Validation of a geostatistical interpolation model using measurement of particulate matter concentration

Lucie Bonvalet[1], Andrée Marijns[1], Benoît Debast[2], Olivier Parks[2], Guy Gérard[3], Grégory Coussement[1]and Charles Passelecq[1]

[1]University of Mons, Faculty of Engineering, Rue du Joncquois 53, 7000 Mons, Belgium
[2]AWAC, Walloon air and climate agency, Rue du Masuis jambois 5, 5100 Jambes, Belgium
[3]ISSeP, Scientific institute of public services, Rue du Chéra 200, 4000 Liège, Belgium

Abstract:
As the air quality has a strong impact on public health, its control is necessary. Several governmental organisations are involved in the ambient air quality control at regional scales. For that reason, the numerical modelling of pollutant transport and diffusion is a research subject of primary importance.

The aim of the paper is to present the validation of a geostatistical interpolation model. This model has been developed and is applied to auto-correlate and estimate continuously the concentration of particulate matter (PM) and gas in the air. The validation was done on particulate matter concentration by comparing measurements with interpolation results.

Due to the duration of transport and diffusion phenomena, the longer the period of comparison, the better the correlation. For this reason, the comparisons were done on daily average and moving average twenty-four hours. For both PM_{10} and $PM_{2.5}$, respectively particles with diameter smaller than 10µm and 2.5µm, the correlation coefficients computed on the orthogonal linear regression are above 0.9.

The *European commission working group on guidance for the demonstration of equivalence* has defined a criterion to validate the data supplied by two measurement instruments. This criterion, consisting of an uncertainty smaller than $2.5\mu g.m^{-3}$, is used to compare the measurements with interpolation results.

The comparison was based on three criteria: uncertainty smaller than $2.5\mu g.m^{-3}$, correlation coefficients close to one and, the difference of the mean of the twenty-four hours moving average between interpolation results and measurements taken during a three months campaign.

Based on these three criteria, the ordinary kriging, which is the geostatistical interpolation model used, is validated for this application within the auto-correlation and estimation of the particulate matter and gas concentration in the air.

Key words: *particulate matter, interpolation, validation, geostatistics*

PAPER PURPOSE

In Belgium, the control of ambient air quality and the alert procedure, except for ozone alerts, are regionalised responsibilities. This paper aims to present a geostatistical interpolation model developed and applied to auto correlate and evaluate continuously the concentration of particulate matter and gas in the air in Wallonia, see Figure 1. This paper describes the method and the way it has been validated on particulate matter concentration by a measurement campaign.

Figure 1: Location of Wallonia in Europe (a) Belgium in Europe, (b) Wallonia in Belgium

INTRODUCTION

Because of the impact on public health (Commission Fédérale de l'Hygiène de l'air, 2007), control of ambient air quality is necessary. It includes the measurement of both particulate matter (PM) and gas concentrations. In Belgium, the agency in charge of this control has also the duty to inform and warn the

public about it. This information is communicated interactively through a website (ISSeP, 2004). The concentration of PM and gas is controlled by a network of twenty-three air quality measurement stations now but fourteen stations when the validation campaign was performed.

The definition of PM_{10} given in the directive of the European Parliament (European Parliament and Council, 2004) is the following : "`PM_{10}' means particulate matter, which passes through a size-selective inlet as defined in EN 12341 with a 50% efficiency cut-off at 10µm aerodynamic diameter;". Accordingly, to determine the concentration of PM in the air, an optical counting method consisting of diffraction laser sensors is used. These measured concentrations are integrated every thirty minutes. The metrological error is linked to a density fixed during the calibration.

To evaluate continuously and reliably the concentrations all over the region, an appropriate interpolation model should be used. Interpolation by geostatistical approach has been chosen, more precisely the ordinary kriging (C. Passelecq and C. Demuth, 1999) model presented in this paper. European regulations concerning the concentration of PM in the air are given for daily means. Therefore, to control the air quality according to these regulations, the results given by a daily mean are the most appropriate.

GEOSTATISTICAL INTERPOLATION MODEL

In 1999, four interpolation models were studied (C. Passelecq and C. Demuth, 1999) on the concentrations of O_3 and SO_2. These models were a method of inverse distance weighting, a method based on Delaunay triangulation, a method of thin plate spline functions and finally the kriging method. The study showed two main strengths of the kriging. Firstly, it makes it possible to get the spatial correlation between the PM concentration measurements at different locations. Secondly it gives an estimation of interpolation error. According to the results of this study, kriging is an appropriate model.

