Available online at www.sciencedirect.com
ScienceDirect
Atmospheric Environment 42 (2008) 4884–4903
ATMOSPHERIC ENVIRONMENT
www.elsevier.com/locate/atmosenv

Spatial interpolation of air pollution measurements using CORINE land cover data

Stijn Janssen[a,*], Gerwin Dumont[b], Frans Fierens[b], Clemens Mensink[a]

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

Received 22 July 2007; received in revised form 31 January 2008; accepted 19 February 2008

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 assessment. Beside this, a reliable interpolation model can be used to produce analysed maps of historical data records as well. Such maps are essential for correctly checking compliance with population exposure limit values as foreseen by the new EU Air Quality Directive. 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.
© 2008 Elsevier Ltd. All rights reserved.

Keywords: Interpolation model; Air quality assessment; CORINE land cover; Kriging

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[1] 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

*Corresponding author. Tel.: +32 14 33 59 67; fax: +32 14 32 11 85.
E-mail address: stijn.janssen@vito.be (S. Janssen).

[1]The telemetric networks are managed by the Flemish Environment Agency (VMM) in Flanders, the Brussels Institute for Management of the Environment (IBGE-BIM) in Brussels and the Institute for Public Service (ISSEP) and the Ministry of the Environment (DGRNE) in Wallonia.

1352-2310/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2008.02.043

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 cost. 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. As expressed by the EU Air Quality Framework Directive 96/62/EC (EC, 1996) and the four emanating Daughter Directives, the actual reporting of this assessment is done—according to the EU Commission Decision 2004/461/EC (EC, 2004)—with the help of a questionnaire listing up the levels and exceedances at the measuring points in order to check compliance with the EU limit and target values. In view of a new Air Quality Directive the Commission tends, in line with the INSPIRE directive 2007/2/EC (EC, 2007), to promote more geographical oriented air quality assessment tools in order to check compliance throughout the zones and agglomerations of the Member States, rather than checking compliance in uplisted data at individual and not always geographically representative points. Moreover, the new Air Quality Directive introduces target and limit values for population exposure to fine particles, which should be assessed on the base of data correctly distributed over the area. 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 analysed 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 geographical 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 (Tombette and Sportisse, 2007; Mensink et al., 2007). Beside this, ambient air pollution can have a distinct local character due to local emission sources and their temporal variability (Vautard et al., 2007). In an urbanized region such as Belgium, the latter effects are significant (Mensink et al., 2005). In this paper, we describe an interpolation model, called RIO, that is developed to incorporate both the regional and local aspects of the air pollution phenomenon and that produces concentration estimates on a $4 \times 4\,km^2$ grid.

The outline of this paper is as follows. In Section 2, the RIO methodology is described in detail. In Section 3, model results are discussed, validated and compared with standard interpolation techniques. The same section contains a subsection on the uncertainty of the RIO results and outlines the operational setup of the model. A conclusion is presented in Section 4.

2. Methodology

The basics of the RIO model are described in depth in the paper by Hooyberghs et al. (2006). Here, we will briefly resume the outline of the methodology with a focus on the extensions of the RIO model.

2.1. Land use patterns and ambient air quality

RIO's fundamental idea is a detrending or removal of the local character of 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. Note that this fundamental requirement of spatial homogeneity of the input data is often disregarded in (air pollution) interpolation analyses. After the interpolation step, a retrending 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. The

methodology results in mathematical relations that establish those links and that can be used in the RIO interpolation model.

In our previous work (Hooyberghs et al., 2006), the driving force (explaining variable) for detrending the measured ozone values was the amount of ozone titration by NO. The strength of the local NO-titration was surrogated by the local population density. However, we have observed that the spatial pattern of population density, although being a good driving force for ozone concentrations, is not necessarily a good and sufficient driving force for other pollutants such as PM_{10} that are not as exclusively linked to population density as ozone. It is expected that the applicability of population density as an explaining variable for the local character of all pollutants (e.g., NO_2, PM_{10}, etc.) is rather limited and could be extended by additional land use information. As such, a land use parameter is defined for each pollutant based on CORINE land cover (CLC) maps (EEA 1995). By relying on the CLC data set, the CLC2000 release is used, all different land cover types (urban, industrial, traffic, agricultural, semi natural, etc.) can be included in the analysis. The CLC data set provides a high-resolution description ($100 \times 100\,m^2$ pixels are used) of land use patterns, making use of 44 different land cover classes. As will be described in the next Section 2.2, the CLC description is transformed into a so-called land use indicator, specifically optimized for each pollutant. By doing so, the high dimensionality of the driving force data (44 CLC classes) is reduced to one single value, representative for the local land use characteristics for a particular pollutant at a given site. A CLC map of Belgium, together with the NO_2 monitoring locations is presented in Fig. 1. The red-coloured densely populated regions in the northern and central part of the country are clearly revealed as well as the green-coloured rural and forest areas in the south.

