H14-247

MAPPING TRAFFIC ATMOSPHERIC EMISSIONS FOR EPIDEMIOLOGICAL STUDIES USING ATMOSPHERIC DISPERSION MODELS AND GEOSTATISTICAL METHODS: A CASE STUDY

Sergio Teggi[1], Grazia Ghermandi[1], Sara Fabbi[1], Carlotta Malagoli[2], Marco Vinceti[2] Luisa Guerra[3], Antonella Sterni[3] ,Giuseppe Maffeis[4]

[1]Department of Mechanical and Civil Engineering, University of Modena and Reggio Emilia, Modena, Italy
[2]Research Center for Environmental, Genetic and Nutritional Epidemiology, Dep. of Public Health Sciences, University of Modena and Reggio Emilia, Modena, Italy
[3]ARPA Emilia Romagna, Sezione di Modena, Modena, Italy. [4]TerrAria s.r.l., Milano, Italy

Abstract: In some cases, epidemiological studies require the air pollutant concentrations at the exposure points. In these cases air dispersion models represent a very important tool. When additional points of exposure are inserted or when some exposure points must be relocated, spatial interpolators can be used in place of new runs of the air dispersion model. In this work the uncertainties and the problematic related to spatial interpolation methods are inspected. The case studied is based on an epidemiological study aimed to study the risk of childhood leukemia associated with benzene exposure due to traffic emissions. The concentration values of benzene computed by the atmospheric dispersion model ADMS are taken as reference and compared with the concentration values computed using several interpolation methods and additional data sets of concentrations computed by ADMS in the same area. The comparison is done following two approaches: the summary statistics of the differences and the correctness of the assignment of the exposure points to the concentration categories used in the epidemiological study. These comparisons show that the values computed by the interpolators are very problematic: important differences and categories assignment and categories uncertainties were found. The main conclusion of this work is that the use of interpolators must be done with extreme caution. Moreover, it is highlighted the importance and the potential pitfalls of exposure modelling methodologies when assessing the health effects of environmental pollutants

***Key words**: Spatial interpolation, exposure level, epidemiological study, traffic emissions, ADMS*

INTRODUCTION

Air concentration of pollutants emitted by traffic strongly influences population exposure and its assessment is of major importance for epidemiological studies.Usually, population exposure is requested for a large number of locations, while direct concentration measurements are typically done at few ground stations. Air dispersion models offer a unique opportunity to estimate the health risk associated with exposure to atmospheric contaminants, while their use in the epidemiologic literature is still limited.

Air dispersion models estimate pollutants concentration at specific locations (receptors) or at locations distributed on spatial grids. In the first case the concentration at the exposure points is computed using interpolation methods. In the second case the receptors can be located on the exposure points. Often, some exposure pointshave to be relocated or additionalexposure points are requested: in these cases air dispersion model must run again, or the concentration on the new locations can be computed by spatial interpolation of the values already available. Therefore, interpolation methods can play an important role in exposure calculation and the reliability assessment of the interpolated values is crucial.

In this work we investigated the performance of several spatial interpolation methods in the framework of an epidemiological study. This study (Malagoli et al., 2011, Vinceti et al., 20011) is aimed to evaluate the risk of childhood leukemia associated with benzene exposure due to traffic emissions in the Modena province (Italy). The concentration of benzene at the exposure points are computed using the ADMS (McHugh et al., 1997) air dispersion model. We computed new concentration values at the exposure points using several interpolation methods and the ADMS values computed at different locations. The performances of the interpolators were evaluated by comparing the concentration values computed directly on the exposure point by ADMS with those interpolated. Moreover, we made some evaluations of the importance of the interpolation errors for the epidemiological study.

ADMS SIMULATIONS AND STUDIED AREA

The epidemiological study done by Malagoli et al. (2011) (see also Vinceti et al., 2011) was based on the air concentrations of benzene from vehicular traffic in the province of Modena and evaluated using two different computational models: ADMS and CALINE. In this work only the ADMS simulation are considered.The assessments of benzene levels were carried out in an area were the risk of exceeding the limit value and/or alert thresholds is high and therefore it is necessary to provide long-term plans and programs. This area is located inthe northern part of the Modena Province (Po Valley, Italy), has an extension of 55 km x 60 km and comprises the most populated zones of the Province.

