RIO: A NOVEL APPROACH FOR AIR POLLUTION MAPPING

Stijn Janssen[1], Frans Fierens[2], Gerwin Dumont[2] and Clemens Mensink[1]

[1] Flemish Institute for Technological Research (VITO), Boeretang 200, B-2400 Mol, Belgium
[2] Belgian Interregional Environment Agency (IRCEL), Kunstlaan 10-11, B-1210 Brussels, Belgium

Abstract: Real-time assessment of the ambient air quality has gained an increased interest in recent years. To give support to this evolution, the statistical air pollution interpolation model RIO is developed. Due to the very low computational cost this interpolation model is an efficient tool for an environment agency when performing real-time air quality assessments. Beside this, a reliable interpolation model can be used to produce analysed maps of historical data records as well. RIO is an interpolation model that can be classified as a detrended Kriging model. In a first step the local character of the air pollution sampling values is removed in a detrending procedure. Subsequently, the site-independent data is interpolated by an Ordinary Kriging scheme. Finally, in a re-trending step a local bias is added to the Kriging interpolation results. As spatially resolved driving force in the detrending process, a land use indicator is developed based on the CORINE land cover data set. The indicator is optimized independently for the three pollutants O_3, NO_2 and PM_{10}. As a result, the RIO model is able to account for the local character of the air pollution phenomenon at locations where no monitoring stations are available. Through a cross-validation procedure the superiority of the RIO model over standard interpolation techniques, such as the Ordinary Kriging is demonstrated. Air quality maps are presented for the three pollutants mentioned and compared to maps based on standard interpolation techniques.

Key words: Air quality maps, interpolation, Kriging, CORINE Land Cover.

1. INTRODUCTION

Ambient air quality is a major concern in highly urbanized and industrialized regions such as Belgium. For its assessment, a dense network of automatic monitoring sites is implemented by the three Belgian regions, collecting real-time data on a half-hourly basis. The real-time measurements of the telemetric networks are used to inform the authorities and the public on actual air quality levels, to trigger a warning mechanism in case of threshold exceedance and to feed short term forecast models which predict air quality up to a few days ahead. The average distance between the nearest measuring stations is about 25 km. In spite of this dense coverage, it remains non-trivial to make an accurate spatial map from these point measurement values. However such a map is of great importance to accurately describe the general features of the spatial patterns of air pollution in Belgium, a need which has been put forward by the Belgian Interregional Environment Agency IRCEL. As an additional requirement, the resulting maps have to be available on-line in real time and at a minimal computational costs. This excludes the use of comprehensive time consuming deterministic air quality models for this purpose.

Apart from real time assessments, EU Member States must assess annual ambient air quality in all air quality zones and agglomerations on their territory. In this perspective, it is essential for the Environment Agencies of the Member States that they can rely on a tool for the production of correctly analyzed air quality maps. So, an accurate spatial interpolation model is an indispensable vital tool. In the recent literature a number of air pollution interpolation models have been presented (Ross et al. 2007, Arain et al. 2007, Denby et al. 2005). In these papers variants of the Kriging interpolation technique are presented as well as various possibilities for the implementation of land use regression models. These models combine in-situ measurements and auxiliary land use data to produce air quality maps.

Pollutants such as ozone, NO_2 or PM_{10} are governed by two different mechanism, each acting on a different spatial scale. On the regional level, fluctuations in the concentration pattern are mainly driven by meteorological (sub-continental) phenomena. Beside this, ambient air pollution can have a distinct local character due to local emission sources and their temporal variability. In an urbanized region such as Belgium, the latter effects are significant. In this paper we describe an interpolation model, called RIO (Janssen et al. 2008), that is developed to incorporate both the regional and local aspects of the air pollution phenomenon and that produces concentration estimates on a 4x4 km² grid.

2. METHODOLOGY

RIO's fundamental idea is a detrending or removal of the local character of the sampling values before they are interpolated. The transformation of the air pollution values into site-independent quantities results in a spatial homogeneous input set which is a prerequisite for a correct use of any interpolation algorithm. After the interpolation step, a re-trending or retransformation is applied in order to incorporate the local character in the final map at places where no monitoring data is available. To deal with the local scale of the air pollution phenomenon, a methodology is developed that links specific (statistical) properties of the air pollution to land use patterns at the same local scale which are described by a so call land use indicator. The methodology results in mathematical relations that establish those links and that can be used in the RIO interpolation model. Before the relations as such are discussed, the definition of the land use indicator will be briefly expounded.

