.4: Example of gstat.exe variogram of PM2.5 on Saturday April, 8th 2017 at 17h

Many other variograms has been made during this study but the shape was always the same, it was a spherical semi-variogram. This result is crucial, indeed, it proves that **the Kriging model that must be used for the predictions is the spherical one**.

Furthermore, the *Figure 2.5* shows a 3D graph of simple Kriging prediction of PM2.5 on a large prediction area above Okayama prefecture. As said before, it does not give the real location of the points, therefore, it is a clear graph that highlights the high concentrated zones in the prediction area. It is possible to see that the medium concentration is around 23 $\mu g/m^3$ but some peaks are at 40 $\mu g/m^3$, this is really serious since as it has been said in the first part, the EPA has established a limit of 35 $\mu g/m^3$ as a daily average. This is just a peak at one precise hour but gstat.exe does not allow to make 24 hours analysis easily, such analysis will be done thanks to Octave. Finally, it is also interesting to see that some areas are more concentrated in PM2.5 than others, it might be due to human activity.

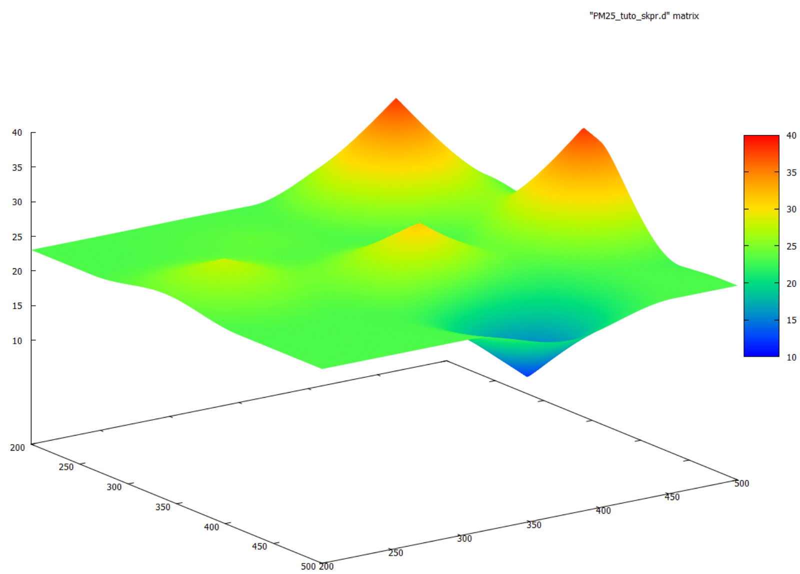


Figure 2.5: 3D graph of PM2.5 predicted SK concentration in microgram/m3 on Sat April, 8th at 12h

2.2.2 Typical results with Octave

Here are some results and explanations from the analysis of the following period : July 17 to 21, 2017. The idea is to make some PM2.5 Kriging predictions, statistical analysis and to study the periodicity of its distribution. During this week, the weather was changing pretty fast and there were no real stability. The SK prediction of the accumulated weight of PM2.5 is very interesting (*Figure 2.6*), it shows two dramatic drops of the PM2.5 weight :

- during the night between Monday and Tuesday and the Tuesday morning, a really raining period
- during the night between Thursday and Friday, a really humid night but not a raining one

It is quiet clear that there is a peak (small or huge) every 40 - 45 hours. This is not a clear proof but it shows that the periodicity of PM2.5 must be between 40 and 45 hours.

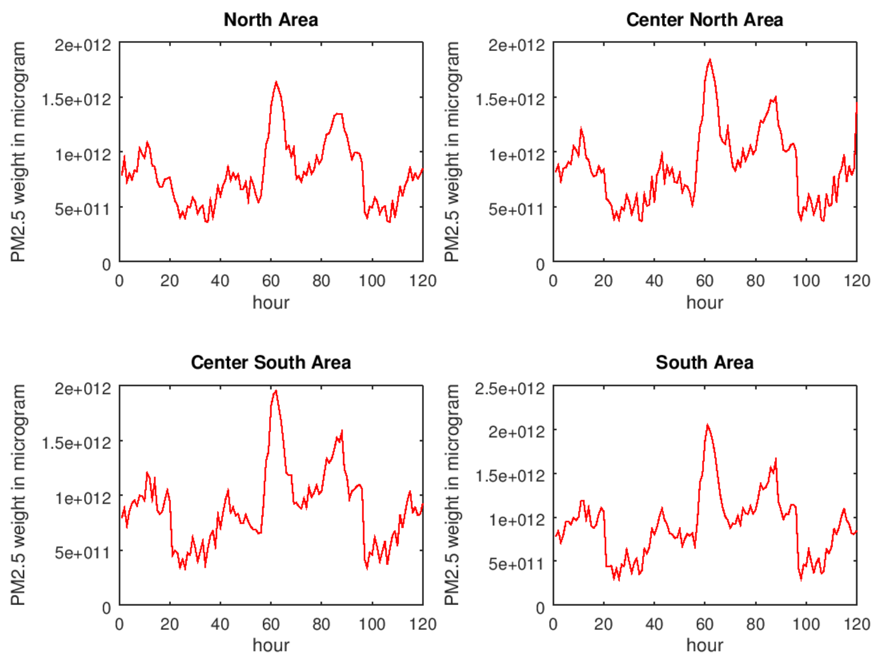


Figure 2.6: Predicted Accumulated weight of PM2.5 during 5 days : July 17 to 21, 2017

This second drop of PM2.5 correspond to a peak of the humidity and also of the atmospheric pressure. This new result is very interesting because it shows that the rain is not the only way to make those particles matters fell down to the ground. It is possible that the humidity increases the density of the air and because of the high pressure of the air, the hot air, which is lighter, goes up in the air creating a convection effect that lifts the PM2.5 because they are very light too. It is just an hypothesis but it can explain this phenomenon.

Furthermore, this week, the weather conditions were changing very quickly and since the distribution of PM2.5 really depends on the atmospheric conditions, it is complicated to have clear conclusions.

It was very interesting to estimate the cross-correlation between last week PM2.5 data analysis and the analysis made last week. Hopefully, the correlation will be high and there will have a clear peak that highlights the periodicity of PM2.5 distribution in the air. (*Figure 2.7*)

Unfortunately, even if the cross-correlation value is very high, there is no clear peak of cross-correlation. There is one small peak for at 40 - 45 hours, this is

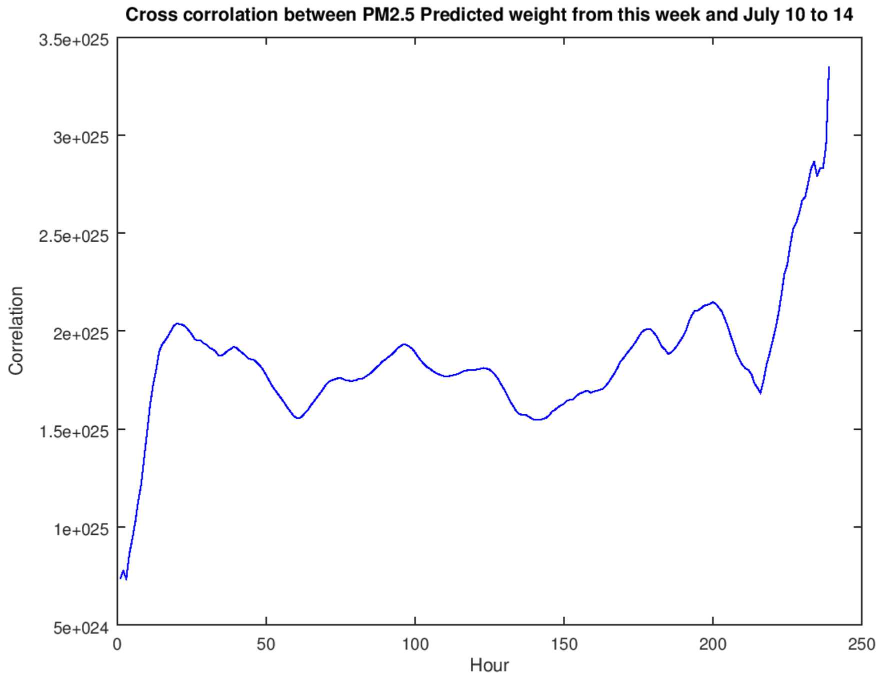


Figure 2.7: Cross-correlation between predictions for two weeks

a very little one but this might be an other proof of the previous hypothesis made. Anyway, it seems to be sure that the value of 40 - 45 hours must be very important for the PM2.5 distribution since it appears every week in our analysis. It will be interesting to make more analysis to see if this peak is really always here. It is possible to suppose that there will have a small or huge peak at this time, the more we will have that kind of results, the more we will be able to make conclusion and say that the periodicity of PM2.5 distribution is about 40 - 45 hours.

In plus, some cross-correlation analysis has be made between the PM2.5 predicted accumulated weight and the weather conditions (Atmospheric pressure, humidity) but those results do not show some interesting peaks that would prove any periodicity. Therefore, the cross-correlation value is very high, that is an other proof of the influence of those two parameters on the PM2.5 distribution.

Finally, it is interesting to see that the value of the predicted accumulated weight of PM2.5 seems to be higher in the crowded urban region (Center North) and the industrial one (South).

2.3 Main problems and achievement

It is now interesting to show some results and some problems faced during this study.

2.3.1 Prediction area

At the beginning, the prediction area was too large and the Kriging predictions were not very accurate, especially for PM2.5 since there are very few observatories for those PM it is shown in the *Figure 2.9*. The *Appendix A* shows the difference between the big area and the small final one. That is why it has been decided to reduce dramatically this prediction area and to center it around an area with a lot of PM2.5 observatories.

Furthermore, at first, the prediction area was made thanks to QGIS and it

Figure 2.8: Shape of Okayama Prefecture with PM2.5 observatories (QGIS)

was complicated to have a precise one, then by using Octave and STK for the Kriging it was possible to generate 2500 prediction points on the area but, unfortunately there were no information about the altitude of those points. Then, with DEM data from Pr Yamakawa and QGIS, it was possible to generate 2500 new points from the prediction area and merge them with their altitude. This is how the XYZ.txt file was created, it contains the prediction points with their altitude. This file would be loaded and the script would use it as new points to make Kriging prediction. Finally, **the final prediction area has a size of 1200** km^2 and the mesh is around 500 m^2.