The ordinary kriging is an interpolation model based on a geostatistical approach (P.S.N. Murthy, T. Arora and S. Ahmed, 2008). In this approach, the difference with a stochastic approach is that the spatial auto-correlation between two neighbouring values is taken into account. This auto-correlation is taken into account by using a variogram.

Model principle

The ordinary kriging model consists in weighting the measures according to the distance between two measurement stations. The station locations for measurement of PM concentration are shown in

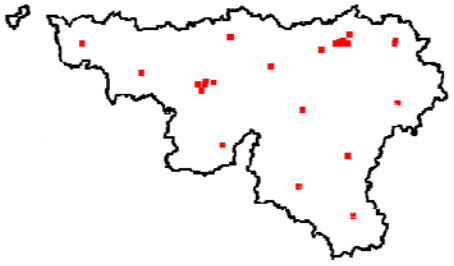

Figure 2: Locations of PM concentration measurement stations

This weighting is performed using a variogram. The determination of the variogram is based on the measurement during one year and a half of PM_{10} and $PM_{2.5}$ concentration, particles with diameter less than 10µm and 2.5µm.

The covariance is computed for each couple of values $z(x_i)$, $z(x_j)$ where z is the concentration of PM and x_i, x_j the location of the measurement stations i and j.

The variogram, γ, is computed from the covariance of station locations, C, (AWAC, 2004), Equation (1), where $h= |x_i - x_j|$ is the distance between the two stations i and j.

$$\gamma(h) = C(0) - C(h) \quad (1)$$

The range of the variogram is the specific distance h_s for which $\gamma(h_s) = C(0)$ i.e. $C(h_s) = 0$. It represents the limit of the spatial correlation between two measurement stations. As for both PM_{10} and $PM_{2.5}$ the range, 271 and 312 km, is higher than the upper distance between two measurement stations, 148 km, then the limit of the spatial correlation is never reached.

APPLICATION

Pollution episodes
Alert procedure activation (AWAC, 2004) in Belgium is based on objective air quality data. These data are the measurements of pollutant concentration. Both the measured values and the evolution forecasting of these data are taken into account to activate the procedure. Three episode types are distinguished. Depending on the value of the PM concentration and on the concerned area, the episode can be local, global or combined when both conditions of global and local episodes are met.

Area concerned
The alert activation procedure works by municipalities. There is a precise value of PM concentration to activate the alert procedure. As the actions are performed for each municipality concerned, the concentration has to be known for each of them. Thus, the measured values have to be interpolated to evaluate the concentration in all of the two hundred and sixty-two municipalities of Wallonia.

An episode is called global when two conditions are met. The first condition is on the measured daily mean of PM_{10} concentration. The limit value in the whole of Wallonia is $70\mu g.m^{-3}$. The second condition is on the forecasting of concentration. This condition is met if such concentrations are forecast for two days. In this case, actions are performed in the whole region.

An episode is called local when three conditions are fulfilled. The first one is on the measured daily mean of PM_{10} concentration in the area where heavy industry is installed. The limit value in this area is $100\mu g.m^{-3}$. The second one is on the measured daily mean of PM_{10} concentration in Northern Wallonia. The last condition, as for global episodes, is a two-day forecast. In this case, actions are performed in the area concerned.

VALIDATION
To validate the interpolation quality, a measurement campaign has been performed using a second network of six additional measurement stations. Both fixed and additional stations use measure the concentrations of PM in the air with a laser diffraction technology, GRIMM. These stations were installed for three months. The concentration of PM was measured for strategic locations. To choose these locations, the error of interpolation was analysed in the Charleroi area, **Error! Reference source not found.**. Once the locations had been chosen and the stations installed, the validation consisted of four steps:

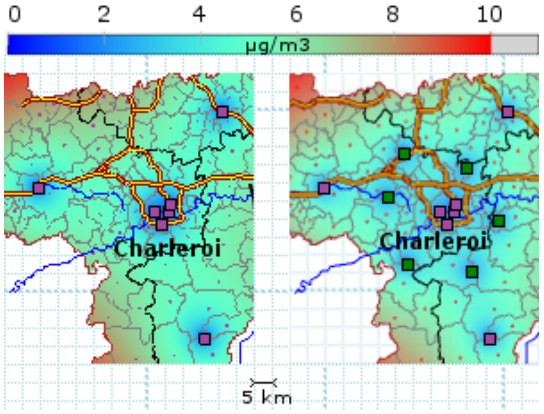

Figure 3: Error of interpolation in $\mu g.m^{-3}$

1. Measurement of PM concentrations at fixed stations
2. Interpolation of these measures to estimate the concentration at the six strategic positions;
3. Measurement of PM concentrations on these positions by using the additional stations;
4. Comparison of the interpolated values with the one given by the mobile measurement stations.