The land use indicator, hereafter called the β-parameter, is calculated from the CORINE Land Cover data set. For a given area (~10 km²) the CORINE Land Cover pixels are determined and classified according to the 44 CORINE classes. The resulting classification histogram can be seen as a spectrum that represents a fingerprint of the land use characteristics in that particular area. It is the aim to define a single value land use indicator that represents the characteristics of the local behaviour of the air pollution phenomenon. Therefore, the CORINE class distribution is transformed into a land use indicator according to the relation:

$$\beta = \log\left[1 + \frac{\sum_i a_i \cdot n_{CLi}}{\sum_i n_{CLi}}\right], \tag{1}$$

In this formula, the index i runs over all CORINE classes. n_{CLi} is the number of pixels of class i inside the specified area and a_i is the pollution coefficient for the CORINE class i. As such, the β-parameter is the logarithm of a normalised sum of the CORINE class distribution in that particular region. The pollution coefficient a_i represents the impact of a particular CORINE class on the air pollution levels. To reduce the number of free parameters in the parameterisation of Equation 1, land cover classes that are expected to have a similar impact on the air pollution phenomenon are grouped together and receive the same pollution coefficient a_i. In addition, some a_i coefficients are fixed to zero or one to further reduce the number of degrees of freedom. As such, only 8 free parameters are left in the parameterisation of the land use indicator β (see Janssen et al. 2008 for more details)

In order to pin down the relation between the mean air pollutant concentrations and the land use characteristics, the β-parameter is determined inside a 2 km buffer around each monitoring site. After all, it is only at those locations that such a relation can be established since measurement data has to be at hand. When the expectation values are plotted versus the β-parameter for all available monitoring sites, a scatter plot as in Figure 1 for NO_2 is obtained. From this figure, a clear trend is revealed. Rural stations (low β) have low mean NO_2 concentrations, in urban or industrial sites (high β) increased NO_2 levels are observed. This relation between land use and air pollution can be approached by a second order polynomial fit. This functional form can then be used as an estimator for the pollutant concentration as a function of the land use indicator β. Similar relations are obtained for the pollutants O_3 and PM_{10} and they constitute the core of this methodology. These relations or so called "trend functions" are optimized by optimizing the a_i coefficients of the land use indicator β. The optimization is obtained by minimizing the RMS error of the polynomial fit.

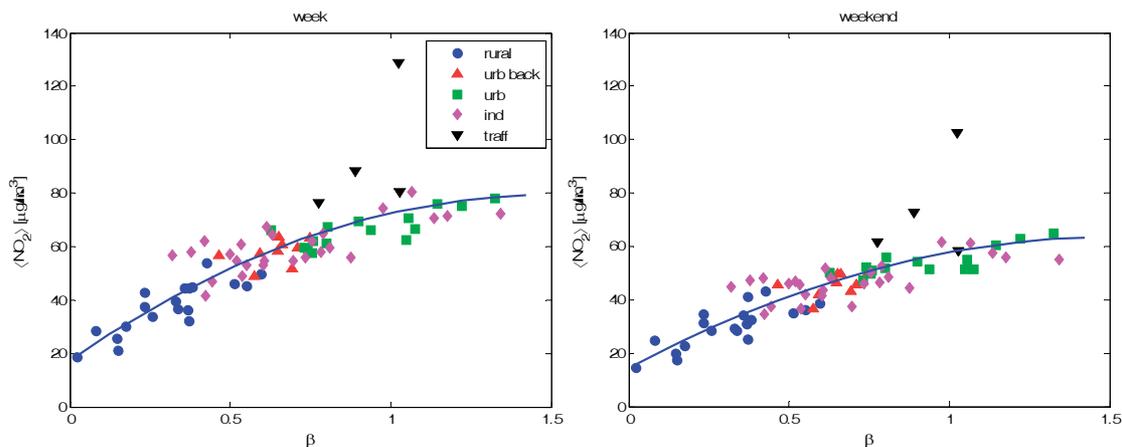Figure 1. Trend function for NO_2: average max 1h NO_2 values as a function of the β-parameter of the Belgian monitoring sites. The left panel shows the results for weekdays, the right panel is based on averages over weekend days. Averages are based on the summer values only between 2001 and 2006. Low β-values correspond to rural areas, high β-values to urban or industrialised sites. Stations are labelled according to their type (rural, urban background, urban, industrial and traffic).