Eventually, it was important to divide the prediction area in function of the human activity and of the traffic. That is why it has been divided in 4 small equal areas *(Figure 2.9)* :

- North Area : an urban area.
- Center North Area : a really crowded and urban area.
- Center South : an urban area.
- South : a calm area with an industrial activity.

It will be very interesting to see the difference of the distribution between those 4 areas and to highlight the influence of the human activity on it.

Figure 2.9: Shape of Okayama Prefecture with the prediction areas (QGIS)

2.3.2 Simple or Universal Kriging ?

One of the most important issue of this study was to decide between two types of Kriging, the Universal Kriging using Simple Kriging Predictions of PM10 for PM2.5 and the Simple Kriging for PM2.5. In order to figure out which method should be used, several analysis has been made during many weeks, here is an example of such an analysis.

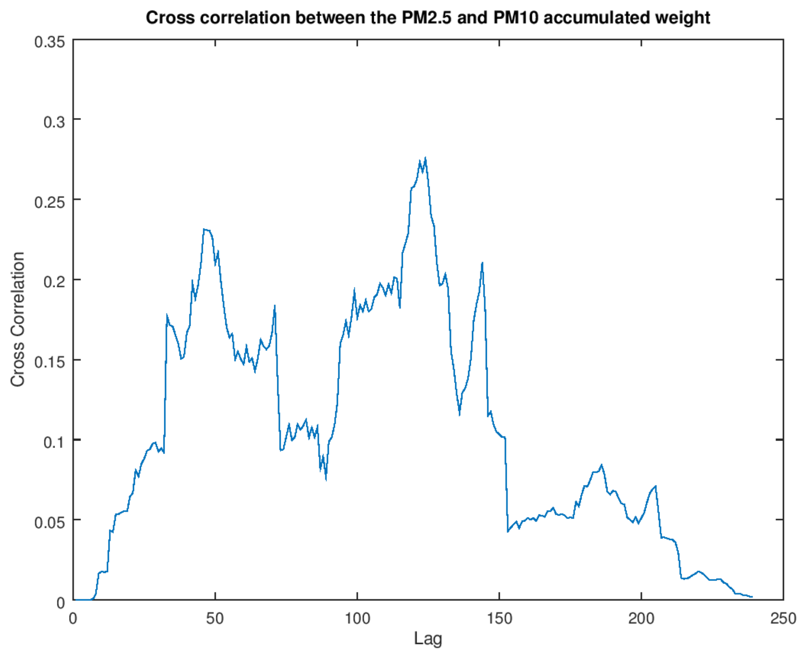


Figure 2.10: Cross-correlation between PM2.5 and PM10 predicted accumulated weight : July 24 to 28, 2017

The *Figure 2.10* shows the cross-correlation between PM2.5 and PM10 simple Kriging prediction during the following period : July 24 to 28, 2017. This graphs reveals that there are several peaks of correlation but the correlation rate is not very high. That is a fist proof that Simple Kriging is a better solution than Universal Kriging. Furthermore, as it is possible to see in *Appendix B*, the evolution of PM10 distribution is very different from PM2.5 one, indeed, it seems that the PM10 distribution periodicity is shorter than the PM2.5 one. It might be due to the fact that PM10 are way heavier than PM2.5 so they can drop down to the ground more easily than PM2.5. That is why PM2.5, which are very light, are more likely to stay in the air longer than PM10.
Finally, some Universal Kriging has been made for PM2.5 but the results were not very interesting since it was too close to PM10 distribution and it seems

that the error was too important. For all those reasons, it has been decided that **Simple Kriging for PM2.5 is the best way to make Kriging prediction for this study**.

2.3.3 Maximum altitude

One of the main goal of this study is to estimate the accumulated weight of PM2.5 in the air above Okayama city. In order to do that, it very important to know the maximum altitude for which PM2.5 can be found, indeed, as it has been explained before, the main idea is to multiply the predicted concentration of the points by a small elementary volume and obtain the weight of PM2.5 in the prediction area.
Also, according to Pr Yamakawa, it is possible to consider that above this altitude value, the PM2.5 concentration is equal to zero. That is why it is very important to figure out the maximum altitude of PM2.5 in the air. The previous maximum used was 500 m, it is a very high value but it was used in order to be sure that there is no mistakes and that all the PM2.5 are involved in the calculation. In order to evaluate this altitude, several variance analysis were made. Two types of variances were used:

- The variance : the regular statistical variance evaluate for predictions between an altitude of [0;50], [50,100], [100,150] until [400,450].
- The "global" variance : it is the regular statistical variance evaluate for prediction results between [0,450], [50,450], [100,450] until [400,450].

An Octave script with several if and for loops was written in order to make those analysis. This script is a little bit complex and since it is now possible to make predictions for 120 continuous hours, the evaluation of the variance takes several minutes, but once it is done, it is very convenient to analyze them on Excel or Octave and to plot the predictions results on QGIS.
Those analysis were made on the data from early April (Sat Sun April, 8-9 2017) and for the Yellow Sand period and the results were obtained very similar results. Given that Yellow Sand is a very unusual event, it proves that those results can be trusted and used.

Those results (*Figure 2.11*) shows that the altitude must be around 250 m, indeed, for the two types of variance, the value drops dramatically from this value. That is why it is possible to say that the weight of PM2.5 will be evaluate for a maximum altitude of 250 m and the predictions will be corrected by imposing to all the points above 250 m high to have a value of PM2.5 concentration equal to zero. This will also add an if loop in our script so it will take more time but our predictions will be more precise and close to the reality.

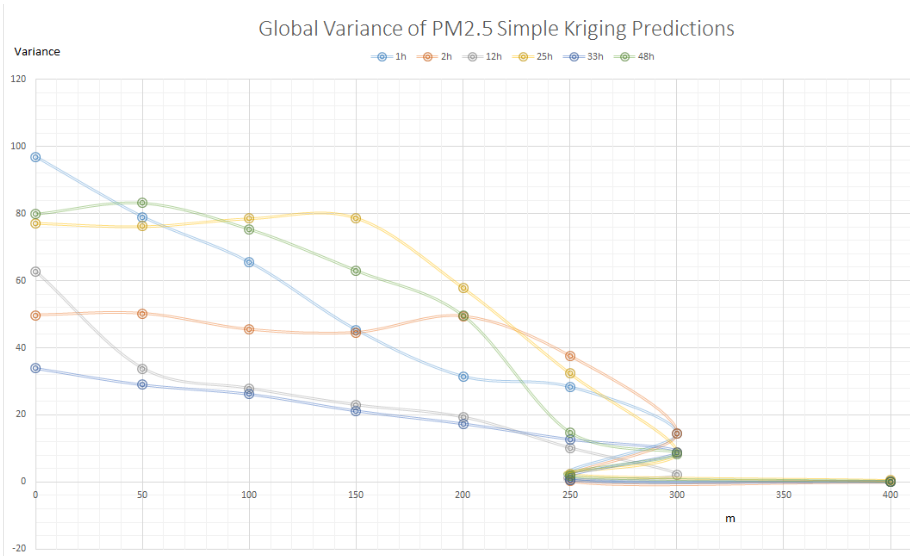


Figure 2.11: Global Variance of PM2.5 Simple Kriging Predictions during the Early April period (excel)

2.3.4 Geographical compression

It was also very interesting to highlight the effect of the geographical compression of the particles due to the altitude, that is why, thanks to QGIS, the altitude and the PM2.5 predicted concentration has been plot on the same map in order to see if the high concentrated areas are close to the bottom of reliefs (*Figure 2.12*). Night predictions has been used to make this analysis, indeed, it is interesting to study this compression phenomenon while the human activity is very calm in order to see the effect of geographical compression.

The red circles show the high concentrated areas and the light blue squares shows the high altitude ones. It it really obvious that there are high concentrated areas at the bottom of the reliefs. It is also interesting to notice that the results are similar for the other nights of the week, that is a clear proof of the geographical compression of PM2.5 due to the altitude. At night, it is clear that there are high PM2.5 concentrated areas at the bottom of those reliefs.

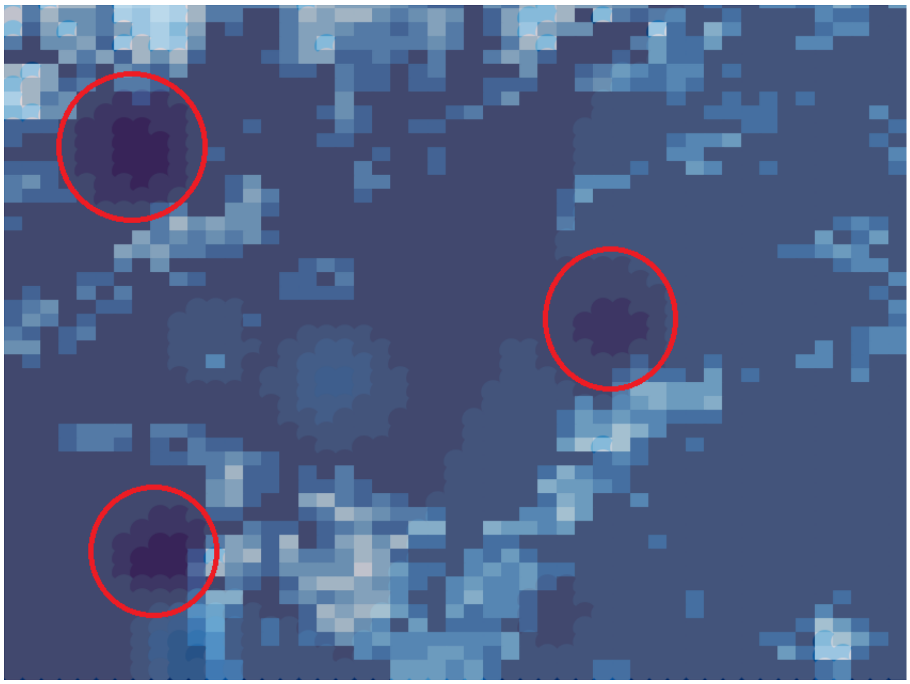


Figure 2.12: PM2.5 concentration in function of the altitude at 1 a.m on Monday July, 17 2017

Now that the procedure is clear and that some main problems have been solved, it is important to show some crucial results.

