Environ Model Assess (2015) 20:583–597
DOI 10.1007/s10666-015-9445-7

Emission Data Uncertainty in Urban Air Quality Modeling—Case Study

Piotr Holnicki · Zbigniew Nahorski

Received: 21 February 2014 / Accepted: 14 January 2015 / Published online: 10 February 2015

Abstract Air pollution models are often used to support decisions in air quality management. Due to the complexity of the forecasting system and difficulty in acquiring precise enough input data, an environmental prognosis of air quality with an analytical model of the air pollution dispersion is burdened with a substantial share of uncertainty, especially as regards urban areas. To ignore the uncertainty in the modeling would lead to incorrect policy decisions, with further negative environmental and health consequences. This paper presents a case study which shows how emission uncertainty of air pollutants generated by the industry, traffic, and the municipal sector relates to concentrations measured at receptor points. The computational experiment was implemented in the Warsaw metropolitan area, Poland. The main source of this adverse environmental impact is the transportation system, including the transit traffic. The Monte Carlo technique was used for assessing the key uncertainty factors. Several types of pollution species that are characteristic for the urban atmospheric environment (e.g., PM_{10}, $PM_{2.5}$, NO_x, SO_2, Pb) were included in the analysis. The results show significant spatial variability of the modeled uncertainty. The reason of this variability is discussed in detail. It depends not only on the category of the emission source but also on the contributing emission sources and their quantity.

Keywords Air quality modeling · Emission inventory · Urban air pollution · Uncertainty analysis

P. Holnicki (✉) · Z. Nahorski
Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland
e-mail: holnicki@ibspan.waw.pl

1 Introduction: Air Pollution Transport Modeling

The strategies which are aimed at reducing the pollution to the required air quality standards are usually based on air pollution dispersion models. A decision support system, intended to help in preparation of abatement plans, often takes a form of an integrated assessment model (IAM) that combines the classical pollution transport model with certain economic, ecological, technological, and other constraints and standards [1–5]. The model provides a tool for conducting a comprehensive analysis of a given environmental policy in order to explore potential strategies to reduce emissions, eliminate violations of air quality limits, and reduce the population exposure. Implementation of such policy often involves a cost-effective approach or optimization [4, 6]. Irrespective of how complex such a system might be, its main component is usually the air pollution dispersion model, with its other components including certain constraints and limits.

Efficient control strategies cannot be elaborated well without clear understanding of imprecisions and uncertainties of the modeling process. To quantify possible ecological, economic, or health benefits of the emission abatement process, the incremental contribution of the respective groups of emission sources to ambient concentrations must be estimated with a reasonable accuracy. Due to a very complex and multidisciplinary structure of such systems, there are many causes leading to imprecision and uncertainty, which include (a) the input data (mainly emissions, meteorological data, boundary conditions), (b) the structure of the mathematical model (simplifications and parameterizations of physical and chemical processes), (c) the numerical scheme, and (d) the uncertainties present across all stages of modeling process and in the interactions with policymakers [7]. All simplifications and uncertainties in the modeling process impact the robustness of the final results [4]. Most of uncertainty studies focus on those arising from the input data and the model parameterizations [8, 9, Warchałowski 2012, Personal communication, Warsaw University of Technology].

Simplifying assumptions or parameterizations becomes a source of certain conceptual uncertainty which is also transposed onto the final results [1, 5]. In particular, such uncertainty is present in the process of deriving trajectories based on the Lagrangian approach or in the sub-grid effects of the Eulerian models [10, 11]. Moreover, due to certain model performance constraints, some atmospheric processes are parameterized or described in a rather simplified way. For example, the height of the mixing layer and atmospheric stability are usually evaluated using imprecise heuristic procedures, yet another source impacting the level of the output uncertainty. However, numerous studies have already revealed that the biggest uncertainty (measurement or estimation errors) is caused by emission inventory and meteorological data (see 4 in [10]).

The main sources of variability (temporal, spatial, or inter-individual differences of input data) and uncertainty (imprecise information or lack of information about unknown quantities) must be identified and assessed [12, 13] to compliment modeling studies. This information is needed for investigating effective strategies of emission abatement and improving air quality. First of all, uncertainty assessment gives an opportunity to check the quality of the modeling results and improves the precision of the modeling. It can also increase policymaker's confidence in robustness of results and recommendations. Finally, such a complex analysis can improve a general stakeholder and public confidence in scientific research [4, 7].