Emissions of benzene from vehicular traffic were calculated using estimates of road traffic flows insistent on the main roads of the Province and a database of emission factors expressed in g/km per vehicle.The emission factors of benzene have been derived from the database “transport data from 1990-2007” developed by ISPRA, using emissions for the most recent year available (2007), resulting from the program COPERT IV (URL: http://www.emisia.com/copert/General.html).The average emission factor associated to the vehicle class (light or heavy) of the vehicular traffic flows, has been calculated as a weighted average based on the product between the number of vehicles registered in the province of Modena to 31/12/2007 (source: ACI, Automobil Club d’Italia) and theirs annual average mileage (source: ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale). The appropriate drive cycle has been assigned to every arch of road network, depending on the type of road and traffic.

ADMS requires, as input, an annual emission value (for example ton/year) for each road. This value has been calculated by multiplying the emission factor associated to the road by the number of vehicles and the road length.Traffic flows were expressed as morning peak hour flows. The daily emission was computed using the appropriate hourly factors derived from measurements of daily traffic. The input annual emission has been calculated assuming the same traffic flow for all days of year.The ADMS input annual emissions were finally modulated by an input modulations file, containing the previous hourly factors.In addition, ADMS requires as input a single set of meteorological data representative of the entire study area. This data set consists on both measured quantities (temperature, wind speed and direction) and derived quantities (e.g. Monin Obukhov length and stability class). This meteorological data set was obtained by the preprocessor CALMET, deployed at ARPA-SIM (Hydro Meteorological Service), representing a suburban domain (URL: http://www.arpa.emr.it). The simulationsweredonefor the whole year 2006 with a time step of one hour and assuming that the turbulence contribution due to urban canyons was neglected and considering a flat terrain (Po valley).The output concentrations considered in this work are the annual mean values calculated from hourly concentrations at different locations (see below).

DATA SET

The data set used for this study is composed by four sets of ADMS simulations: a) 2077 points located at the regular grid (SA, Figure 1-left); b) 19777 points located at the intelligent grid (SB, Figure 1-center); c) 4220 points obtained by spatial aggregation of SB points (SC); 240 validation points located at the exposure points (or receptors, SR, Figure 1-right).The spatial aggregation of SB (SC data set) is done in order to reduce the spatial distribution dissimilarity between regular and intelligent grids. The aggregation is done using a blocking method: the domain is dived into 150 m x 150 m cells (blocks) and then all the points contained in a cell are replaced by a point located at mean position and with the mean concentration value of the original points.