Apart from trend functions for the mean values of the pollutant concentration, trends can also be established for the variance or standard deviation of the time series of the monitoring data. As a consequence it can be concluded that also the variance in the concentration values depends on the land use.

The basic idea of the RIO model is to apply those relations (for both the mean values and standard deviations) in a detrending procedure of the air quality sampling values. This detrending procedure is an essential step to obtain spatial homogeneity of the sampling values before they are used in the interpolation scheme. The detrending step is schematically depicted in Figure 2. According to a station's β-value, a theoretical shift ΔC is calculated as the difference between an arbitrary reference level (in the case of NO_2 this is 70 µgm⁻³) and the trend function. This

residue ΔC is then added to the sampling value to establish the detrending. By doing so, the local character of the monitoring data is removed and all values are transformed to a site-independent (β-independent) reference level. Of course, local differences between the various monitoring sites still exist in the detrended sampling values (as they were already present before detrending) due to local variability of the general air quality pattern over the region.

In a subsequent step the detrended values are interpolated by an Ordinary Kriging schema resulting in a high resolution map. When this Kriging interpolation is performed, a subsequent re-trending procedure has to be performed in order to restitute the impact of the local land use characteristics in each spatially interpolated point. This re-trending is carried out by making use of the specific β-value of the interpolation grid cell. Based on this grid cell specific β-value a concentration shift ΔC can be determined as the difference between the reference level and the trend function. This concentration residue ΔC is then subtracted from the grid cel concentration value obtained in the Kriging interpolation. The procedure re-introduces the local character in the concentration levels and the general idea of the re-trending is presented in Figure 3.
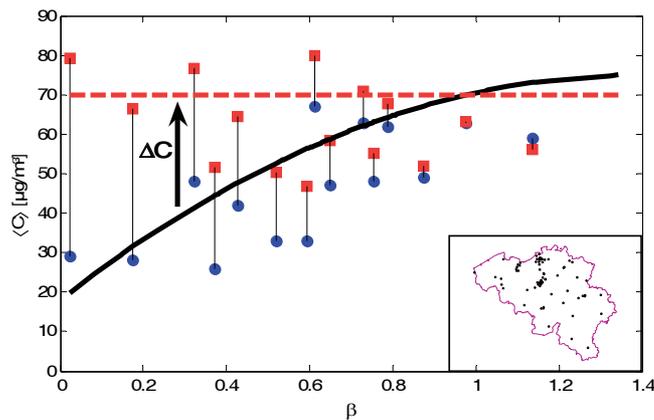

Figure 2. Schematic outline of the detrending procedure applied to measurements. The blue dots represent a selection of measurement values. The red squares are obtained after the detrending step. A concentration residue ΔC is added to the measured values. ΔC is given as the difference between the reference level and the trend function for the appropriate β-value. The trend function is depicted as the solid line, the reference level is given as the dashed red line.
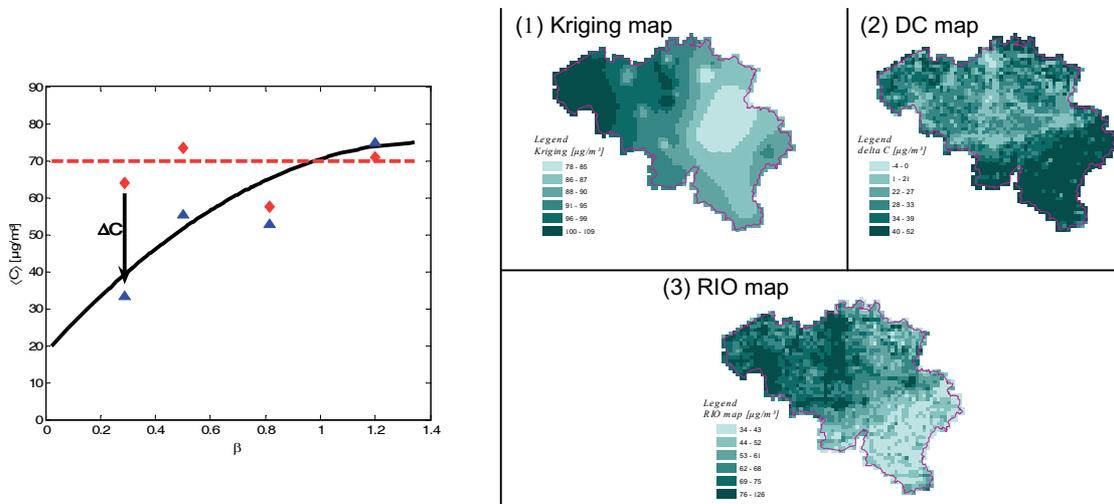