The problem related to air pollution in urbanized areas is ranked high on the list of priority environmental concerns [4, 6, 14]. Urban-scale pollution estimations are a sophisticated modeling issue for computational reasons because of complexity of emission field, the complicated building orography, and wind-field effects. Also, the uncertainty analysis of emission data is challenging especially in the case of urbanized or industrial areas [2, 8, 15]. Emission inventory of such areas usually includes different emission source categories, each with specific emission parameters. Here, the emission field means the spatial concentrations of a large number of emission sources with different technological characteristics, fuel type (and related fuel parameters), composition of polluting chemicals, and emission intensities, with the consequence being the varying range of emission uncertainty. The available emission data are not accurate because of inventory level uncertainties. The emissions of major power plants can be treated as relatively accurate because of the well-described parameters of the combustion process and the fuels used. On the other hand, the emission data for residential areas or the urban transport system are usually based on some aggregated and averaged information related to fuel consumption, and then to disaggregation parameters. These categories of data neither reflect precisely the real temporal variability nor chemical constitution of polluting chemicals, and as such are very uncertain. In complex uncertainty analysis, the correlation between the various pollutants generated by a source must be taken into account [16, 17]. Moreover, the variety of the primary pollutants generates secondary compounds through chemical transformation processes, often even more dangerous for the environment. Due to the excessive population density, the exposure to the urban air pollution is a crucial factor associated with numerous adverse health effects. In particular, many research results indicate that a considerable harm of public health is caused by traffic emissions of fine particulate matter, for example PM_{10} or $PM_{2.5}$ [2, 4, 8, 12, 15, 16].

The analysis presented in this study concentrates on the impact which the major traffic emissions have on the uncertainty of the resulting concentrations. The Monte Carlo algorithm was used as the key tool to assess the level of uncertainty. The discussed issues include the spatial distribution of the output level of uncertainty, its correlation with road structure, and the intensity of traffic. The key factors determining the output level of uncertainty are elaborated in detail.

The main point of this paper, as presented in the next sections, is the assessment of uncertainty in the modeled forecasts in relation to the input uncertainty of the emission data set. This analysis is based on certain historic air quality forecasts for Warsaw, as shown in [18]. This section shortly recalls the basic facts concerning the model, the emission data, computational parameters, and the air quality results. Also, we added information on the model performance and a comparison of results with observations. These outcomes form a basis for the complex uncertainty analysis which is presented in Section 3. The computations performed within the framework of this study relate to the analysis of air quality in the Warsaw metropolitan area.

2 Air Pollution Simulation: Selected Results

The regional scale Gaussian puff dispersion model CALPUFF [19] was used to simulate the air pollution movement and transformations within the domain. The forecasting model is integrated with the meteorological module CALMET which includes a diagnostic wind field generator.

CALPUFF/CALMET system has been used in a number of studies to investigate gas [20, 21] and particulate matter [22–24] dispersion, also in urban areas. Validation studies showed good correlation with the observations, especially for annual mean concentrations [25, 26]. This fact has been confirmed also within this study by the model performance estimates presented below (see Table 2 and Fig. 2).

In this study, the aggregate emission field was divided into four basic categories, mainly based on technological parameters and the intrinsic uncertainty. The distinguished emission categories, including the quantity of the individual sources in each category, are (i) 16 high-point sources (the energy sector

mainly)—the level of uncertainty is relatively low since both the combustion process and the fuel parameters are precisely defined and stable; because of the high stacks, the modeling procedure must include the initial plume development near the source; (ii) 1002 other point sources (other industrial sources)—a higher level of uncertainty because the technological parameters and fuel parameters are not described so precisely; (iii) 872 area sources (the residential sector and distributed industrial sources)—high level of uncertainty; the emission data is usually only an estimate calculated based on the fuel type and the fuel consumption; and (iv) 1157 linear sources (urban transport)—high level of uncertainty; the emission data is estimated based on several traffic parameters, traffic types and intensity, the fuel used (its quality and consumption), the age of vehicles, and their technological parameters.

The analysis covers a rectangular domain, approximately 30 km×40 km of Warsaw metropolitan area (about 520 km^2 within the administrative borders of Warsaw) as shown in Fig. 1. For computational purposes, the domain is divided using a homogeneous grid with the step size of h=1 km. The locations of all spot sources are defined based on their geographical coordinates, while the area and linear sources are characterized by the respective spatial mesh elements 1 km× 1 km that coincide with the domain grid. The time resolution step size is τ=1 h and is assumed to include temporal variability of the input set of meteorological and emission data. The input dataset for 2005 was selected for the analysis mainly because of its representative nature in terms of the meteorological conditions and the comprehensiveness of the emission data inventory. Moreover, the air quality estimates for 2005 can be used as the reference data for assessing environmental effects of economic and technological changes which occurred in the later years (e.g., economic crisis, modernization of energy/industry, development of the urban transport). The results discussed below relate to the annual mean concentrations of the main urban pollutants, which are recorded at 563 fictitious receptors (the central points of the mesh elements).