2.2. Land use indicator

In the assessment procedure of the land use indicator, the first step is a characterisation of the

Fig. 1. CORINE land cover map for the Belgian region. NO_2 monitoring stations with label are indicated by the dots. Densely populated and industrialised areas show up as red-coloured while rural and forest areas are green.

land cover in the direct vicinity of the monitoring sites. After all, it is only at those locations that both air pollution concentrations and land use information is at hand and that relations between both can be formulated.

For each monitoring station, a buffer zone dimensioned by a fixed radius is created (see Fig. 2). Inside each buffer zone, the CLC pixels ($100 \times 100\,m^2$) are determined and classified according to the 44 CLC classes. The resulting histogram can be seen as a spectrum that represents a fingerprint of the land use characteristics of the surroundings of the monitoring stations. Fig. 3 shows that these land use fingerprints are quite different in the vicinity of a rural, an urban background and an urban station.

Fig. 2. Monitoring station 43R240 (Engis) with a 2 km radius buffer. CLC describes the land use in the direct vicinity of the monitoring location.

The radius of the vicinity buffer is a free parameter in the methodology but it turned out that a radius of 2 km produces the best results. When the buffer is chosen too small, it becomes difficult to discriminate between, e.g., a small town and a larger city. When the buffer is too large, the site-specific character of the CLC class distribution disappears and the distribution evolves to a general spectrum with very little discriminating power.

2.2.1. Establishing the land use indicator β

In order to reduce the high number of degrees of freedom in the land use description, the 44 CLC classes are grouped into 11 more general land use

Table 1
Definition of the 11 RIO classes (RCL) as a combination of the 44 CLC classes

RIO class	Description	CLC classes
RCL1	Continuous urban fabric	1
RCL2	Discontinuous urban fabric, green and sport	2, 10, 11
RCL3	Industrial or commercial units	3
RCL4	Road and rail networks and associated land	4
RCL5	Port areas	5
RCL6	Airports	6
RCL7	Mine, dump and construction sites	7–9
RCL8	Arable land	12–14
RCL9	Agricultural areas	15–22
RCL10	Forest and semi natural areas	23–34
RCL11	Wetlands and water bodies	35–44

Fig. 3. CLC class distribution in the close vicinity of three typical monitoring locations in Belgium (rural, urban background and urban). For each CLC class, the number of grid cells inside the station vicinity buffer is indicated.

classes, hereafter called the RIO classes (RCL). The definition of those RIO classes is given in Table 1. Note that especially for the urban (CLCI), industrial (CLC3) and traffic (CLC4, 5 and 6) classes, the highest resolution of the CLC classes is maintained in the RIO set.

In order to define a single value indicator that correlates the local land use characteristics to the local air pollution levels, the RIO class distribution in the vicinity of the station is transformed into a single land use parameter β, according to the relation:

$$\beta = \log\left[1 + \frac{\sum_i a_i \times n_{\mathrm{RCL}i}}{\sum_i n_{\mathrm{RCL}i}}\right]. \quad (1)$$

In this formula, the index i runs over all RIO classes. $n_{\mathrm{RCL}i}$ is the number of pixels of class i inside the buffer and a_i is a pollution-related coefficient for the RIO class i. As such, the β-parameter is the logarithm of a weighted and normalised sum of the RIO class distribution. The pollution-related coefficients a_i are used to weigh the importance of a particular RIO class on the concentration in the air of the pollutant under investigation. The β-indicator can then be optimized for each pollutant individually by choosing a best set of a_i coefficients for the RIO classes as will be described in Sections 2.2.2 and 2.2.3.

2.2.2. First guess of the land use indicator β

A zero value for a particular a_i coefficient in Eq. (1) means that the corresponding RIO class i is of no importance for the determination of the pollutant concentration in that area. Generally this might be the case for the RIO classes RCL10 and RCL11 which correspond to rural or natural land and water bodies. The RIO class RCL2, on the other hand, (discontinuous urban fabric) can be considered as a land use type of major importance for the air pollution. Its corresponding coefficient a_2 can be set at a reference value of 1 for normalisation purposes. By fixing the a_i coefficients for RIO classes RCL10 and RCL11 at zero and for class RCL2 at the value of 1, the β-parameter is forced to glide between 0 and an acceptable upper value (somewhere below 2). Moreover, by setting and keeping three out of the 11 a_i values at fixed values, the number of degrees of freedom for the parameterisation of the β-value (Eq. (1)) is reduced from 11 to 8.

A numerical procedure will further on optimize the 8 remaining a_i coefficients. However, a detailed analysis pointed out that the optimal solution is sensitive to the initial guess of the a_i set. As can be expected, this optimal set is probably only a local minimum in an eight-dimensional parameter space. In order to avoid finding in an irreproducible way a local minimum as an optimal set in the optimization, the procedure should start from a well-founded air quality-related initial guess for the a_i coefficients. For this, we rely on the distribution of the Belgian national air pollution emission data over the 11 EMEP sectors (EMEP, 2007). To do so, a pragmatic relation between the 11 EMEP sectors and the 11 RIO classes was needed. The relations are presented in Table 2 together with the corresponding Belgian emission totals of NO_x and PM_{10} for the year 2004.