Chapter 3

Important results and discussions

3.1 Reference week

Once the script and the Kriging procedure made, the main task of this study was to collect 5 days continuous data of PM2.5 and PM10 concentration and make those analysis. The goal was to highlight a clear periodicity of the evolution of PM2.5 distribution in the air above Okayama. That is why, during several weeks (June, July, August), the work was to collect all the concentration data. Therefore, this period was a very complex one for Japan, indeed, the weather conditions were changing very fast and some special events such as typhoons, Yellow Sand or Oxidant have been experienced during this period. The main issue was to find a **Reference Week** with a relatively stable week and luckily, such a week happened on **July 24 to 28, 2017**.

As usual, it was very interesting to make Simple Kriging Predictions of the PM2.5 concentration above Okayama city. It has been decided that Simple Kriging is the best solution because PM2.5 and PM10 have different behavior, it is due to the difference of weight between those two types of PM.
The SK prediction of the accumulated weight of PM2.5 above Okayama city (*Figure 3.1*) shows a really clear periodicity of 40 - 45 hours, indeed, there is a clear peak every 40 -45 hours, that proves that this week might be a real reference for the study.

This very good results can be explain by the fact that last week was a very stable one, as it has been said in the introduction of this part, there was no dramatic change of the weather for this week so it is very interesting to study the behavior of PM2.5 during this period since it has been seen during this study that they highly depend on the weather. It shows also that the concentration value of PM2.5 above Okayama city is under the maximum concentration value

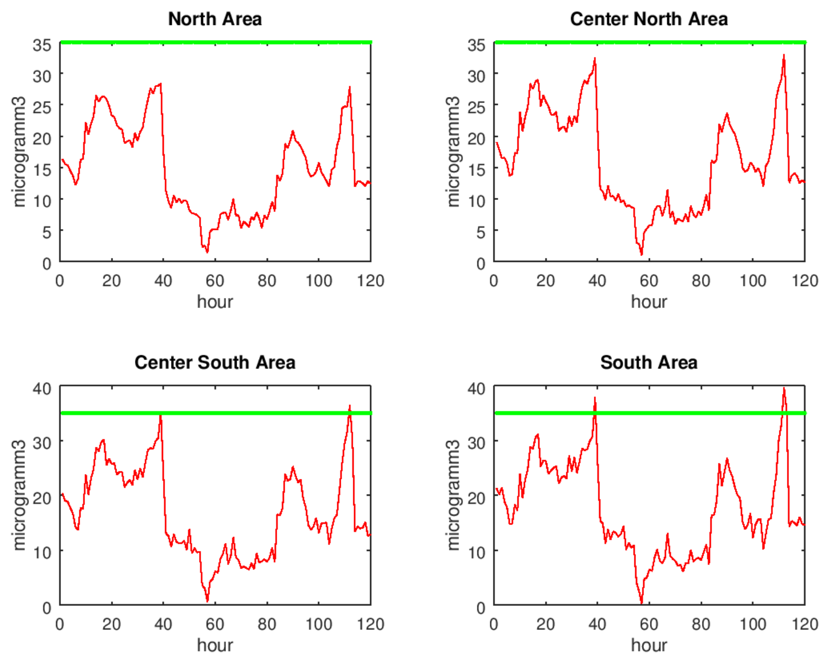


Figure 3.1: Average Predicted Concentration of PM2.5 for the 4 areas during 5 days : July 24 to 28, 2017

set by the EPA (green line), it proves that the air quality of Okayama city is acceptable but, during some periods, the PM2.5 average concentration is close to this maximum value, some measures have to be taken to protect the population. The evolution of the atmospheric pressure and of the humidity (*Appendix C*) shows that this week was a stable one: the atmospheric pressure was stable compare to the previous weeks and there was one raining period on Wednesday (60h) which corresponds to a drop of the atmospheric pressure and humidity at the same time.

Given that this week was a really interesting one, it is normal to hope that the correlation analysis will show some clear results this time. That is why some auto-correlation analysis of the predicted accumulated weight of PM2.5 above the 4 prediction areas has been made (*Figure 3.2*). Luckily, the results show a clear peak for 40 -45 hours proving once more that the periodicity of the PM2.5 distribution must be close to those values.

This auto-correlation result is crucial since it is the best proof of the PM2.5 distribution periodicity so far. That is why this week should be a real reference for the rest of the study.

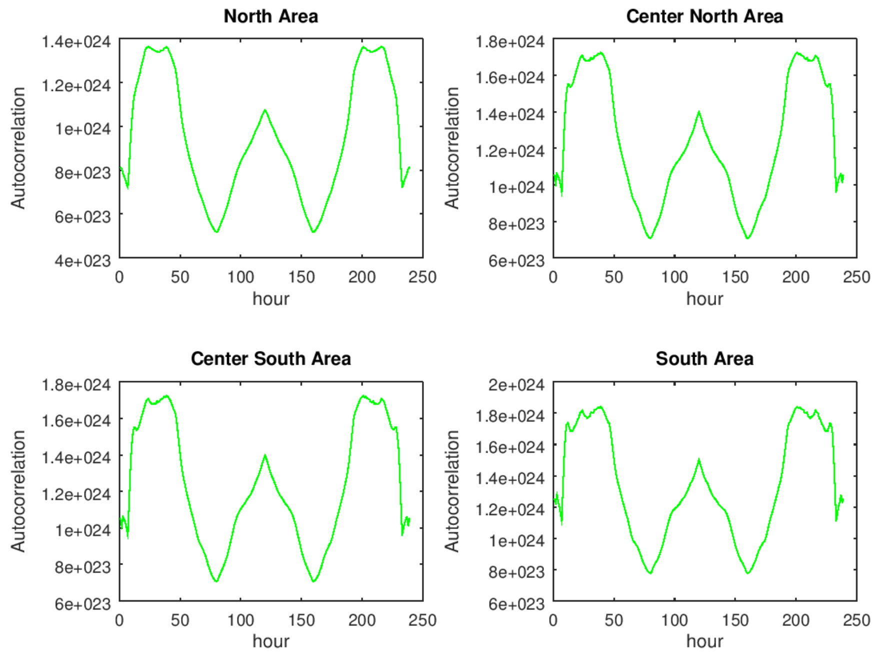


Figure 3.2: Auto-correlation of the PM2.5 predicted weight for 4 areas : July 24 to 28, 2017

Furthermore, it was important to study the error of prediction, that is why the evolution of the mean of the prediction variance during those 120 hours has been plotted and analyzed (*Figure 3.3)*. In this graph, it is possible to see that the value of the prediction variance is almost stable and does not show many dramatic change of it except for two hours (hour 85 and hour 115). Therefore, those two peaks of variance are very thin and the stability of the rest of the graph proves that the Kriging prediction of this week can be trusted. The value of the variance is pretty high, it is due to the fact that there are only 15 observatories for 2500 predicted points and also because it is a average value of the error for those points, but it is important to notice that the evolution is really stable, except for two strange peaks.

Finally, it is important to notice that this 40 - 45 hours periodicity value appears for the 4 small areas. This shows the that even if the variance value is relatively high (because it is an average one), this value of 40 - 45 hours of

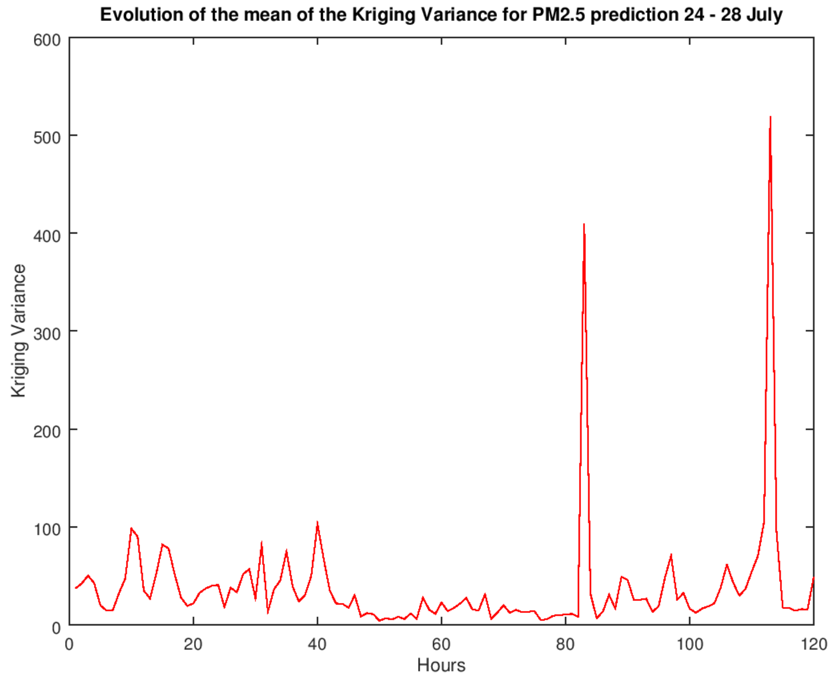


Figure 3.3: Mean of the Prediction Error for the Reference Week (Octave)

periodicity is very important and significant for the PM2.5 distribution above Okayama city.

This result is clearly the most important one of this study because as it has been explained earlier (part 2.3.2), the PM2.5 distribution really depends on the weather conditions because they are very light, indeed, the rain remains the most efficient way to drop down to the ground those particles. This result shows the behavior of those PM2.5 during a stable week so it is the "natural" behavior of those particles and as it has been foreseen thanks to the firsts obtained results, it is now clear that **the time dependent change of PM2.5 distribution above Okayama city is around 40 - 45 hours**. This result will be discuss in the part 3.3.

3.2 Specific Results

3.2.1 Traffic activity

Finally, this week, thanks to Pr Yamakawa, it was possible to collect some data about the traffic in Okayama city. It is data from 2015 that is why the idea was to estimate the average of several traffic data at different corners and estimate the shape of the daily traffic periodicity in the city and to compare it with the predicted weight of PM2.5.