Measurement locations
The location of the fixed stations in the Charleroi area made it geometrically interesting for the measurement campaign. The fixed stations in this area are located as follow:

- One station in the town centre;
- Four stations around at a distance of four kilometres;
- Three stations forming a triangle around at a distance of thirty-five kilometres.

In order to have significant results, the additional stations were installed where the error of interpolation is maximal and reaches $5\mu g.m^{-3}$, left part of **Error! Reference source not found.**, i.e. halfway to fixed stations, right part of **Error! Reference source not found.**.

Results

The concentrations measured at the additional stations were compared to the values interpolated from concentrations measured at fixed stations. Table 1 shows the orthogonal linear regression results for daily means at the six additional stations for PM_{10} and $PM_{2.5}$.

European legislation aims at defining common methods and criteria to access air quality data (European Parliament and Council, 2004 and 2008). The European committee for standardization set the principles for the uncertainty assessment in order to homogenize the various approaches of the different ambient air working groups (European committee for standardization, 2002).

To compare and validate the data supplied by two measurement instruments (European Commission Working Group on Guidance for the Demonstration of Equivalence, 2010), the between-sampler uncertainty u_{bs} is used. This uncertainty is calculated from the differences of results of two devices operated in parallel. The instrument data is validated if the uncertainty is less than $2.5\mu g.m^{-3}$. As there are no criteria defined to validate an interpolation model, this uncertainty is used. It is computed from the differences between interpolated and measured daily means, Equation (2), where u_{bs} is the between-sampler uncertainty, $z_{i,meas}$ and $z_{i,int}$ are the daily means of measured and interpolated concentrations for day i and n the number of days.

$$u_{bs}^2 = \frac{\sum_{i=1}^{n} \left(z_{i,meas} - z_{i,\text{int}} \right)^2}{2n} \tag{2}$$

As the transport and diffusion phenomena have a certain duration, the longer the period of comparison, the better the correlation. For this reason, comparisons on daily averages and moving averages twenty-four hours are more conclusive.

Table 1 : Orthogonal linear regression and between-sampler uncertainty

Stations		S1	S2	S4	S5	S6
Number of days		71	70	78	78	59
Correlation coefficient	PM_{10}	0.9773	0.9684	0.9745	0.9851	0.9573
	$PM_{2.5}$	0.9878	0.9862	0.9839	0.9920	0.9767
u_{bs}	PM_{10}	2.77	3.13	2.41	2.19	2.32
	$PM_{2.5}$	1.95	1.71	1.67	1.99	1.58

For both PM_{10} and $PM_{2.5}$ the correlation coefficients are above 0.9. The uncertainty u_{bs} is less than $3.2\mu g.m^{-3}$ for PM_{10} and less than $2.0\mu g.m^{-3}$ for $PM_{2.5}$. The transport and diffusion phenomena do not have the same duration for PM_{10} and $PM_{2.5}$. Furthermore, PM_{10} and $PM_{2.5}$ do not have the same composition. $PM_{2.5}$ behaves almost as a gas. This difference of behaviour explains the difference obtained for PM_{10} and $PM_{2.5}$ for the correlation between measured and interpolated values.

Based on the criterion of u_{bs} less than $2.5\mu g.m^{-3}$, it can be said that the ordinary kriging model is validated for $PM_{2.5}$. For PM_{10}, this criterion is respected for three stations and the correlation coefficients are above 0.9 for all of them, so the model can also be validated for PM_{10}.