Table 2
Pragmatic relation between the RIO classes and the EMEP sectors

RIO class	Pollutant coefficient, a_i	EMEP sectors	Emission totals 2004 (Gg yr^{-1})		Relative emissions with respect to RCL2 (a_i initial guess)	
			NO_x	PM_{10}	NO_x	PM_{10}
RCL1	a_1	S2	25.84	2.84	1.0	1.0
RCL2	a_2	S2	25.84	2.84	1.0	1.0
RCL3	a_3	S3 + S4	71.91	19.18	2.8	6.8
RCL4	a_4	S7	133.53	9.33	5.2	3.3
RCL5	a_5	S8	21.61	23.69	0.8	8.3
RCL6	a_6	S8	21.61	23.69	0.8	8.3
RCL7	a_7	S1 + S4 + S5 + S9	66.71	20.41	2.6	7.2
RCL8	a_8	S10	0.00	3.06	0.0	1.1
RCL9	a_9	S10	0.00	3.06	0.0	1.1
RCL10	a_{10}	S11	0.00	0.00	0.0	0.0
RCL11	a_{11}	S11	0.00	0.00	0.0	0.0

Corresponding EMEP emission totals for the year 2004 and their ratio to the RCL2 value. The latter ones will be used as initial guess for the a_i-parameters in the β-optimisation.

In order to transform emission data into a_i coefficient values while keeping the already fixed values of zero for the a_{10} and a_{11} coefficients as well as the reference value of 1 for the a_2 coefficient, the sector emission data is normalised with respect to the RCL2 values. In this way, a justified set of a_i coefficients is obtained which can be used as initial guess for a reproducible β-optimization procedure for each pollutant. The NO_x set is used as start value for both O_3 and NO_2 since both pollutants are strongly correlated through atmospheric photochemistry.

Based on the first guess parameterisation of the β-parameter, a β-value can be assigned to each air quality monitoring station. The corresponding values give an indication of the land use characteristics in the direct vicinity of the sampling site and will be used to classify the sites.

2.2.3. Optimization of the land use indicator β

Once the stations are characterised by a β-value, we can try to formulate the relations between land use and air pollution properties. Therefore, the long-term average of the pollutant value in each monitoring station is plotted versus the corresponding β-value of the station. This plot represents what will be called further on the "trend" function: it is the correlation between the mean pollutant value and the land use β-parameter for the near vicinity (e.g., see Fig. 4). In this analysis, a time averaging period from 2001 to 2006 is used for all pollutants.

So far, the β-value is determined by the first guess parameterisation of the RIO class distribution. At this stage, it is essential to first optimize this parameterisation for each pollutant individually. This can be done by optimizing the correlation between the average air pollution value and the β-parameter. Indeed, a perfect (or high) correlation between the average air pollution values and the β-parameter can be understood as an optimal parameterisation of the local character of the spatial distribution of the air pollution. Here, the optimization is obtained by minimizing the RMS error of a second-order polynomial fit. For the numerical optimization, a Generalized Reduced Gradient nonlinear optimization code is used, standard available in a numerical software package. Based on the initial guess values formulated in Section 2.2.2, optimal solutions were identified for O_3 and NO_2. For PM_{10}, the solver did not find the best solution automatically. With some minor manual corrections and by further optimizing only a subset of the a_i set, a slightly lower RMSE value for the fit could be identified. However, in general, we can conclude that the procedure for the initialisation of the a_i parameters was successful and gives rise to acceptable results which can easily be reproduced. The resulting optimal weights a_i for the β-parameterisation are summarized in Table 3.

The second-order polynomial fit that is obtained in the optimization procedure can be used as a mathematical expression for the trend function

Fig. 4. Trend function for ozone: average max 1 h ozone 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).

Table 3
Optimised set of a_i-parameters for the β-indicator

RIO class	Description	O_3		NO_2		PM_{10}	
		a_i	Fraction (%)	a_i	Fraction (%)	a_i	Fraction (%)
RCL1	Continuous urban fabric	1.01	5.9	2.71	4.3	1.16	3.6
RCL2	Discontinuous urban fabric	1.00	35.4	1.00	34.2	1.00	38.9
RCL3	Industrial or commercial units	1.41	8.6	1.23	11.2	2.07	11.2
RCL4	Road and rail networks	4.30	2.1	6.12	2.6	2.23	2.5
RCL5	Port areas	8.38	0.2	3.24	4.1	2.65	2.1
RCL6	Airports	0.97	0.1	1.41	0.9	1.01	1.1
RCL7	Mine, dump and construction sites	1.96	0.7	1.58	0.8	10.99	0.8
RCL8	Arable land	0.80	13.5	0.33	10.9	0.64	10.7
RCL9	Agricultural areas	0.35	22.3	0.00	17.7	0.64	17.1
RCL10	Forest and semi natural areas	0.00	9.5	0.00	8.1	0.00	8.0
RCL11	Wetlands and waterbodies	0.00	1.7	0.00	5.2	0.00	4.0

Parameter values are given for each pollutant. The column "Fraction" contains the relative occurrence of a RIO class in the buffer zones of all monitoring stations.

describing the land use–air quality relations. If the polynomial fit turns out to be non-monotone (e.g., Fig. 4—right panel), a constant level is imposed for the lower β-values to describe the plateau which is reached. As an alternative, an asymptotic exponential function can be used to describe the trend. In a cross-check analysis, it was concluded that this alternative function gives similar results.