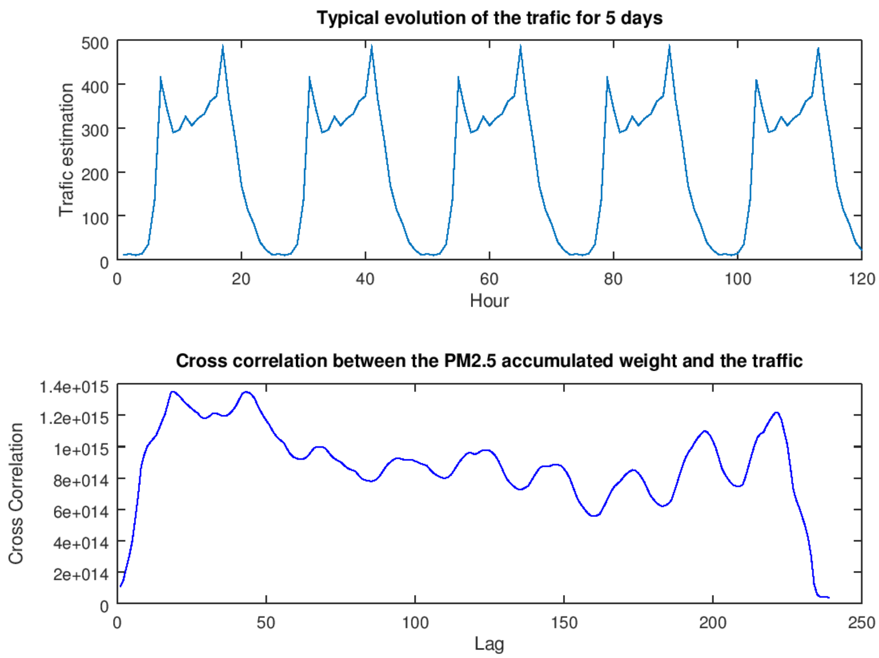


Figure 3.4: Influence of the traffic on the PM2.5 concentration : July 24 to 28, 2017

It is clear that when there is a peak of traffic in the city there is also a peak of PM2.5 concentration in the air. Furthermore, the cross-correlation between the traffic and the predicted weight shows that there is a clear relation between those two data, indeed, there are many peak, there is a peak of traffic in the morning between 8 am and 10 am and in the evening between 5 pm and 8 pm. There is a 12 hours periodicity of the traffic activity. So it is easy to conclude that the human traffic has a serious impact on the PM2.5 but since the PM2.5 are very light their own periodicity is different from the traffic periodicity. The PM stay a long time in the air and need hot air to go up and rain to drop to the ground. It is interesting to see that the PM2.5 distribution periodicity that has been highlighted earlier is longer than the 12 hours periodicity of the traffic activity. It must be due to the fact that those PM2.5 are so light and thin that they stay in the air for a longer time and they do not drop down to the ground fast enough.

3.2.2 Yellow Sand

The Yellow Sand last from the 6th of May to the 8th of 2017. This phenomenon comes from China. It is sand from Chinese industry that is brought in Japan by the wind.

> Yellow sand is a meteorological phenomenon which affects much of East Asia year round but especially during the spring months. The dust originates in the deserts of Mongolia, northern China and Kazakhstan where high-speed surface winds and intense dust storms kick up dense clouds of fine, dry soil particles.
>
> An analysis of Asian Dust clouds conducted in China in 2001 showed them to contain high concentrations of silicon (24 - 32 per cent), aluminum (5.9 - 7.4 per cent), calcium (6.2 - 12 per cent), and iron, numerous toxic substances were also present, although it is thought that heavier materials (such as poisonous mercury and cadmium from coal burning) settle out of the clouds closer to the origin. *Wiki*

This is a real issue in the country, indeed:

> Yellow sand from China reduced visibility in many parts of the nation Saturday and will continue to drift over the archipelago through Sunday afternoon, the Meteorological Agency said. *Japantimes*

The *Figure 3.5* show the mean of the predicted concentration of PM2.5 for each area during the Yellow Sand period. It is important to notice that the 2 first days of this period were a Saturday and a Sunday and the last one was a Monday, it clearly explains the increasing of PM2.5 weight during Monday morning, indeed, the human activity has a serious influence on its distribution. Therefore, it is very interesting to see that the value of the PM2.5 weight is a little higher than it is during other periods. This result is disturbing because it was expecting to have a very clear increasing of PM2.5 . Furthermore, the *Appendix D* shows that the error for those predictions is low and that proves that those results can be trusted. It is possible to say that the Yellow Sand phenomenon has no big effect on the PM2.5 distribution.

Therefore, there is no clear periodicity in those predictions, it might be due to the fact that the period is shorter (only 3 days) and also due to the Yellow Sand phenomenon.

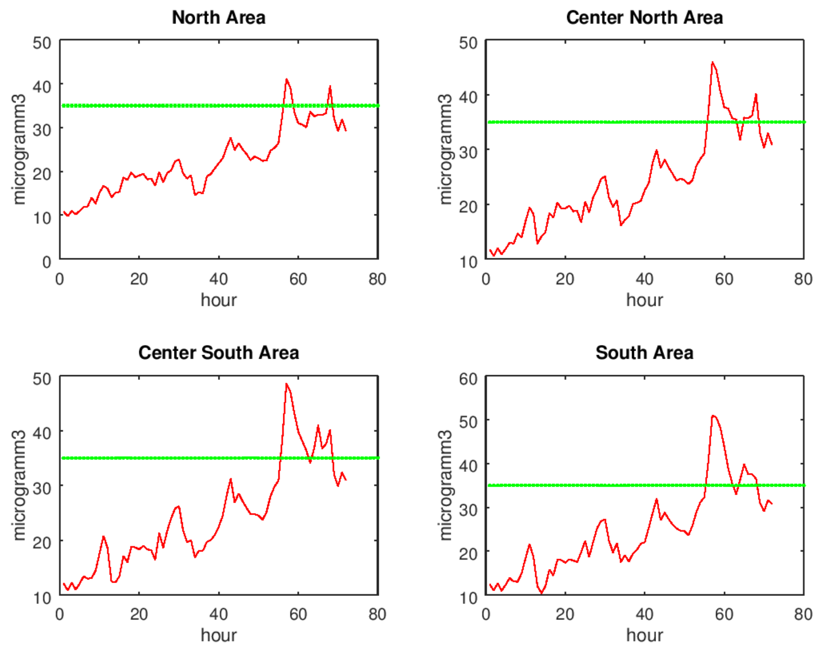


Figure 3.5: Simple Kriging Prediction (accumulated weight) during Yellow Sand

3.2.3 Typhoon

As said before, the rain is clearly the most efficient way to make particulate matters drop down to the ground. Some raining periods have been studied and it seems that it makes the concentration of PM2.5 dramatically decrease. Okayama prefecture experienced a Typhoon on : Early August (July 31st - August 04th, 2017). This is a real great opportunity to study the effect of a huge raining period on the PM2.5 distribution. The expected result is very clear: a dramatic drop of the PM2.5 concentration. The Typhoon appeared from Tuesday night to the end of the week.

The *Figure 3.6* shows the SK results for this period. The beginning of the week has a classic evolution but from Tuesday evening and night, the concentration of PM2.5 is dramatically decreasing to reach very low level (10 $\mu g/m^3$), very far from the maximum value of the EPA. This result shows the huge effect of the rain on those particulate matters.

Furthermore, it seems that there is no periodicity in this case, indeed, the autocorrelation (not in this report) does not show any clear peak and the SK result also proves it. It seems that the typhoon was so intense and long and because the natural periodicity of PM2.5 distribution is pretty high (40 -45 hours), the intense rain made all the PM2.5 drop down the ground very quickly.

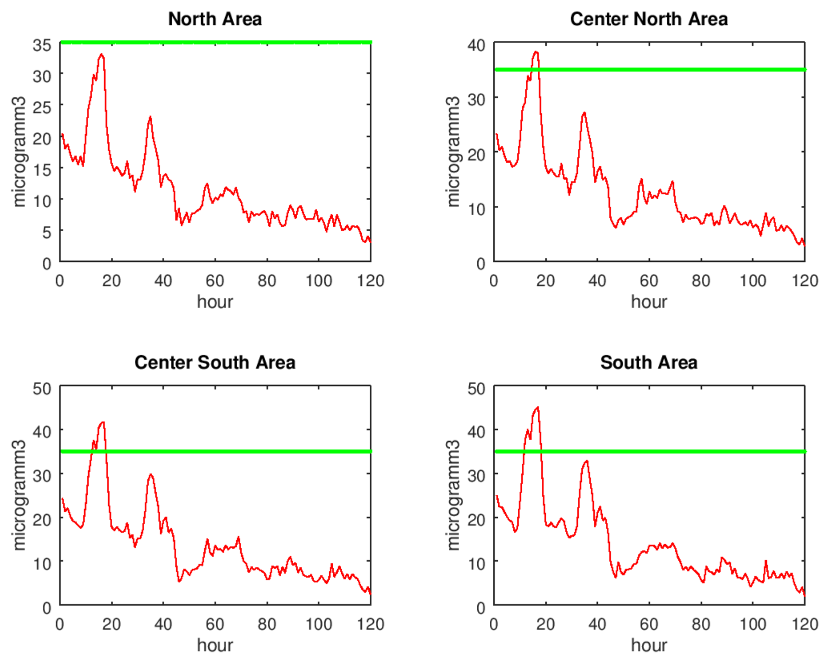


Figure 3.6: Simple Kriging Prediction during the Typhoon period

As expected, the analysis of this typhoon period highlighted that **the rain is the best way to decrease the PM2.5 concentration in the air**. This dramatic period of rain keeps the air quality very good but it does not allow to see the periodicity of PM2.5 distribution.

3.3 Electrochemical modeling

It is very interesting to notice the strong correlation between the evolution on PM2.5 concentration during this reference week and the evolution of the potential of batteries Li-ions with the so-called *Memory Effect*[10]. This effect can be seen when users start charging their electronic devices whereas the batteries are not completely down, the devices think that the 0 per cent of state of charge is reached although there is some energy left. Since the state of charge of the batteries is given by a cycling potential, it is possible to have an error of the estimation of the state of charge. This effect can be seen in the *Figure4.1*.