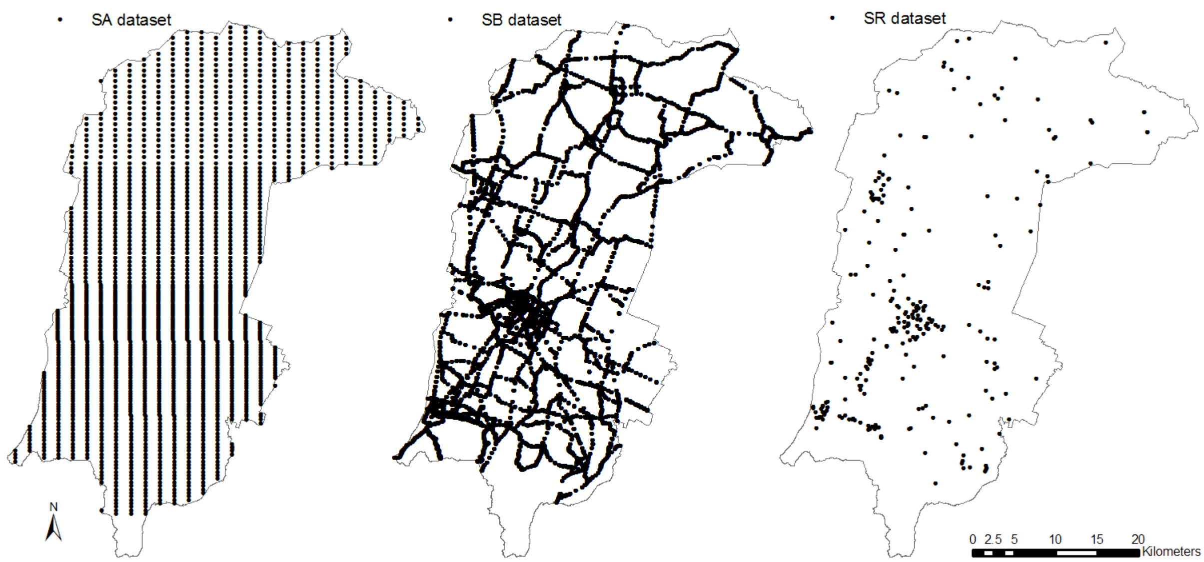

Figure 1. Principal datasets used: SA

Table 1shows the statistics of the four datasets. It can be observed that the datasets SA and SR are composed by values lower than those of SB and of SC datasets. This is due to the higher proximity of SB and SC to the emission sources (roads). Moreover, it is important to notice that none of the datasets show normal distribution (high Kurtosis and Skewness values, high differences between means and medians).

Table 1. Summary statistics of the datasets

Data Set	SR	SA	SB	SC	S1	S2
N	240	2078	19777	4281	21855	6359
Min [(a)]	0.02	0.	0.02	0.02	0.	0.
Max [(a)]	4.59	4.97	11.13	6.44	11.13	6.44
Mean [(a)]	0.40	0.13	0.66	0.54	0.61	0.41
St. Dev. [(a)]	0.53	0.24	0.78	0.66	0.76	0.59
Median [(a)]	0.24	0.08	0.39	0.27	0.33	0.16
Kurtosis [(b)]	19.7	117.1	12.6	8.9	13.3	12.7
Skewness [(b)]	3.7	8.7	2.7	2.5	2.8	3.0
[(a)] Values in µg / m^3 ; [(b)] Kurtosis and skewness of a normal distribution are both 0.						

INTERPOLATION

In this work five spatial interpolators, between the most quoted in literature (Isaaks and Srivastava, 1989; O'Sullivan and Unwin, 2003), wereused to compute the concentration at the exposure points (SR dataset): Voronoj polygons (VO) (actually, this is not a true interpolation method); Inverse Distance method (ID); local Linear Interpolation (LI, first order); S-Pline

(SP); Kriging (KR); Co-Kriging (CK). All these interpolators, but CK, have been applied to three data sets: the SA data set; the union of SA with SB (S1, Table 1); the union of SA with SC (S2, Table 1). CK was applied in two different configurations:in both the cases the SA data set was used as the principal variable, whilethe correlated variable was set to SB in the first case (CK1) and to SC in the second case (CK2).

COMPARISON METHODOLGY

The validation of the concentrations calculated by the considered interpolators is based on the comparison with the concentrations directly evaluated by ADMS at the exposure points. This statistical comparison is made difficult by the fact that the concentration datasets are not normally distributed and in some cases they are multimodal. For this reason we computed separate statistics for each of the concentration categories (I_i) used in the reference epidemiological study: I_0 = [0, 0.1] $\mu g/m^3$, I_1 =]0.1, 0.5] $\mu g/m^3$, I_2 =]0.5, 1.0] $\mu g/m^3$ and I_3 =]1.0, ∞] $\mu g/m^3$. The first category has been added with respect to those used in the reference study in order to consider a null (insignificant) concentration level. For each of these categories several statistical parameters have been computed using as reference the ADMS data set :

- N_{Ci}: the number of exposure points correctly assigned by a given interpolator to the categoryI_i;
- Δ_i, $RMSD_i$: the bias and the root mean square difference between the concentrations interpolated and those computed by ADMS of the exposure points belonging to the ADMS data set I_i.
- Median, 0.16 and 0.84 quantiles (these quantiles include the 68% of the cases) of category number, determined by the interpolated value, of the exposure points belonging to the ADMS I_i data set (this statistic inspects the category assignment error);

RESULTS AND DISCUSSION

The validation results of the 17 different cases considered are summarized in Table 2 and in Figure 2. Table 2 reports the values of the statistical parameters N_{Ci} , Δ_i and $RMSD_i$. Figure 2 shows the plots of medians and of 0.16 and 0.84 quantiles (error bars, assignment error) of category assignments.