Based on the optimized a_i parameters, a β-indicator can be calculated for each monitoring station and for each type of pollutant separately and a corresponding best fit for the trend function is obtained. Results for the summer average of the maximum daily 1 h ozone values are presented in Fig. 4 where the average is given separately for week and weekend days. As becomes clear from the plots, a clear trend can be observed for the average daily maximum ozone value as a function of the β-parameter. This Figure also indicates the well-known characteristics of the ozone phenomenon in Belgium, which is typical for regions with a high population and traffic density (Vanderstraeten, 1996). For low β-values (in rural areas), average ozone values are high and equal for week and weekend days. For high β-values corresponding to urban or industrialized regions, lower average ozone values are observed and much more difference is revealed between week and weekend days. This phenomenon can be explained by the so called "NO-titration effect". This effect is known for a long time (Jiménez et al., 2005; Sillman, 1999) and was also already described by Hooyberghs et al. (2006).

For NO_2, the mean values show a more intuitive behaviour with respect to the β-values. This is presented in Fig. 5 where higher daily maximum values are observed for highly urbanised or industrialized locations. Note that also here a difference is observed between week and weekend days. The results for PM_{10} are presented in Fig. 6. From this plot, it is clear that PM_{10} shows the weakest relation of the three pollutants considered in this study. This is not a complete surprise, since PM_{10} is a rather erratic pollutant with many unknown or poorly characterised emission sources (Fuller and Green, 2004; Almeida et al., 2005) and complex secondary formation processes (Sillanpää et al., 2006).

2.2.4. β-Land use indicator maps

Up to now the optimized parameterisation of the β-indicator was applied to establish the land use indicator in the vicinity of the monitoring sites. However, the β-indicator can also be calculated on a regular grid for the entire territory of the country. To obtain such a gridded β-map on the base of the CLC map shown in Fig. 1, the contribution of each RIO class (or equivalently CLC) is calculated in each grid cell and subsequently converted into a unique β-value according to the definition in Eq. (1). The resulting β-maps for ozone, NO_2 and PM_{10} are presented in Fig. 7. The plots in this figure indicate the similarities and the differences in the general spatial distribution pattern of the three air pollutants. The spatial distribution pattern of ozone and NO_2 are rather similar. This can be explained by their strong (but opposite) relation to NO_x emission

Fig. 5. Trend functions for NO_2: same as in Fig. 4 but for the average max 1 h NO_2 concentrations on weekdays (left panel) and on weekend days (right panel).

Fig. 6. Trend functions for PM_{10}: same as in Fig. 4 but for the daily mean PM_{10} concentrations on weekdays (left panel) and on weekend days (right panel).

Fig. 7. β-Land use indicator maps for ozone (left), NO_2 (middle) and PM_{10} (right). β-Values are calculated on a $4 \times 4\,km^2$ regular grid.

sources in urbanised and industrialised areas. It must indeed be repeated that due to the NO-titration effect, high β-values in the ozone map correspond to low ozone concentrations whereas the opposite is true for NO_2 concentrations. In the β-map for PM_{10}, the major urbanised regions are recognised as well but higher β-values are now also observed in some rural areas of the country. Those regions are characterised by important agricultural activities which have a non-negligible impact on the PM_{10} formation due to primary emissions and emissions of PM_{10} precursors such as NH_3 (Erisman and Schaap, 2004; Pinder et al., 2007; Mensink et al., 2007).

In the CLC data set as well as in the RCL dataset, traffic is available as an individual class (class 4, see Table 1) provided that this is the dominant land use in the $100 \times 100\,m^2$ pixel. It turns out that only a small fraction of the total number of pixels can be attributed to a pure traffic class. Nevertheless, the effect of traffic is expected to be large in the vicinity of the major highways, in particular for NO_2. To upgrade the current RIO methodology, a combination of CLC information and traffic volumes was investigated to deduce a more complex traffic-enhanced β-parameter along the same lines as presented above. It turned out that the addition of an optimized traffic term contributed to the NO_2 β-values for some 5% as an average over all grid cells, although for some cells the contribution was more important. For ozone and PM_{10}, no contribution at all was observed after the traffic-enhanced optimization procedure. From this analysis, it can be concluded that most of the information of traffic impact on ambient air quality levels is already contained in the CLC data set.

2.3. Detrending of measured values

So far, only trend functions for the mean values of the pollutant concentration at the measuring sites are presented. However, similar trends can be established for the variance or standard deviation of the time series of the monitoring data (Fig. 8). A net trend is observed for the standard deviation of the distribution of NO_2 and PM_{10} values. As a consequence, it can be concluded that also the variance in those NO_2 and PM_{10} concentration values depends on the land use. For ozone, it turns out that the variance is rather independent of the land use characteristics.