Figure 3. Schematic outline of the re-trending procedure applied to grid cells. The panel left gives the transformation for 4 different grid cells. The red diamonds are the result of the Kriging interpolation. The blue triangles are the result of the re-trending step. A concentration residue ΔC is subtracted from the Kriging result. ΔC is given as the difference between the reference level and the trend function for the appropriate β-value. The trend function is depicted as the solid line, the reference level is given as the dashed red line. The panels on the right give a Kriging map of detrended measurements (1), a map of the ΔC residue (2) and the final RIO result (3).

3. RESULTS

The performance of the RIO interpolation model is examined in a cross-validation analysis making use of the "leaving-one-out" principle. For one particular station each value in the historical time series is recalculated making use of all available monitoring data at that time step except the sampling value of that particular station. This procedure is repeated for the entire time series and for every monitoring station. The years 2003–2006 are used as a fixed time period in this analysis. As a validation procedure, the interpolated time series obtained in this way can then be (statistically) compared to the genuine measured values. Here, the quality of the model results is expressed by the Root Mean Square Error (RMSE), the bias and the Mean Absolute Error (MAE).

In order to further evaluate the overall performance of the RIO model, RIO results are compared to two standard interpolation techniques. For this purpose, a fourth power Inverse Distance Weighting (IDW) interpolation scheme is applied together with the Ordinary Kriging (OK) interpolation methodology. With the exception of a few monitoring locations, the performance of the RIO model is in almost all monitoring stations better than the standard technique. Averaged RMSE, bias and MAE values over all available Belgian monitoring locations are summarized in Table 1. It is clear from the table that RIO outperforms the standard interpolation techniques as for each of the three pollutants the three quality indicators (RMSE, bias and MAE) improve. These RMSE correspond to rather accurate estimates (relative errors of 11% for O_3, 26% for NO_2, 28% for PM_{10}) which can clearly compete, as far as accuracy is concerned, with the results of much more sophisticated deterministic models used for air quality assessments (Van Loon et al. 2007, Vautard et al. 2007)

Table 1. Model performance indicators for the IDW, OK and RIO model and for the three pollutants O_3, NO_2 and PM_{10}. Values are averaged over all available Belgian monitoring locations. The mean value for the corresponding time averaging value (e.g. max 1h) is given between brackets $<X>$. All values are in μgm^{-3}.

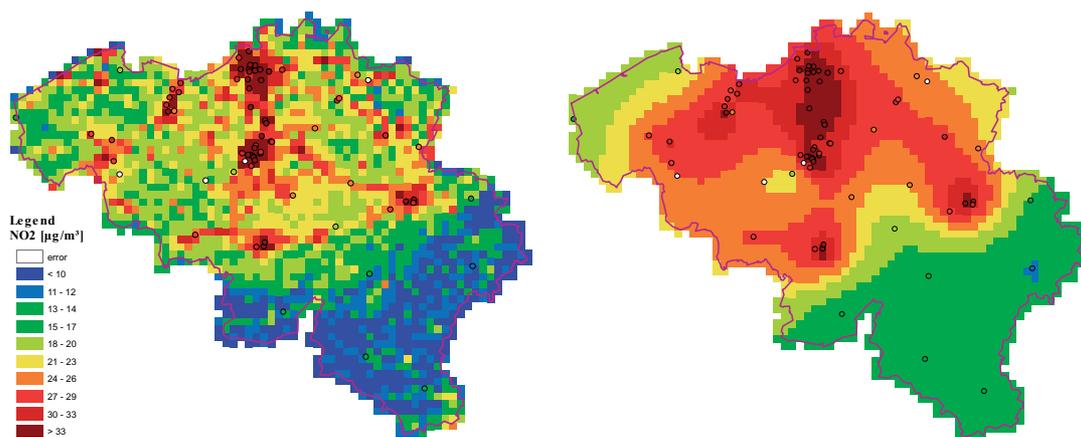| model	O_3 (max 1h, summer) < 90.9 >			NO_2 (max 1h) < 54.8 >			PM_{10} (day avg) < 35.9 >		
	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE
IDW	10.97	-1.70	8.16	18.17	4.74	14.43	12.12	1.70	8.49
OK	10.37	-0.44	7.66	16.85	1.45	13.11	11.65	1.22	8.05
RIO	9.56	-0.08	6.89	14.45	-0.67	11.23	9.89	0.01	6.98

Figure 4. Annual mean NO_2 map for the year 2006 obtained by the RIO model (left) and Ordinary Kriging (right). The small coloured circles on the map represent the ozone monitoring locations and the corresponding annual mean value at that station.