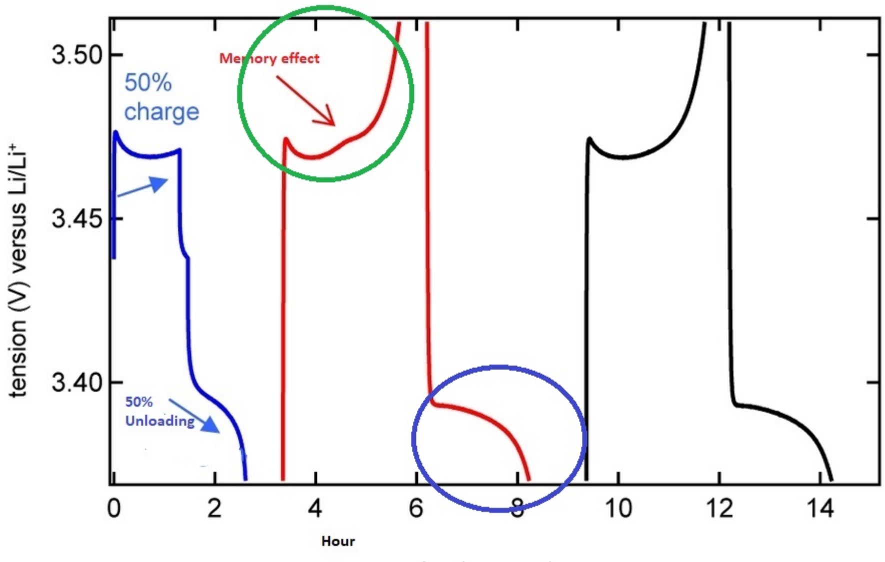


Figure 3.7: Electrochemical Memory effect

In the *Figure 4.2*, it is possible to see that the shape of the memory effect is the same of the charge of PM2.5 in the air (green circle) and that the drop of the PM2.5 concentration due to rain (blue circle) is the same as the unloading of the batteries. This correlation is very interested since it might be possible to talk about an analogy between the PM2.5 distribution and the life of batteries.

In the case of PM2.5, because of the traffic (12 hours of periodicity) and the fact that those PM2.5 stay in the air for a long time (it is due to their very low density), when the traffic starts again the air is filled by PM2.5 and they have no time to drop down to the ground before an other peak of traffic. It needs a raining period to drop the PM2.5 concentration in a significant way. This memory effect can also explain the periodicity of 40 - 45 hours for PM2.5 distribution.

Furthermore, the analogy between the PM2.5 distribution and electrical batteries can been seen like this:

- The 40 - 45 periodicity of PM2.5 distribution is the equivalent of the time of an electrochemical cycle for batteries.
- Peaks of pollution that increase the PM2.5 concentration during a cycle can be the equivalent the charge of batteries while it is not completely down.

- The drop of PM2.5 concentration due to rain can be the equivalent of a fall of the energy in the batteries due to the Memory Effect.

In order to prove this analogy and to use many results from the electrochemical field on the study of PM2.5 it is important to make several analysis. Indeed, in the future, it will be interesting to make those analysis in an other period (winter) to check if the results are similar. In plus, it will be crucial to see if the analogy can still be made in other regions in Japan or in the world (Grenoble ?) to make sure that this analogy is strong because Okayama has very specific geography and weather conditions. It is also very important to find with which parameter(s) of an electrochemical system the weather conditions will be linked with since it has been seen that PM2.5 distribution really depend on those weather conditions.

More proofs of this analogy will be very useful, indeed, the theory of an electrochemical is already well known and referenced and it will be very interesting to use this knowledge for the understanding of PM2.5 distribution in the air. Some theoretical and experimental works should be done to make such analysis.

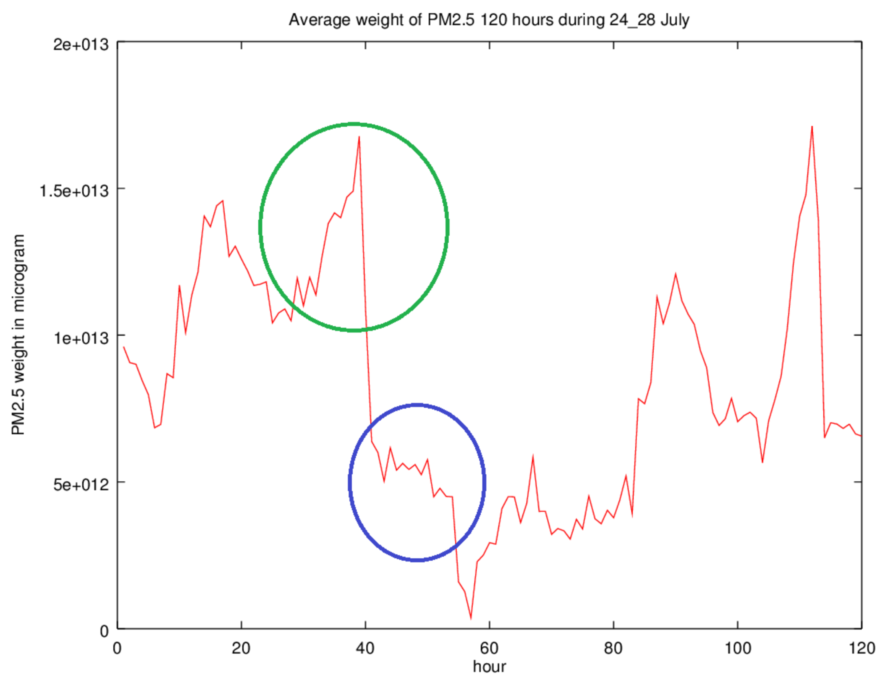


Figure 3.8: PM2.5 distribution during the reference week

Of course, this analogy is just an hypothesis but it is interesting to notice the similarities between those two phenomena and to highlight it. This analogy does not explain the behavior of PM2.5 during specific periods but it will be very interesting to study this analogy in the future.

3.4 Conclusions, discussion and prospective

As said before, the result of 40 - 45 hours as PM2.5 distribution periodicity is very disturbing, indeed, it was expected to be closer to the traffic activity periodicity which is 12 hours. The variance of the Kriging and the fact that this periodicity has been highlighted with average values allow to criticize this result.

Therefore, there are some arguments that prove that this result must be very close to the reality and even if it is not a very precise one, it is accurate enough to have an idea of the general PM2.5 distribution periodicity above Okayama city. Indeed, as it has been explained before, this result has been seen for several weeks and analysis during this study.
The fact that the result is the same for each area is also an other proof that it might be the good one since it is very unlikely that there is the same result for 4 areas just by chance. This was the main proof that this result can be trusted.
Finally, it is very interesting to analyze the auto-correlation of the PM2.5 concentration at several observatories to see if the value of 40 - 45 hours is close to the reality (*Figure3.9*). It seems that there is always a peak of auto-correlation between 40 - 45 hours, it is nos the highest peak for every observatories but it is always here. This also proves that even if the final result is evaluated tanks to average values, it is trust-able and gives a real idea of the time dependent change of PM2.5 distribution above Okayama city.

Even if this result is not the most accurate one due to the Kriging method and the use of average values to make conclusions, it is clear that 40 - 45 hours is a significant value for the evolution of PM2.5 distribution. This result is surprisingly different from 12 or 24 hours, it might be due to the fact that PM2.5 are very light and can stay very long in the air. They really depend on the weather conditions and on the traffic activity, the combination of those parameters can explain this surprising result.

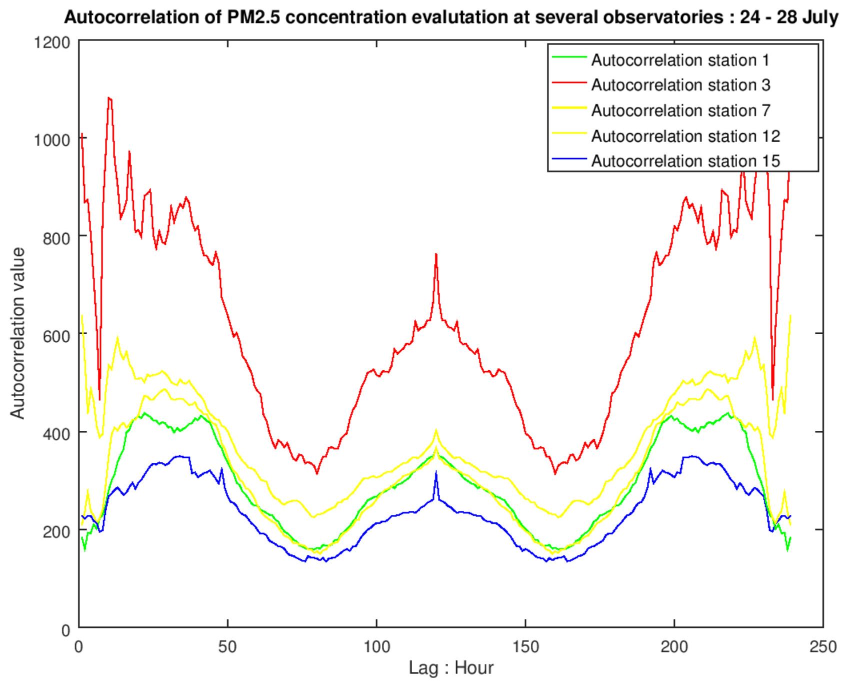


Figure 3.9: Auto-correlation for observatories values

Finally, the main perspective of this study might be the analogy between PM2.5 distribution and the memory effect of electrochemical batteries. This interesting analogy should be analyzed more deeply in the future and as it has been said before, the study should use data from an other period and also from other regions since Okayama has a very specific topography.

Chapter 4

Work organization

Now that the scientific aspect of this internship has been explained, it is important to present the organization of the work in Japan and in this laboratory.

4.1 Gantt diagram

First of all, the Gantt diagram is a very useful tool to explain the time management of a project. In that case, this diagram (*Figure 4.1*) shows the main tasks that I had to achieve during this internship.
As it is explained is the Gantt diagram, the first 3 weeks were discovery weeks, indeed, the main goal of my supervisor was that I became familiar with the Kriging process, some new software (QGIS, gstat.exe, Power BI, Google earth). Once this preliminary work done, I could make some Kriging predictions using the classical method. Then, the main task of this internship could start, it was to write an Octave script that would allow us to make statistical analysis and predictions for several hours. It was necessary to use the STK package, therefore, this package was completely new for me and for Pr Yamakawa. It took me almost two weeks to write a simple script that could make all the calculation for the predictions. Since it was new for me and that such package is like a black box, it was very complicated to understand how it works at the beginning, then after several tries and having study examples, I could make a simple script that makes Kriging predictions. Then, the main task was to improve this script so that it could make more statistical analysis and gives more results. At the same time, I was always collecting more concentration data and make the predictions every week. Unfortunately, when the script was ready, the raining season was about to start so we had to wait several weeks to have a stable week and have reference results. Nevertheless, this long period was the opportunity to improve the script and to study the air pollution during some specific period such as Yellow Sand or very raining period. Finally, the last weeks of this internship were used to analyze more precisely the reference week and to make useful figures for the understanding of this very report.