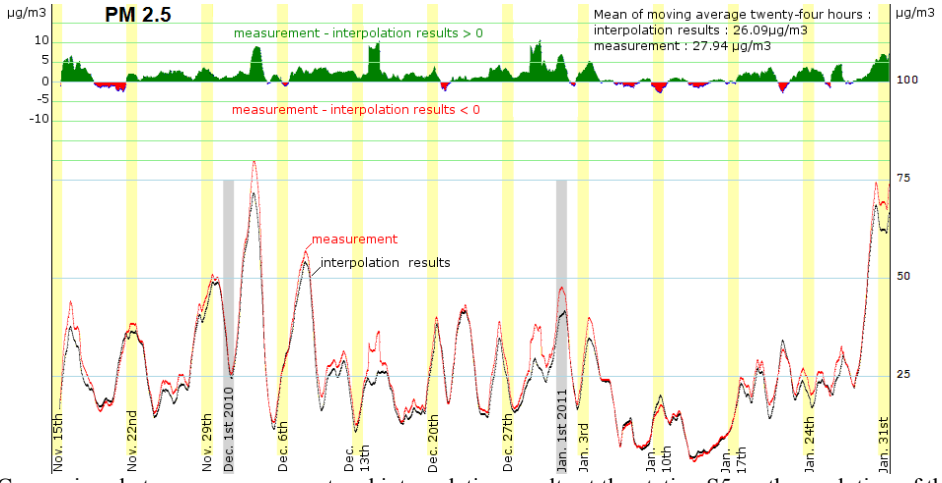

Figure 4: Comparison between measurement and interpolation results at the station S5 on the evolution of the moving average twenty-four hours

The evolution of the moving average twenty-four hours on the whole period of the campaign, Figure 4, confirms that the interpolation is close to the measurement with a difference between the means of moving average twenty-four hours less than 1.9µg.m^{-3}. The mean of moving average twenty-four hours is 26.09µg.m^{-3} for the interpolated values and 27.94µg.m^{-3} for the measured values. Furthermore, globally on the whole campaign period interpolation underestimated the measurement.

The comparison between the measured and interpolated values for both PM_{10} and $PM_{2.5}$ give correlation coefficients which are above 0.95. This leads to the conclusions that the model is finally well validated.

Furthermore, the between-sampler uncertainty, used to compare and validate the data supplied by two measurement instruments (European Commission Working Group on Guidance for the Demonstration of Equivalence, 2010), less than 3.2µg.m^{-3} and 2.0µg.m^{-3} for PM_{10} and $PM_{2.5}$ support this conclusion.

The measures are gross data and are not corrected by applying coefficients. The results are gross results.

CONCLUSION

The geostatistical approach has two advantages: it gives the spatial correlation between PM concentration and locations of measurement stations, and it gives an estimation of the interpolation error. This model has been validated by a measurement. During this campaign, the measured values of PM concentration were compared to interpolated values based on three criteria. The first one is the correlation coefficients close to 1. The second one is the between-sampler uncertainty u_{bs} less than 2.5µg.m^{-3}, wich is the criteria used to compare and validate the data supplied by two measurement instruments (European Commission Working Group on Guidance for the Demonstration of Equivalence, 2010). The last one is the difference of the mean of moving average twenty-four hours between interpolation and measurement during the three months of the campaign. So, with correlation coefficients which are above 0,9 for both PM_{10} and $PM_{2.5}$ for five stations, with u_{bs} less than 3.2µg.m^{-3}for PM_{10} and less than 2.0µg.m^{-3}for $PM_{2.5}$ and with a difference between the means of moving average twenty-four hours less than 1.9µg.m^{-3}, the ordinary kriging model used is validated.

ACKNOWLEDGEMENT

The Walloon air and climate agency (AWAC), for financial support and proof reading, and the scientific institute of public services (ISSeP), for proof reading, are gratefully acknowledged.

REFERENCES

AWAC, Méthode d'interpolation géostatistique, last accessed September 2011: http://airquality.issep.be/WebAirQuality/MicroParticules.aspx.

AWAC, Plan d'actions en cas de pic de pollution par les poussières fines, last accessed September 2011: http://airquality.issep.be/WebAirQuality/PicPollutionPoussieres.aspx.

C. Passelecq and C. Demuth., 1999: Développement d'une méthode d'interpolation spécifique aux polluants atmosphériques mesurés dans les réseaux automatiques. Contract AS/18/009.
Commission Fédérale de l'Hygiène de l'air (CFHA), 2007: Les poussières fines en Suisse. Rapport de la Commission Fédérale de l'Hygiène de l'air (CFHA).
European committee for standardization, 2002: report CR 14377:2002E, Air quality - Approach to uncertainty estimation for ambient air reference measurement methods.
European Commission Working Group on Guidance for the Demonstration of Equivalence, 2010: Guide to the demonstration of equivalence of ambient air monitoring methods.
European Parliament and Council, 2004: Directive 2004/107/EC relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air.
European Parliament and Council, 2008: Directive 2008/50/EC on ambient air quality and cleaner air for Europe.
ISSeP, last accessed September 2011: http://airquality.issep.be.
P.S.N. Murthy, T. Arora and S. Ahmed, 2008: Groundwater Dynamics in Hard Rock Aquifers, Springer, pp. 150-171.17.