Apart from an evaluation on a statistical basis the performance of the interpolation model can also be evaluated by means of the actual annual air quality maps for 2006 obtained with the different interpolation techniques. The maps of the annual mean are obtained by an hour by hour interpolation of all available (hourly) measurements in 2006. Then, for each pixel in the final map all hourly results are averaged over the year. In Figure 4 the annual mean NO_2 maps for the year 2006 are presented for both the RIO and Ordinary Kriging interpolation techniques. A comparison of the two maps clearly shows the strength of the RIO interpolation scheme as it is able to introduce land use based local variations in the NO_2 concentrations fields at places where no monitoring data is available. Whereas standard techniques have to rely only on the monitoring data themselves and their relative distances and correlations. The differences between the Ordinary Kriging map and the RIO map are due to the fact that a great deal of the NO_2 monitoring stations is located in urbanised or industrialized areas, more than it is the case for the ozone sampling

sites. As a consequence, a standard interpolation technique extrapolates these enhanced NO_2 levels from the urban sites to the rural areas. RIO is able to deal with these urban stations in a much more accurate way by limiting the high values to only those areas with an urban character. Similar differences are observed for the ozone and PM_{10} maps.

4. CONCLUSIONS

In this paper the RIO model is presented as an "intelligent" interpolation model for air pollution data. The methodology makes use of a land use indicator (the β-parameter), which offers a maximum flexibility in the development of the relations between land use and air pollution levels. Those relations are the core of the RIO model and they are used to remove the local character of the air quality monitoring data before they are used in the interpolation scheme. In the RIO model, a detrending of the monitoring data is applied according to the observed trend in the mean values and the standard deviation.

The evaluation of the RIO model in a cross-validation procedure clearly revealed that the new methodology produces better results compared to standard interpolation techniques such as IDW or Ordinary Kriging. The improvements of the RIO model can be attributed to the detrending procedure which is applied on the air quality monitoring data and the re-trending step that is imposed on the Kriging map. It was shown that RIO performed best on the basis of all indicators. Subsequently, the interpolation model was further evaluated based on the resulting maps. It was illustrated that RIO is able to produce local variations in the concentration fields at places where no monitoring data is available. This is because of RIO's ability to incorporate land use information and not just monitoring data. This is especially important if the air quality gradients at the urban-rural interface have to be described.

REFERENCES

Arain, M.A., Blair R., Finkelstein N., Brook J.R., Sahsuvaroglu T., Beckerman B., Zhang L. and Jerret M., 2007: The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment*, **41**, 3453-3464.

Denby, B., Horálek J., Walker S.E., Eben K. and Fiala J., 2005: Interpolation and assimilation methods for European scale air quality assessment and mapping, Part I: Review and recommendations. *ETC/ACC Technical Paper*, **2005/7**.

Janssen, S., Dumont G., Fierens F. and Mensink C., 2008: Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment*, **42**, 4884–4903.

Ross, Z., Jerrett M., Ito K., Tempalski B., and Thurston G.D., 2007: A land use regression for fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, **41**, 2255-2269.

van Loon, M., Vautard R., Schaap M., Bergstrom R., Bessagnet B., Brandt J., Builtjes P.J.H., Christensen J.H., Cuvelier C., Graff A., Jonson J.E., Krol M., Langner J., Roberts P., Rouil L., Stern R., Tarrason L., Thunis P., Vignati E., White L. and Wind P., 2007: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmospheric Environment*, **41**, 2083-2097.

Vautard, R., Builtjes, P.H.J., Thunis, P., Cuvelier, K., Bedogni, M., Bessagnet, B., Honoré, C., Moussiopoulos, N. Pirovano, G., Schaap, M., Stern, R., Tarrason and L., Van Loon, M., 2007: Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry-transport models over 4 European cities within the CityDelta project. *Atmospheric Environment*, **41**, 173–188.