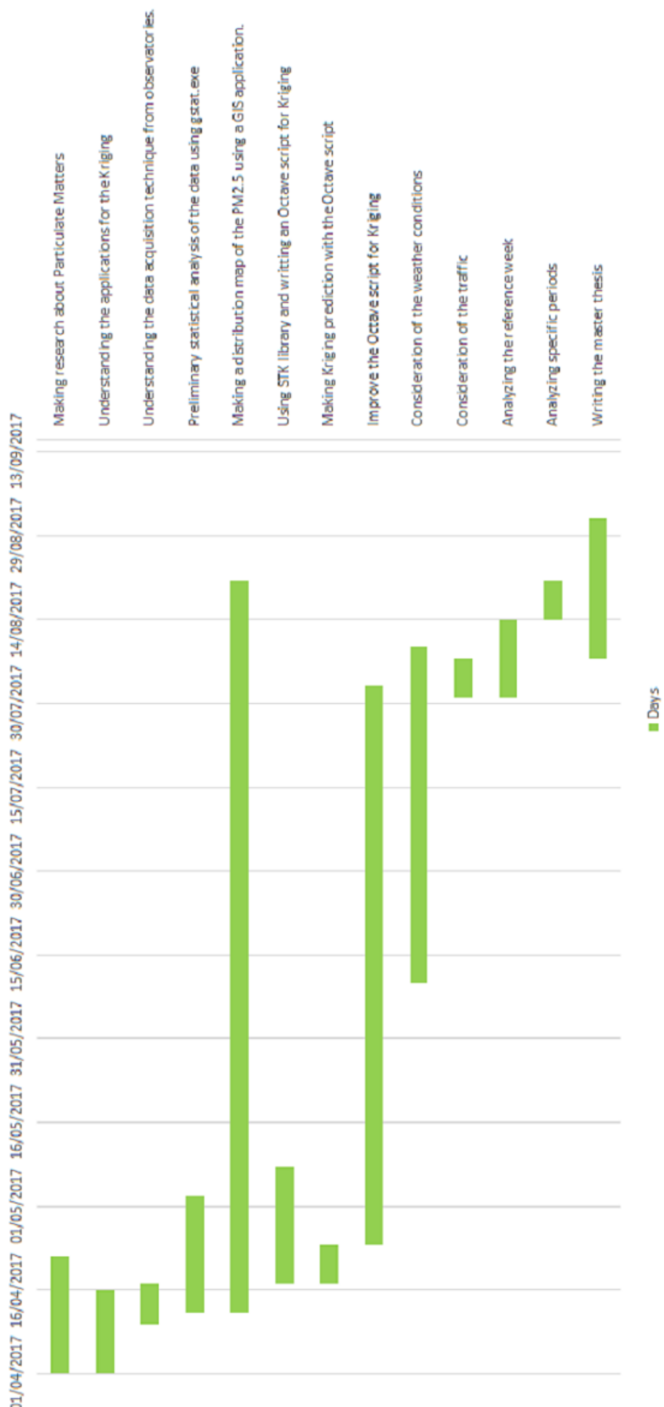


Figure 4.1: Gantt diagram of this internship

4.2 Organization

This internship organization was very simple, indeed, there were two daily meeting, one at 9 am and another at 1.30 pm. The first one was made to show the results from last day and to give the directions for the day whereas the second one was mostly to check the results and to see if everything is going fine. The most important thing was the Free Management of the work, indeed, after those two meetings, I was free in the laboratory so I could manage my time the way I wanted. I could stay or leave the laboratory whenever I want as long as the work was done. In my opinion this was a good organization since I was responsible of my time but I also had some deadlines and it shows that Pr Yamakawa trusted me on my work.

Furthermore, there was a weekly oral presentation of 5 or 10 minutes to explain the method, the most important results or issues of this week and to prepare the directions for the next one. This was a very important moment because it showed the main achievement of the week and it was a good way to end it.

Finally, in addition of the weekly presentation, I had to make a small weekly report (between 4 and 6 pages) that sums up my work. It was the opportunity to show all the important figures and to explain them. That was very important to have a real state of advancement of the study. In *Appendix E* there is one example of such a report.

General Conclusion

Particulate matters are a real issue for developed societies and especially for crowded cities. It has been explained that those PM could have terrible effects on human health and that is why it was necessary to study their evolution and distribution in order to be able to foresee high peaks of pollution. The use of geostatistical methods for such analysis is new but it was very interesting and efficient. The Kriging appeared to be very interesting and helpful to make PM2.5 concentration predictions. It took several weeks to understand this new method and to learn how to use it.

The main mission of this internship was to create an Octave script using the STK library that would be able to make Kriging predictions for several continuous days and allow to make statistical analysis to highlight a periodicity in the PM2.5 distribution above Okayama city. This mission took several weeks also since this STK-package was totally new for this laboratory. There is now a clear and very efficient script that achieve all those goals.

Before making clear conclusions, it was very important to answer to two very crucial questions. First of all, a choice between UK using PM10 SK results and only SK had to be made. After several tries and variograms comparisons, it has been decided to use SK for the predictions. Then it was crucial to set the prediction area, it has been decided to reduce it to an area with many observatories in order to reduce the Kriging variance and have more accurate results. This area was divided in 4 equal areas with different human activities.

Then, it was possible to make predictions for 5 continuous days (from Monday to Friday because those days are more significant). During several weeks, PM2.5 concentration data were acquired and Kriging and analysis were made. Unfortunately, the weather conditions were changing very fast and it appeared that PM2.5 distribution really depends on the weather so it was complicated to highlight a clear periodicity. Some specific weather periods like Yellow Sand, a Typhoon or the raining season were experienced and it was interesting to notice that rain is definitely the best way drop down to the ground PM2.5 and make the amount of PM2.5 in the air really decreases. It also seems that the PM2.5 concentration is usually under the EPA maximum value of 35 $\mu g/m^3$, so the air quality in the prediction area is acceptable. The influence of the traffic activity on the PM2.5 distribution was also highlighted.

Luckily, in late July a very stable week was experienced, indeed there were no dramatic change of the weather conditions, and it was possible to set this week as a reference week. After Kriging predictions and statistical analysis the value of 40 - 45 hours periodicity that had been seen in the previous weeks was very clear for this one and the fact that it was the same for the 4 divided areas proved that this value is very significant for PM2.5 distribution. This result is very disturbing because it was expected that the PM2.5 periodicity would have been close the the traffic activity one (12 hours) but it seems that those particulate matters are so light that they stay in the air for a long time before dropping down to the ground. The value of 40 - 45 seems to be very important for the PM2.5 evolution and knowing that, the authorities can now take measures to reduce the traffic activity in some high concentrated region (like the Center North one) at some periods to reduce the risks of exposure.

Furthermore, it has been very interesting to notice the correlation between the PM2.5 distribution and the electrochemical cycle of batteries. This analogy can been very useful to understand better the pollution in the air and to explain this 40 - 45 hours periodicity. It will be important to make more analysis to prove this analogy and to formalize it.

Eventually, this internship was also the opportunity to discover a new country and a new culture. The Japanese way of work is very different from the French one and this experience was very interesting and will clearly be important in the future.

Bibliography

[1] Xavier Querol, *PM10 and PM2.5 source apportionment in the Barcelona Metropolitan area, Catalonia, Spain*, Atmospheric Environment, 2001

[2] Christoph Hueglina, *Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland*, Atmospheric Environment, 2004

[3] Barbara J. Turpin and Ho-Jin Lim, *Species Contributions to PM2.5 Mass Concentrations: Revisiting Common Assumptions for Estimating Organic Mass*, Aerosol Science and Technology, 2001

[4] J.S Bailly, *Geostatisque applique. Chapitre 1 : Analyse et modellisation de structures spatiales*, UMR TETIS - ENGREF

[5] Ecole Polytechnique Montral, *Gostatistique*, Ecole Polytechnique Montral

[6] https://www.mailman.columbia.edu/research/population-health-methods/kriging

[7] Michael Jerrett, *Spatial Analysis of Air Pollution and Mortality in Los Angeles*, Epidemiology 2005

[8] Kevin Fong-Rey Liu, Ming-Jui Hung, Jong-Yih Kuo, Han-Hsi Liang, *Using GIS and Kriging to Analyze the Spatial Distributions of the Health Risk of Indoor Air Pollution*, Journal of Geoscience and Environment Protection, 2015, 3, 20-25

[9] Bernardo S. Beckerman, *A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in the Contiguous United States*, Environmental Science and Technology, 2013, 47, 72337241

[10] https://www.psi.ch/media/un-effet-memoire-decouvert-egalement-dans-les-batteries-li-ion

Appendix A

Prediction Area

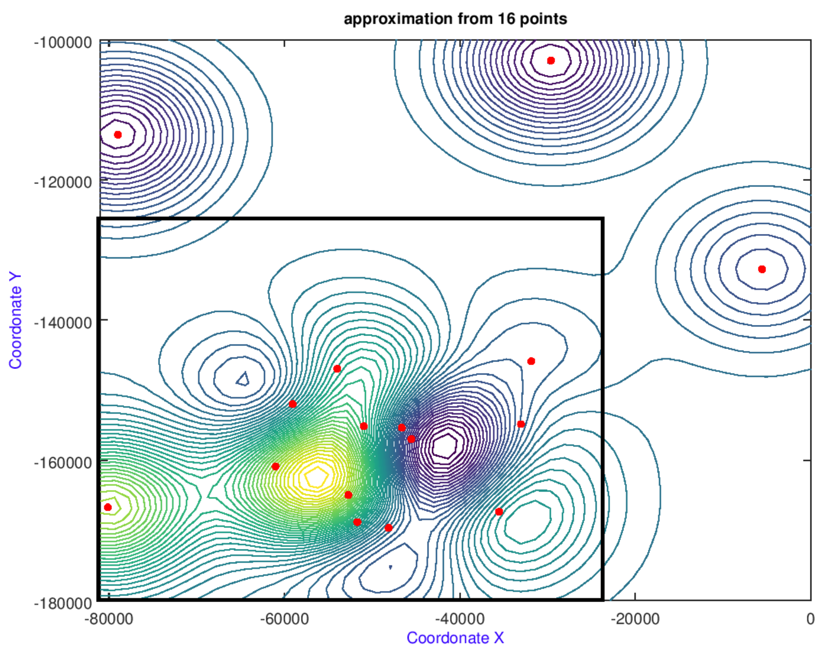


Figure A.1: Difference between the big prediction area and the small final (Octave GNU)

Appendix B

Universal Kriging or Simple Kriging ?