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 Fig. 9. 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 $70\,\mu g\,m^{-3}$) 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

Fig. 8. Trend functions for the standard deviation: standard deviation of the sampling values (week days only) as a function of the β-parameter for max 1 h ozone (left), max 1 h NO_2 (middle) and daily mean PM_{10} concentrations (right). The ozone results are obtained from summer data (April until September) only.

Fig. 9. Schematic outline of the detrending procedure applied to measurements. The dots represent a selection of measurement values. The 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 line.

Fig. 10. NO_2 max 1 h probability distribution for the urban station 41B006 and the rural station 44N050, before (left) and after (right) the detrending procedure.

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.

The detrending of the mean value is performed by a simple linear translation as presented in Fig. 9. As mentioned above, for some of the pollutants a clear trend is also observed in the standard deviation of the monitoring time series. In order to remove also the discrepancies in the variances of the sampling data at different locations, a detrending (or rescaling) is also applied for the standard deviation according to the transformation:

$$x_{\mathrm{det}} = (x - \langle x \rangle)\frac{\sigma_{\mathrm{ref}}}{\sigma(\beta)} + \langle x \rangle, \tag{2}$$

where σ_{ref} is the reference level of the detrending, $\sigma_{(\beta)}$ is the theoretical standard deviation corresponding to the local β-value of the station and $\langle x \rangle$ is the overall mean of the time series. This transformation ensures a reshaping of the data to obtain a uniform standard deviation σ_{ref}. The result of the detrending procedure for two typical stations is presented in Fig. 10. From those graphs, it

Fig. 11. Mean values (upper panel) and standard deviations (lower panel) of the max 1 h NO_2 values in all stations before (left bar) and after (right bar) the detrending procedure. The reference level used in the detrending is given as the straight line. Stations are ordered according to their β-value, rural stations to the left, urban and industrial stations to the right.

becomes clear that initially both stations have a totally distinct character and distribution. Those differences are almost completely removed after the detrending operation. An overall appreciation of the quality of the detrending procedure can be obtained from Fig. 11, where the mean values and standard deviations for all NO_2 stations are presented before and after the detrending procedure. From this figure, it is clear that a large part of the local character of the measurement time series is removed by the detrending, both in the mean value and the variance of the probability density. This is a crucial achievement in the RIO methodology.

2.4. Kriging interpolation of detrended values

The interpolation routine of RIO is based on an Ordinary Kriging (OK) scheme (Isaaks and Srivastava, 1989). This technique was originally developed in geostatistics but has proven to be successful in other domains as well (Van Leeuwen et al., 1996; Denby et al., 2005). The Kriging technique used in this work differs from the standard technique on one important point. In standard Ordinary Kriging, the variogram or spatial correlation function is solely determined on the basis of the input values of the interpolation problem at hand. However, for air quality interpolation problems the number of input values for a particular situation (e.g., a 1 h averaged real-time data set) is rather limited. For a country such as Belgium, typically 50 sampling sites are available. This limited number of sampling values can compromise the statistical significance of the correlation function or variogram. Moreover, air pollution sampling values can be subject to rather large variations on a 1 h time step. As a result, the correlation function will be subject to the same (erratic) variations. To tackle this problem and to improve statistics, the correlation functions in RIO are constructed on the basis of the entire historical set of measurement time series. In this way, a larger data set over different years can be used for the construction of the correlation function. In addition, this procedure allows for a unique correlation function for each time averaging value that is to be calculated by the interpolation model. Correlation functions can for example be constructed for, maximum 1-h values, for daily mean values, for maximum daily 8-h mean values or for each specific hourly value of the day. The correlation functions are not calculated for week and weekend days separately as was done for the trend functions. In a cross-check, only minor differences were observed when both data sets are handled separately. This is because the correlation functions are established on the base of already detrended time series where week and weekend days differences are already levelled out since both data sets are transformed to the same reference level.

Examples of the spatial correlation functions are given in Figs. 12 and 13. The points in the plots represent the correlation values defined as the

Fig. 12. Spatial correlation functions for detrended max 1 h summer ozone values (left), max 1 h NO_2 (middle) and daily mean PM_{10} concentrations (right). The spatial correlation function is printed as a solid line.

Fig. 13. Spatial correlation function for hourly ozone values, specifically determined for each hour of the day.

covariance of two time series in a pair of two detrended monitoring stations, separated by a distance r. The red lines in the plots represent the best (exponential) fit and are used as a spatial correlation function in the Kriging interpolation scheme.

From Fig. 12, it is clear that a significant difference is observed between the three pollutants. It turns out that (detrended) ozone is highly correlated, up to very long distances. For PM_{10} and NO_2, much more scatter is observed in the spatial correlation plots and correlation functions with a shorter range are obtained.