It can be observed that in most of the cases all the interpolators furnished similar performances when applied at the same dataset.

The exposure points with “not significant” concentration levels (I_0) are better approximated using the SA dataset only. The use of the other two datasets produces overestimation of the concentration (high Δ_0 values). In this case the VO method performs slightly better than the others. When the interpolation is done using the SA dataset more than 60% of the exposure points belonging to the I_0category are correctly assigned (78% in the VO case). This percentage lowers to less than 50% when using the S1 or the S2 datasets. The exposure points not correctly classified are mostly assigned to the I_1category (Figure 2).

Table2. Comparison results. All the concentration values are in µg / m³

Category	I_0			I_1			I_2			I_3		
$Range_{SR}$	0.06 ± 0.02			0.24 ± 0.11			0.68 ± 0.13			1.79 ± 0.88		
N_{SR}	52			127			41			20		
	N_{C0}	Δ_0	$RMSD_0$	N_{C1}	Δ_1	$RMSD_1$	N_{C2}	Δ_2	$RMSD_2$	N_{C3}	Δ_3	$RMSD_3$
VO-SA	41	0.03	0.14	92	0.01	0.21	7	-0.03	0.63	7	-0.66	1.5
LI-SA	33	0.03	0.09	106	-0.01	0.12	11	-0.06	0.49	9	-0.84	0.91
SP-SA	33	0.03	0.06	107	-0.01	0.12	11	-0.05	0.5	10	-0.74	0.98
ID-SA	33	0.03	0.06	106	0	0.12	12	-0.05	0.5	10	-0.76	0.97
KR-SA	33	0.02	0.04	112	-0.02	0.11	14	-0.11	0.48	6	-0.83	0.93
VO-S1	26	0.06	0.1	79	0.25	0.37	17	0.51	0.62	16	0.08	0.82
LI-S1	21	0.12	0.21	52	0.42	0.39	8	0.77	0.67	18	0.37	0.9
SP-S1	17	0.04	0.06	91	0.15	0.19	30	0.21	0.41	14	-0.25	0.71
ID-S1	14	0.15	0.15	37	0.48	0.37	9	0.81	0.74	18	0.29	0.94
KR-S1	14	0.05	0.06	79	0.23	0.24	23	0.39	0.53	17	0.02	0.81
VO-S2	24	0.09	0.15	69	0.32	0.4	12	0.75	0.89	18	0.28	0.86
LI-S2	18	0.11	0.15	48	0.4	0.32	8	0.68	0.63	18	0.16	0.9
SP-S2	16	0.1	0.12	56	0.37	0.38	12	0.66	0.66	19	0.22	0.84
ID-S2	14	0.14	0.15	50	0.4	0.33	9	0.7	0.61	18	0.22	0.86
KR-S2	14	0.09	0.1	65	0.34	0.38	13	0.67	0.8	19	0.24	0.98
CK1	14	0.02	0.06	115	0.01	0.12	16	0.01	0.54	11	-0.56	0.83
CK2	14	0.03	0.06	112	0.04	0.13	13	0.09	0.64	12	-0.59	0.93