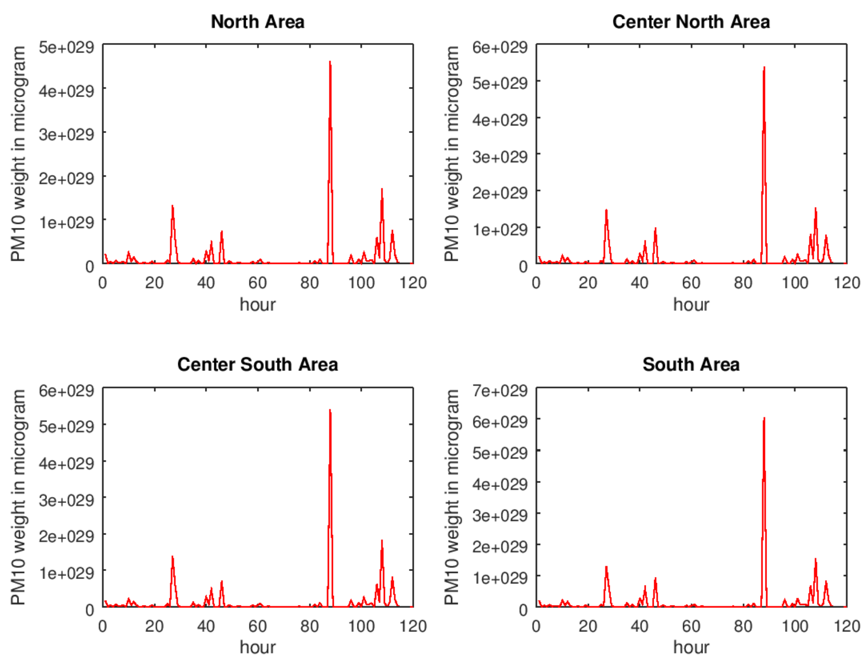


Figure B.1: Predicted accumulated weight of PM10 : July 24 to 28, 2017

Appendix C

Weather Conditions during the reference week

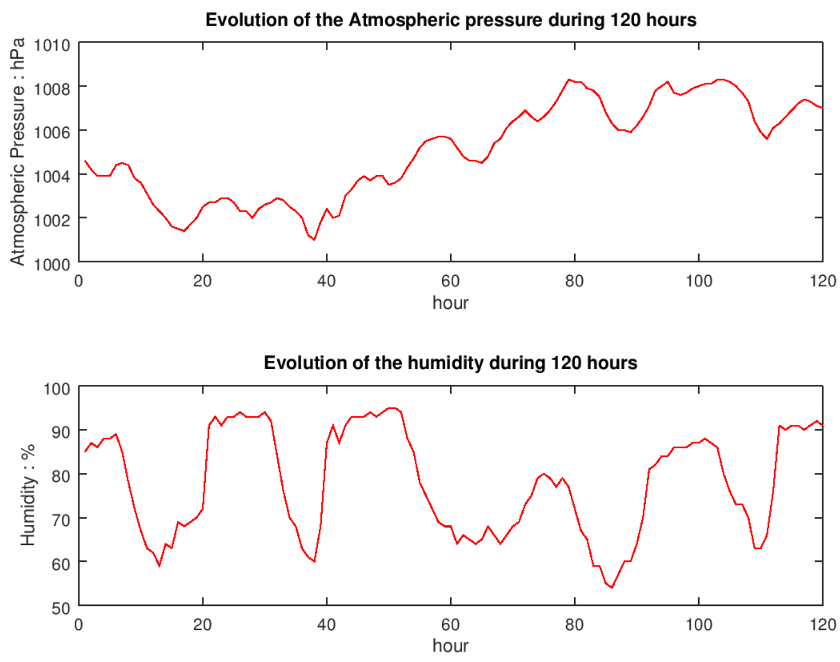


Figure C.1: Evolution of the weather conditions during 5 days : July 24 to 28, 2017

Appendix D

Yellow Sand

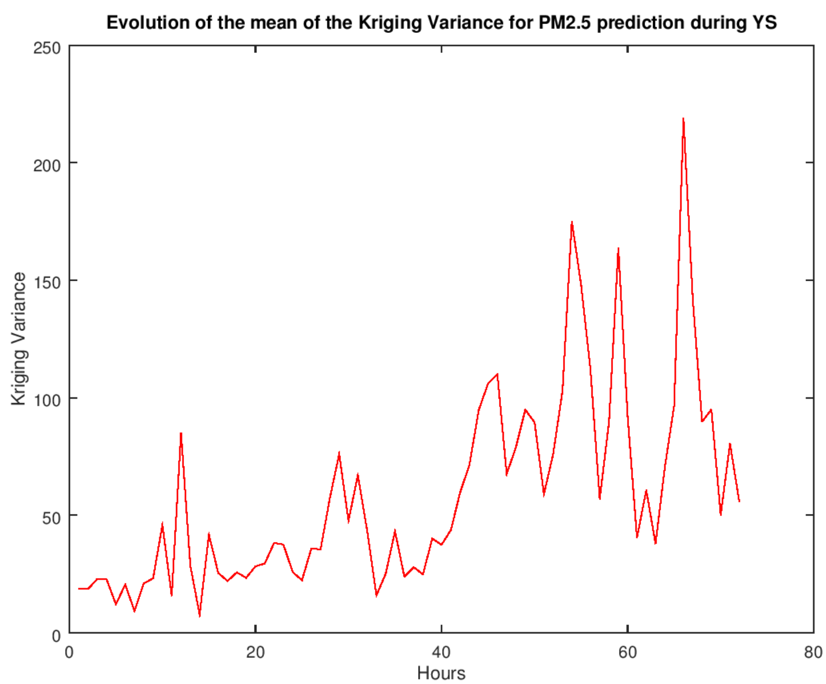


Figure D.1: Mean of the Prediction Error for Yellow Sand period prediction (Octave)

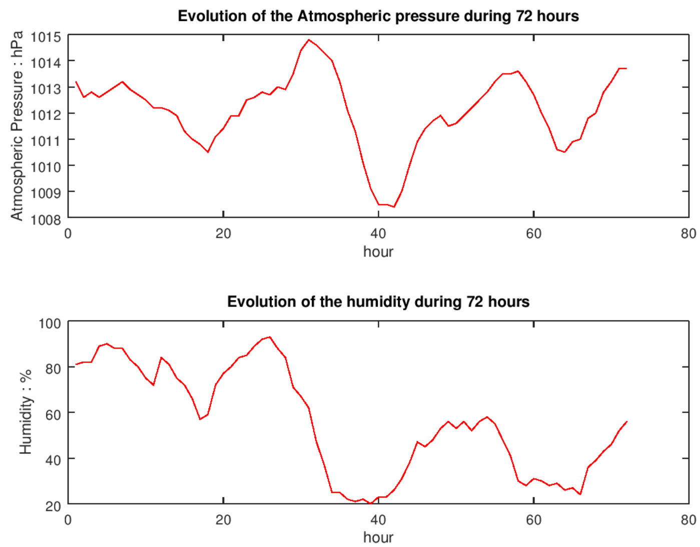


Figure D.2: Weather conditions during Yellow Sand

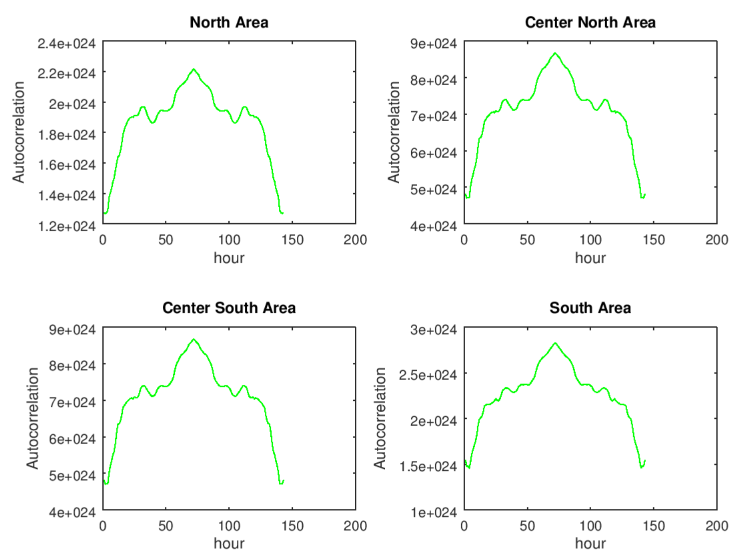


Figure D.3: Auto-correlation of the Kriging Prediction during Yellow Sand

Appendix E

Example of Weekly Report

Weekly report : 15/20 : July 17 to 21, 2017 : Kriging estimation of the time dependent distribution of PM2.5 above Okayama city, Japan

Ariel Chiche*
The University of Okayama, Japan

July 21, 2017

List of Figures

1 Introduction

This week main object was to collect more data from last week : PM2.5 and PM10 concentration but also the Atmospheric Pressure and the humidity. Given that we have a clear script to make our predictions and statistical analysis, we had more time to study our results and make hypothesis about them. We are expecting very similar results between this week : July 10 to 14, 2017 and the early July week we studied last week, indeed, if the results are very similar it will be easier for us to make conclusions about the time dependent change of PM2.5 distribution in the air above Okayama city. We already know that the PM2.5 distribution is really depending on the weather and especially on the rain : atmospheric pressure and humidity.

Therefore, the week we are studying is very specific, indeed, it was the "'oxidant"' period this week. This natural phenomenon is likely to increase the PM concentration in the air. It will be interesting to study the air pollution during this period but we think that, unfortunately we will have some differences with the last week results.

*Thanks to Yu Kurozumi and Pr Yamakawa

2 Predictions and evolution of the weather conditions

As usual, the first task was to make Simple Kriging Predictions of PM2.5 and PM10. Since we have proved in the previous weeks that Universal Kriging (using PM10 predictions) for PM2.5 is not the right solution for our predictions, we will only show and discuss about the PM2.5.
Last week weather was really hot and wet with a very little wind. It was raining so we can a dramatic drop of PM2.5 weight during the raining time (*Figure 1*). Unfortunately, the beginning of the week does not show us a 40 or 45 hours periodicity, we think it is due to the change of weather. Luckily, we can see this 40/45 hours periodicity between two peaks at the end of the week but we can not conclude yet about the time dependent change of PM2.5 distribution.