As stated before, the RIO methodology can be applied for any time aggregation value, e.g. for the maximum daily 8-hour mean concentration. Therefore, the correlation functions are also derived for each hourly value of the day. For ozone, the results are presented in Fig. 13 and similar results are obtained for NO_2 and PM_{10}. In this plot, it can be

noticed that for morning and evening ozone values the spatial correlation is rather low. For ozone values in the (late) afternoon, high correlations are observed in agreement with the maximum 1-h correlation function in Fig. 12.

2.5. Re-trending the Kriging results

When the Kriging interpolation of the detrended air quality values 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 (see Fig. 7 for the β-maps). 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 cell concentration value obtained in the Kriging interpolation. The procedure re-introduces the local character in the concentration level. The general idea of the re-trending is presented in Fig. 14. Although not presented in this figure, a re-scaling of the Kriging map is also applied in order to restore the scaling transformation as presented in Eq. (2).

3. Results and discussion

3.1. Model validation

In this section, the performance of the RIO interpolation model is examined in a so called jack-knife or 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 interpolation methodology as

Fig. 14. Schematic outline of the re-trending procedure applied to grid cells. The panel left gives the transformation for four different grid cells. The diamonds are the result of the Kriging interpolation. The 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 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).

available in, e.g., commercial GIS packages. The IDW scheme is retained in this analysis since it was commonly used in the past as the standard technique for the production of air quality maps. For the Ordinary Kriging, the special RIO technique of detrending before Kriging and re-trending afterwards is not applied. Furthermore, the correlation functions as described in Section 2.4 are not applied but a unique variogram is determined for each time step based on the spatial distribution of the corresponding input values.

Fig. 15 presents the results of the cross-validation with the Ordinary Kriging and the RIO model for ozone. From the figure, it is clear that the Ordinary Kriging model produces larger RMSE values in the majority of the sampling sites compared to the RIO model. Thereby, the RIO model results are in general less biased compared to the standard interpolation technique. Similar validation results are presented for NO_2 and PM_{10} in Figs. 16 and 17, respectively. 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 4. 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 values correspond to rather accurate estimates (relative errors of 11% O_3, 26% NO_2, 28% 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).

3.2. Interpolation maps

In this section, the performance of the interpolation model is not evaluated on a statistical basis as in the previous section, but by means of actual annual air quality maps for 2006 obtained with the two 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. Note that in this hour-by-hour interpolation RIO is using the hourly parameterisations as derived in the statistical analysis (see Fig. 13 in Section 2.4).

In Fig. 18, the annual mean ozone 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 ozone concentrations fields at places where no monitoring data is available. Standard techniques have to rely only on the monitoring data themselves and their relative distances and correlations.

Similar maps for the 2006 annual mean concentration of NO_2 and PM_{10} are presented in Figs. 19 and 20. For NO_2, the difference between the

Fig. 15. RMSE (left) and bias (right) for max 1 h ozone values. Results are presented for the RIO model (upper bars) and for standard interpolation technique Ordinary Kriging (lower bars). Summer ozone data is used from the 41 ozone-monitoring stations in Belgium between 2003 and 2006.

Fig. 16. Same as in Fig. 15 but for the max 1 h NO_2 values. All available data between 2003 and 2006 is used in the validation. Note that the indicators of station 41B003 go beyond the range of the plot.

Fig. 17. Same as in Fig. 16, but for the daily mean PM_{10} values.

Ordinary Kriging map and the RIO map is even more pronounced. This is 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 (like Ordinary Kriging) would extrapolate these enhanced NO_2 levels from the urban sites to the rural areas. RIO however 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. The same conclusions can be drawn for the PM_{10} maps. Here, the effect is even more explicit than for the NO_2 case. It turns out that almost all of the PM_{10} sampling sites operational in 2006 can be classified as urban, industrial or traffic stations. Based on those measurements only, standard interpolation techniques will assess

Table 4
Model performance indicators for the IDW, OK and RIO model and for the three pollutants O_3, NO_2 and PM_{10}

Model	O_3 (max 1 h, summer) ⟨90.9⟩			NO_2 (max 1 h) ⟨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

Values are averaged over all available Belgian monitoring locations. The mean value for the corresponding time averaging value (e.g., max 1 h) is given between brackets ⟨ · ⟩. All values are in $\mu g\,m^{-3}$.

Fig. 18. Annual mean ozone 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.

Fig. 19. Annual mean NO_2 map for the year 2006 obtained by the RIO model (left) and Ordinary Kriging (right).

Fig. 20. Annual mean PM_{10} map for the year 2006 obtained by the RIO model (left) and Ordinary Kriging (right).

PM_{10} pollution over a large part of the Belgian territory according to these enhanced PM_{10} levels. The RIO model however restricts, the high PM_{10} levels to the urbanised regions, whereas lower levels are introduced in the rural areas. This result radically changes the outcome of any assessment of the population exposure or areas with concentration exceedances of the European limit values.

3.3. Uncertainty

When interpolation maps are used for air quality assessments, it is important to have an idea of the uncertainty of the presented maps. Therefore, the errors introduced by the RIO interpolation scheme are further explored in this Section.