Similar considerations can be done for the estimates of exposure points belonging to the I_1category. In this case the percentage of exposure points correctly assigned suing the SA dataset is 88% for the KR interpolator and 84% for the others but VO (72%). The same performances are achieved using the CK method. The very low values of Δ_1 and of $RMSD_1$ and the plots on Figure 2 indicated that the exposure points not correctly assigned were mostly assigned to the I_0category (underestimation). The use of the S1 and of the S2 data sets produces percentages of correct assignment lower than 70% and Figure 2 indicates that the remaining exposure points are preferentially assigned both to the I_1category and to the I_2category. The estimations for the exposure points of the third category, I_2 , are the most problematic. In most of the cases the percentage of exposure points correctly assigned to the category is lower than 42%. The only two exceptions are the SP-S1 (73%) and the KR-SP1 case (56%). The estimation obtained using the SA dataset are spread in the I_1, I_2 and I_3categories. The concentrations obtained with the S1 and S2 datasets are overestimated in most of the cases.

Concerning the category with the higher values, I_3 , the better performance was obtained using the S1 and the S2 datasets. Excluding the VO-S1 and the SP-S1 cases, the percentages of correct assignments range from 85% to 95%. These high percentage values are probably due to the fact that using the S1 and S2 datasets all the interpolators overestimate the concentrations and that the I_3 interval is not top limited. The performances of the interpolators applied to the SA dataset and using the CK methods are considerably lower. For example, the percentage of exposure points correctly assigned is lower than 60% in all the cases.

Moreover, it can be observed that neither the spatial aggregation of the SB dataset (S2) nor the use of Co-Kriging method did not produces appreciable difference with respect to the S1 dataset or the classic Kriging method.