The emission field includes sources which are located within the administrative borders of Warsaw and, in some cases of major sources, also outside Warsaw but within the computational domain, as shown in Fig. 1. The transboundary inflow is the background for the pollutants generated by the local sources. The key inflowing pollutants (sulfur and nitrogen oxides, particulate matter, aerosols) adopted as the boundary conditions for CALPUFF simulations are taken from the European scale EMEP model forecasts, spatial resolution 50× 50 km. The results enclosed in [18] show some preliminary accuracy and uncertainty estimates. The analysis refers to the annual mean concentrations of the polluting components which are listed in Table 1 and recorded at the receptor points shown in Fig. 1. The paper also presents some of the results of the simulation.

To assess accuracy of CALPUFF model simulations, the calculated annual mean concentrations of the main pollutants are compared with the data from monitoring stations which are located as shown in Fig. 1. The air quality measurements were conducted by several monitoring stations (automatic or manual) where only selected pollutants (very limited in some cases) were observed. Table 2 presents comparison of the measured and calculated annual mean concentrations of the main gaseous pollutants: NO_x, PM_{10}, and SO_2, as well as heavy metals: Pb, Cd, and Ni at the measurement points.

Moreover, Fig. 2 uses graphics to show the calculated annual mean concentrations versus measurements in the case of the gaseous pollutants (where data feedback was sufficient). The values for particulate matter, PM_{10}, nitrogen oxides, NO_x (adopted from [18]), and sulfur dioxide, SO_2 are demonstrated. The dotted lines in Fig. 2 represent the limits of the standard domain where the model performance quality (FAC2 [27]) was met.

The calculated PM_{10} and SO_2 concentrations match the measured values well enough, while some NO_x forecasts are underestimated. It results from the fact that nitrogen oxide pollution is generated mainly by moving traffic sources (compare Section 3 for details). The calculated values of concentration are averaged per mesh element (1×1 km), while some measurements were taken at the monitoring stations which were located near the road axis where concentrations have maximal values.

3 Uncertainty Analysis

Monte Carlo algorithm was applied to assess the resulting uncertainty of the concentration forecasts in relation to the uncertainty of the input emission data, similarly as in [9, 13]. For all the sources and pollutants, 2000 random sets of emission data were generated within the assumed ranges of uncertainty. To avoid generating unrealistic emission episodes, a correlation between emission intensities of key individual pollutants from each emission source was used (compare [17, 18]).

The key factors in this approach are the ranges of uncertainty in the input emission data. Table 3 presents the ranges, for 95 % confidence interval and four categories of emission sources. The uncertainty ranges were applied as they have been recognized in expert opinions (Warchałowski 2012, Personal communication, Warsaw University of Technology). Recently, levels of sectorial emission uncertainty in Poland have been assessed in [28]. Although this report does not consider the specific urban emissions in Warsaw, the results are close to those presented in Table 3. The normal distributions of the input emission data were assumed.

Selected results were demonstrated in the previous study [18]. Below, a more comprehensive uncertainty analysis is presented for five main pollutants which are specific to air