For now, the only conclusion we can make is that PM2.5 are so light that weather conditions have a really serious impact on their distribution. We can see that it is when the humidity rain is low and the atmospheric pressure is just getting important, so at the end at a raining period, that the PM2.5 weight is the lowest.

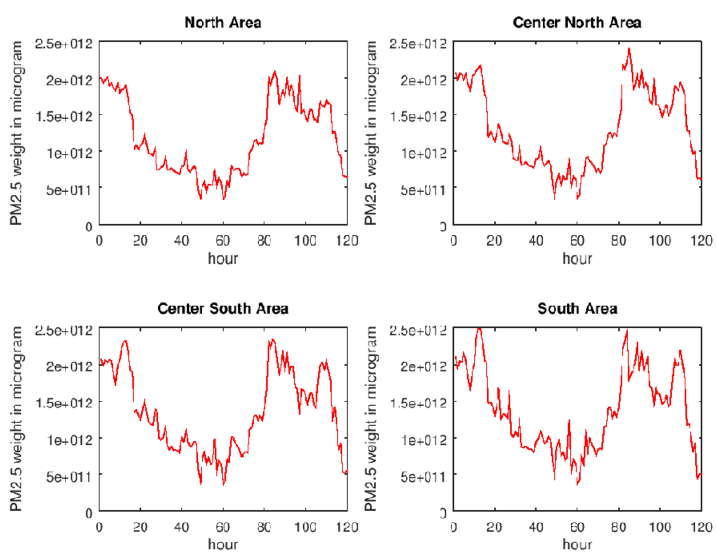


Figure 1: Predicted Accumulated weight of PM2.5 during 5 days : July 10 to 14, 2017

It is also interesting to study the evolution of the Atmospheric pressure and the humidity during this period (*Figure 2*). We can see that the humidity has a really clear 24 hours periodicity. The dramatic drop of Atmospheric pressure correspond to the beginning of the rain. The atmospheric pressure also has a clear 24 hours periodicity. As we said before, this phenomenon is clearly the most important to decrease the concentration of PM2.5. Making correlation analysis might help us having clear results for PM2.5 distribution.

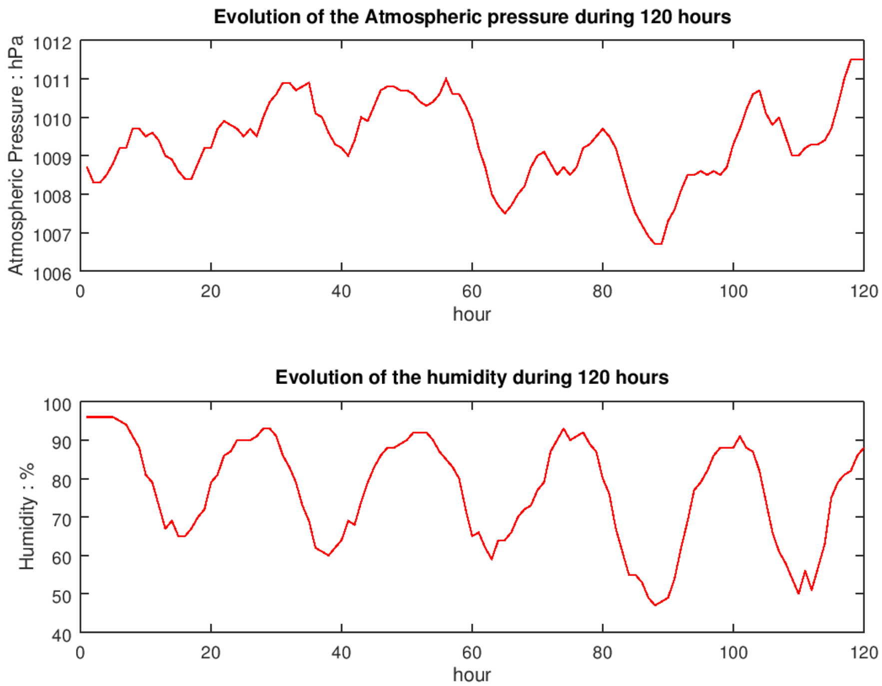


Figure 2: Evolution of the weather conditions during 5 days : July 10 to 14, 2017

3 Correlation analysis

As usual, we hope that auto-correlation and cross-correlation analysis will help us determine the time dependent change of PM2.5 distribution and show us a clear periodicity.

At first, it was really interesting to make cross-corrolation analysis between the predicted weight of PM2.5 from last week (July, 10-14) and the early July week (*Figure 3*). The idea is to look for similarities between our two predictions. Unfortunately, given that the weather between those two weeks was really different, indeed, the early July week was the opportunity to study a typhoon and its influence on PM2.5 whereas, last week was really hot and wet with a very changing weather.

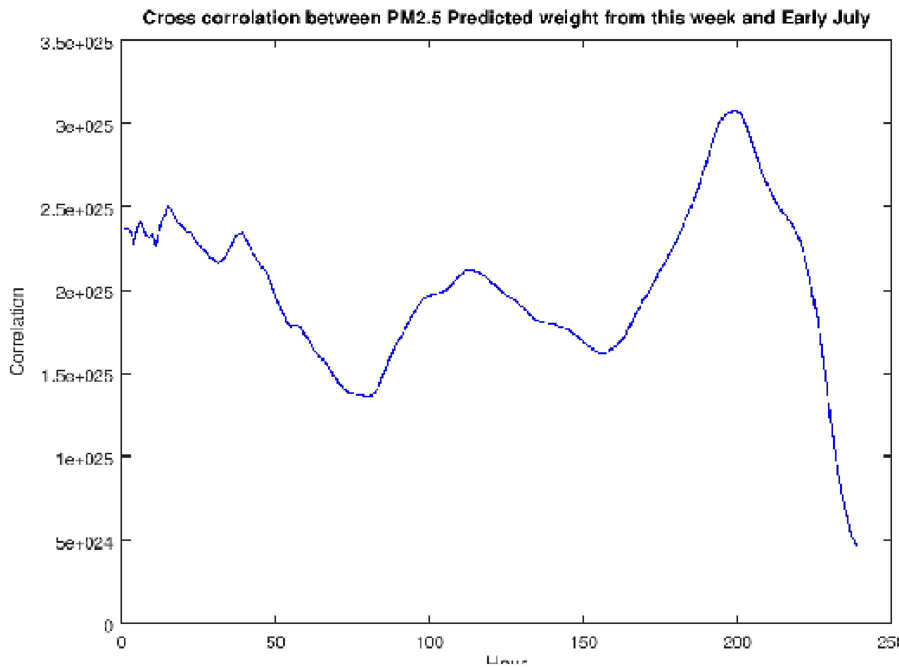


Figure 3: Cross-correlation between predictions for two weeks

As we can see in the figure, the cross-correlation rate is very high but the peak of correlation is very late : 200 hours lag, which is very far from the 40 45 hours periodicity we could determine last week. This is clearly due to the different conditions of atmospheric pressure and humidity change between those two weeks.
Therefore, there is a little peak at this time but this is not the highest peak. This might be a new clue that proves that the time dependent change of PM2.5 distribution is around 40 and 45 hours.

Finally, we have made some cross-correlation analysis between our PM2.5 predicted accumulated weight and the atmospheric pressure and the humidity (*Figure 4*). We want to show again the important influence of those two parameters on the PM2.5 distribution. We can see that the cross-correlation value is very high and we have a peak of correlation for a 200 hours lag for those two parameters. This might be due to the "‘oxidant"’ phenomenon, indeed, it was at the end of the week and it is likely to increase the amount of PM2.5 in the air. Unfortunately, in those analysis, we can not see a 40 or 45 hours periodicity, we need to make more analysis and to have a more stable week.

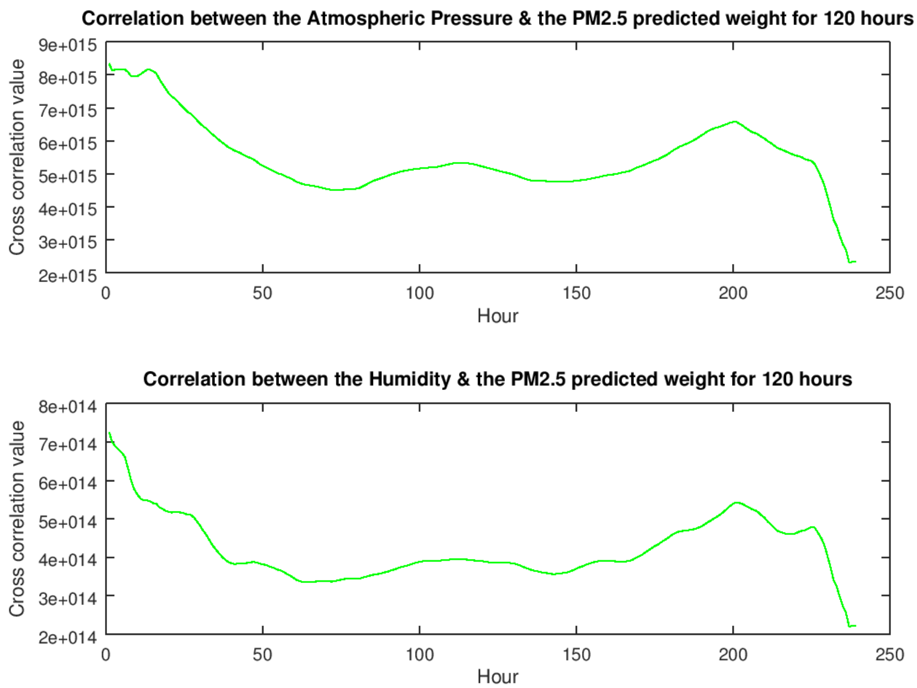


Figure 4: Cross-correlation between predicted weight and weather conditions

4 Conclusion

This week was also a very interesting one, we were able to collect a lots of data and make several analysis. Unfortunately, we can not make really clear conclusion about the time dependent change of PM2.5 yet but we have an efficient procedure that allows us to make many analysis and gives us time to understand all the phenomena that are involved in PM2.5 distribution.

For now, we can say that PM2.5 are so light that rain is the only way to make those PM fell down to the ground, indeed, gravity power is way too weak on those PM. It is also clear that it is just at the end of a raining period that the accumulated weight of PM2.5 is the lowest in the air above Okayama. The newt weeks will be important because we expect to have more stable weather conditions and to have a clear periodicity.

Last but not least, we will soon start to write a report for this internship (for Grenoble INP - PHELMA) and also an article to present our results and conclusions.