3.3.1. Kriging error

When solving the Ordinary Kriging equations, a value for the error variance can be obtained at the same time (Isaaks and Srivastava, 1989). This error variance σ_R^2 is a measure for the uncertainty (1 σ_R) of the interpolation result. The parameters that determine the spatial pattern of σ_R^2 are obtained by the Kriging solution. The overall scale of σ_R^2 is fixed by the variance σ^2 of the so-called random function model which is the only quantity that needs to be specified. A most obvious choice would be to take the average of the variances of the time series obtained in the different monitoring stations over the country. This choice would be acceptable if, and only if spatial patterns of air pollution around all monitoring stations would be homogeneous. However, due to local and regional pollution sources and sinks, the air pollution phenomenon does not obey at all the necessary requirement of spatial homogeneity and as a consequence, a unique variance σ^2, representative for any location cannot be used. To tackle the problem, a β-depended variance $\sigma^2(\beta)$ will be introduced in the σ_R^2 calculation, based on the trend analysis of the variance as presented in Fig. 8. These plots indicate that the variance is positively correlated with the β-parameter and that (with an exception for ozone) increased β-values correspond to increased σ's. This dependency and the fact that $\sigma^2(\beta)$ is now used as a scaling factor for $\sigma_R^2(\beta)$ makes that for regions with a low β-value, the Kriging interpolation error is smaller compared to a result that would have been obtained with a unique σ averaged over all stations whereas in urbanised and industrialised regions, characterised by higher β-values, the Kriging error is higher.

3.3.2. Re-trending error

Besides the error due to the Kriging interpolation scheme, an error is also introduced by applying the "trend function" during the re-trending procedure. As this trend function is only a best estimator for the β-dependency of the pollution data (see Fig. 21), the use of it during re-trending gives rise to an additional uncertainty. Therefore, the error on the polynomial fit through the scattered data is used as an additional and cell specific error term $\sigma_{\text{trend}}(\beta)$.

Fig. 21. Trend function (see left panel of Fig. 5) with indication of the error boundaries (dashed lines) for max 1 h NO_2 concentration as a function of the β-parameter.

Fig. 22. Relative error map for a max 1 h NO_2 interpolation result. The error is composed of the standard Kriging error where a β-dependent σ is taken into account and an error due to the uncertainty of the trend function.

3.3.3. Total error map

The total error attributed to the final RIO interpolation result is then calculated as

$$\sigma_{\text{tot}}(\beta) = \sqrt{\sigma_R^2(\beta) + \sigma_{\text{trend}}^2(\beta)}, \tag{3}$$

where both the Kriging error $\sigma_R(\beta)$ and the error introduced by applying the trend function $\sigma_{\text{trend}}(\beta)$ are taken into account.

As an illustration, Fig. 22 presents a relative error map for a maximum 1-h NO_2 map of a particular day. Low interpolation errors (~18%) are found in the vicinity of the monitoring stations whereas higher values (~25%) are found far away from any sampling site. This typical circular error pattern is predominantly caused by the Kriging error. This pattern, however, is locally distorted according to the corresponding land use characteristics introduced by the β-dependencies of both error terms in Eq. (3). It is interesting to note that the relative higher errors in the Northeastern part of the country are caused by the fact that for the three monitoring stations in this region no valid measurement values were present for this particular day. When measurement data is available in this region, the error values are further reduced.

As a closing remark, it must be said that the introduction of a supplementary driver (de- and re-trending) on top of the Ordinary Kriging for the spatial interpolation, slightly increases the overall total error on the final result. This disadvantage, however, is largely compensated by the benefits of a much more reliable spatial prognosis obtained in the RIO scheme by introducing the de- and re-trending intelligence, as it is driven by the nature of the pollution itself.

3.4. Operational application

From the viewpoint of applicability, it is emphasised again that RIO is able to describe different time aggregation values of one pollutant. For the Belgian Interregional Environment Agency (IRCEL), daily maximum 1-h values, maximum daily 8-h means and daily mean values are implemented in the model code together with the possibility of an hour-by-hour interpolation of a set of (real time) hourly sampling values. This latter option allows maximal flexibility and the possibility for calculation of any time aggregation value over any time window from a 1-h map up to a map of annual mean concentrations. The former types of maps are published in real-time on-line on the website (www.irceline.be) in order to visually inform the authorities and the public about the current and forecasted air quality levels. The annual mean concentration maps are used for assessing air quality levels and checking the compliance with limit and target values within the context of the reporting obligations in the EU directives.

Finally, it is stressed that the computational cost of RIO is many orders of magnitude lower than the cost needed by deterministic air quality models. The Fortran code of the RIO model produces air quality

maps on a standard PC within 2 s. This extremely low computational time (compared to deterministic air quality models) is an important asset for the RIO model in the operational setup at IRCEL.

4. Conclusions

In this paper, the RIO model is presented as an "intelligent" interpolation model for air pollution data. The model was initially developed for the spatial interpolation of ozone data and the population density was used as a spatially driving force. In this work, the model is extended for NO_2 and PM_{10} and a land use indicator based of the CLC data set is introduced. This new land use indicator (the β-parameter) 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. Especially for NO_2, the latter effect is substantial.