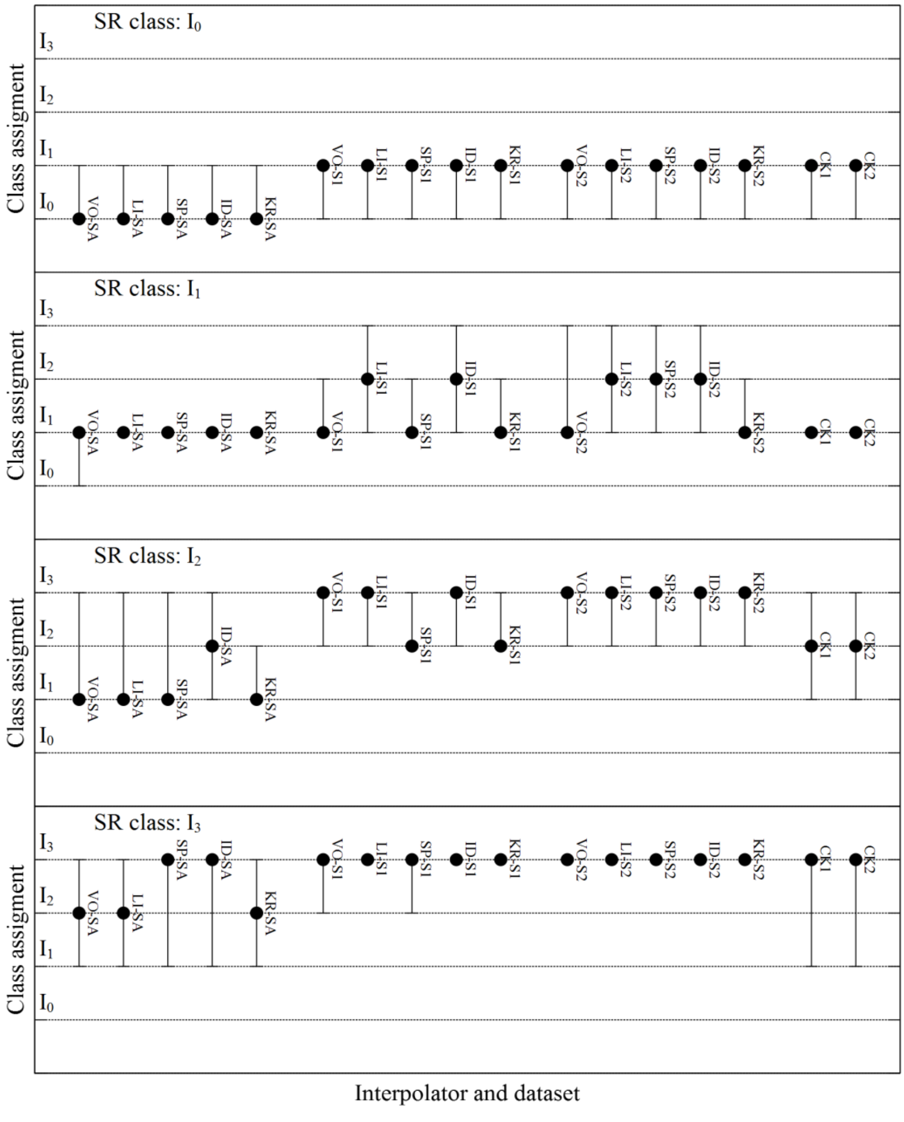

Figure 1. Statistics of the category assignments of the exposure points to the different concentration categories (see text).

Finally, we tried to make some considerations on the effects that the above uncertainties would have on epidemiological studies like the one that originated this work. The exposure misclassifications arising from the different interpolation methods examined in the present study would have marked effects on computation of health risks attributable to the benzene emissions. In addition, since the misclassification of exposure occurring with the different methods would be differential, i.e. not characterized by a uniform pattern across the different strata I_i, these effects would not be a simple reduction in amount and statistical stability of the risk estimates, as expected in case of non-differential exposure misclassification, but the induction of severe bias in estimates computed for specific strata. In fact, the baseline exposure category (I_0), to which all the remaining groups are compared, would be substantially narrowed (Figure 1), as it would occur to the third and fourth exposure stratum (I_2 and I_3), due to the shift of several subjects to the upper one. This differential misclassification would introduce instability in calculation of all risk estimates, due to the narrowing of the reference category I_0, and will substantially alter the risk estimates computed for I_2 and particularly the I_3 strata, also considerably affecting the possibility to detect dose-response relations in leukemia risk.

CONCLUSIONS

The use of atmospheric dispersion modelling is an important tool for the calculation of concentration levels of pollutants at the exposure points required by epidemiological studies like the one taken as reference in this work. When additional exposure points are required or when the exposure points have to be relocated to approaches are commonly used: to make new runs of the model or to use spatial interpolators. This work was addressed to inspect the problematic related to the second choice.

The comparisons between the concentration values at the exposure points directly computed by ADMS and those obtained by the considered interpolators lead to the conclusion that the use of interpolators must be done with extreme caution. The numerical comparisons of the two set of data showed substantial differences (bias and root mean square differences). Using the interpolated values, the assignment to the exposure categories, utilized in the epidemiological study, showed important discrepancies and uncertainties. These considerations can be carried out for all the interpolation methods used (except for the Voronoj method in pejorative sense) and for all the data sets used.

Overall, these considerations highlight the key importance and the potential pitfalls of exposure modelling methodologies when assessing the health effects of environmental pollutants such as benzene.

REFERENCES

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

Malagoli, C., Cherubini, A., Maffeis, G., Sterni, A., Guerra, L., Fabbi, S., Teggi, S., Ferretti, E. and Vinceti, M., 2011: Stima della esposizione a benzene da traffico veicolare nelle provincie di Modena e Reggio Emilia, XII Conferenza Nazionale di Sanità Pubblica, Rome, 12-15 October.

McHugh, C., Carruthers, D. and Edmunds, H., 1997, ADMS-Urban: an air quality management system for traffic, domestic and industrial pollution, *International Journal of Environment and Pollution*, 8, 666-674

O'Sullivan, D. and Unwin, D. J., 2003, Geographic Information Analysis, John Wiley & Sons

Vinceti, M., Malagoli, C., Sterni, A., Guerra, L., Cherubini, A., Maffeis, G., Fabbi, S., Teggi, S., Ferretti, E., De Girolamo, G., Palazzi , G. and Paolucci, P., 2011: Esposizione a benzene da traffico e leucemia infantile: influenza delle modellistiche di dispersione atmosferica sulla valutazione del rischio, XII Conferenza Nazionale di Sanità Pubblica, Rome, 12-15 October.