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 mostly 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 in a cross-validation analysis that the RIO technique, compared to standard interpolation techniques, has a clear better result for all performance indicators (RMSE, bias and MAE). In this respect, the small bias values of the RIO model were most striking. 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.

At present, the RIO model is being extended along the same lines to incorporate SO_2 measurements. Furthermore we will explore the possibilities to use satellite retrievals such as the Aerosol Optical Thickness parameter as an alternative driving force for the CLC β-indicator in the RIO interpolation scheme for PM_{10} (Koelemeijer et al., 2006).

The accuracy of the RIO model in combination with its extremely low computational cost makes it a valuable tool for the Belgian Interregional Environment Agency IRCEL in their air quality assessments both in real-time on-line web applications as in checking compliance with EU air quality limit values and population exposure indices.

References

Almeida, S.M., Pio, C.A., Freitas, M.C., Reis, M.A., Trancoso, M.A., 2005. Source apportionment of fine and coarse particulate matter in a sub-urban area at the Western European Coast. Atmospheric Environment 39, 3127–3138.

Arain, M.A., Blair, R., Finkelstein, N., Brook, J.R., Sahsuvaroglu, T., Beckerman, B., Zhang, L., 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., 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.

EC, 1996. Council Directive 96/62/EC of 27 September 1996 on ambient air quality assessment and management.

EC, 2004. Commission Decision 2004/461/EC of 29 April 2004 laying down a questionnaire to be used for annual reporting on ambient air quality assessment under Council Directives 96/62/EC and 1999/30/EC and under Directives 2000/69/EC and 2002/3/EC of the European Parliament and of the Council.

EC, 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

EEA, 1995. CORINE Land Cover Project, published by Commission of the European Communities.

EMEP, 2007. WebDab 2007 emission database, EMEP, Co-operative programme for monitoring and evaluation of long range transmission of air pollutants in Europe. 〈http://webdab.emep.int〉.

Erisman, J.W., Schaap, M., 2004. The need for ammonia abatement with respect to secondary PM reductions in Europe. Environmental Pollution 129, 159–163.

Fuller, G.W., Green, D., 2004. The impact of local fugitive PM_{10} from building works and road works on the assessment of the European Union Limit Value. Atmospheric Environment 38, 4993–5002.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., 2006. Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. Journal of Environmental Monitoring 8, 1129–1135.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Jiménez, P., Parra, R., Gassó, S., Baldasano, J.M., 2005. Modelling the ozone weekend effect in very complex terrains:

a case study in the Northeastern Iberian Peninsula. Atmospheric Environment 39, 429–444.

Koelemeijer, R.B.A., Homan, C.D., Matthijsen, J., 2006. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. Atmospheric Environment 40, 5304–5315.

Mensink, C., Cosemans, G., Pelkmans, L., 2005. Dynamic modelling of transient emissions and concentrations from traffic street canyons. International Journal Environment and Pollution 25, 118–130.

Mensink, C., Deutsch, F., Janssen, L., Torfs, R., Vankerkom, J., 2007. Comprehensive Modelling of PM_{10} and $PM_{2.5}$ Scenarios for Belgium and Europe in 2010. Numerical Methods and Applications 4310, 466–474.

Pinder, R.W., Adams, P.J., Pandis, S.N., 2007. Ammonia emission controls as a cost-effective strategy for reducing atmospheric particulate matter in the Eastern United States. Environmental Science and Technology 41, 380–386.

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

Sillanpää, M., Hillamo, R., Saarikoski, S., Frey, A., Pennanen, A., Makkonen, U., Spolnik, Z., Van Grieken, R., Braniš, M., Brunekreef, B., Chalbot, M.C., Kuhlbusch, T., Sunyer, J., Kerminen, V.M., Kulmala, M., Salonen, R.O., 2006. Chemical composition and mass closure of particulate matter at six urban sites in Europe. Atmospheric Environment 40, 212–223.

Sillman, S., 1999. The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments. Atmospheric Environment 33, 1821–1845.

Tombette, M., Sportisse, B., 2007. Aerosol modeling at a regional scale: model-to-data comparison and sensitivity analysis over Greater Paris. Atmospheric Environment 41, 6941–6950.

Vanderstraeten, P., Willette, F., Dumont, G., 1996. Working day versus non-working day ambient ozone concentrations in Brussels and in Belgium. In: Borrell, P.M., Borrell, P., Kelly, K., Cvitaš, T., Seiler, W. (Eds.), Proceedings of the EUROTRAC Symposium, vol. 1, pp. 899–903.

Van Leeuwen, E.P., Draaijers, G.P.J., Erisman, J.W., 1996. Mapping wet deposition of acidifying components and base cations over Europe using measurements. Atmospheric Environment 30, 2495–2511.

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., 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, L., Van Loon, M., 2007. Evaluation and intercomparison of ozone and PM_{10} simulations by several chemistry-transport models over 4 European cities within the CityDelta project. Atmospheric Environment 41, 173–188.