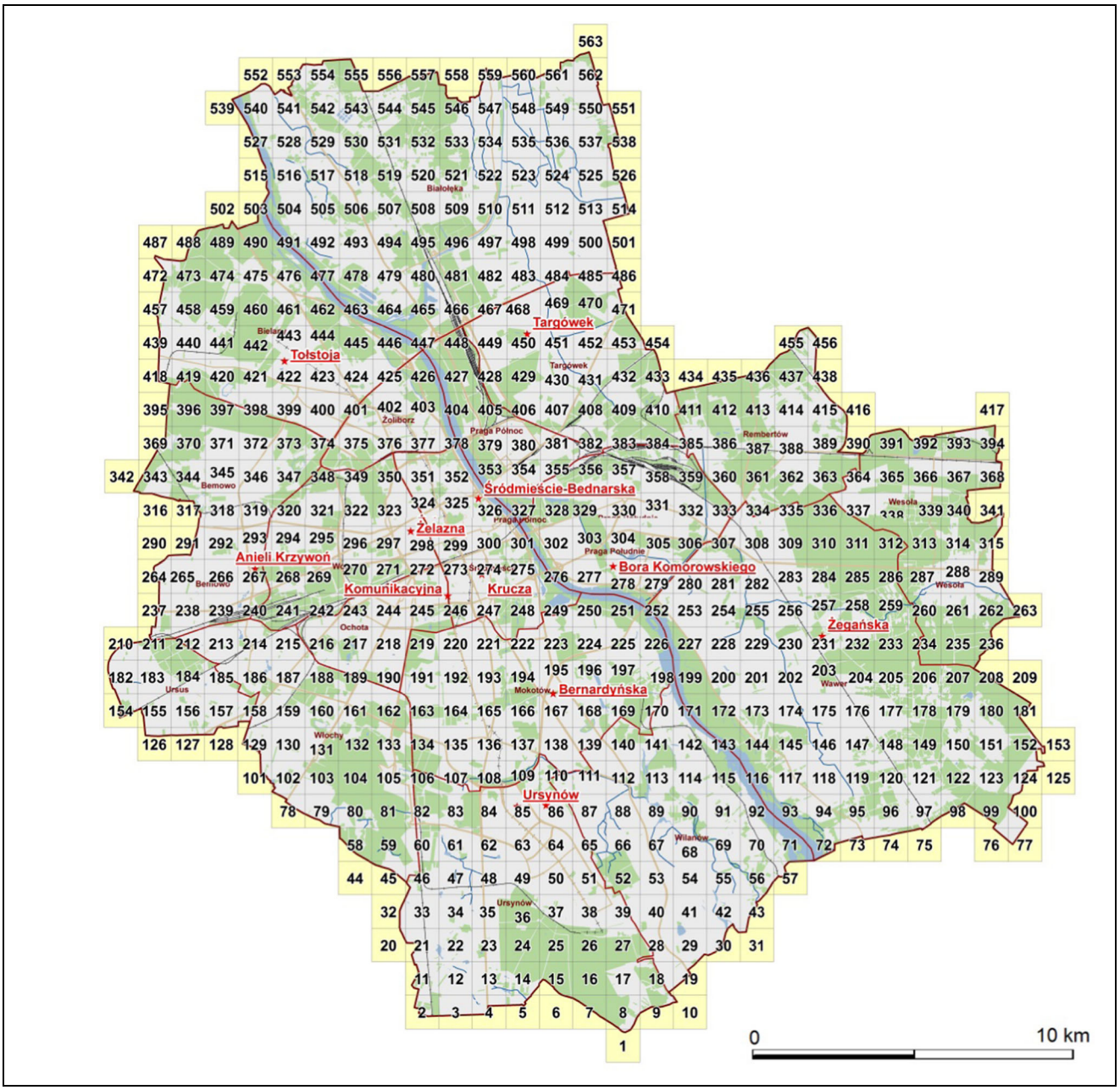


Fig. 1 Computational domain and locations of the receptor points (according to [18])

quality in urban areas, namely NO_x, PM_{10}, $PM_{2.5}$, Pb, and SO_2. The level of uncertainty is related very directly to the aggregate influence of the linear sources (urban transport system) and the local point and area sources. Next, the discussion focuses on the spatial variability of the resulting uncertainty, depending on the resulting concentrations of the above pollutants and certain other factors.

Figure 3 presents the distributions of the standard deviation (left) and the relative uncertainty range (right) versus the annual mean concentrations calculated at 563 receptor points (as shown in Fig. 1) for NO_x, PM_{10}, SO_2, and Pb, respectively. In all cases, the values of standard deviation increase faster (compared to the linear pace) as a function of the mean concentration. The relative uncertainty range at a receptor point is calculated as a ratio $(C_{97.5} - C_{2.5})/C_M$, where $C_{2.5}$ is the 2.5 and $C_{97.5}$ is the 97.5 percentile concentration value, and C_M is the mean value. Strongly dispersed points in uncertainty graphs (Fig. 3 right) suggest the impact of some other factors (besides the mean concentration) which determine the resulting uncertainty.

A more cluster-shaped distribution of SO_2 standard deviation is a result of a relatively homogeneous spread of SO_2 concentration over the Warsaw domain, with the mean values much below the admissible levels. This results from the

Table 1 The polluting components included in the study (according to [18])

Emission/primary pollution	Secondary pollution
SO_2 (sulfur dioxide)	$SO_4^{=}$ (sulfate aerosol)
NO_x (nitrogen oxides)	NO_3^{-} (nitrate aerosol)
	HNO_3 (nitric acid)
PPM_{10} (primary PM, diameter ≤10 μm)	
PPM_{10}_R (PPM_{10} re-suspended by road traffic—secondary emission)	$PM_{10}=PPM_{10}+PPM_{10}\_R+SO_4^{=}+NO_3^{-}$
$PPM_{2.5}$ (primary PM, diameter ≤2.5 μm)	
$PPM_{2.5}$_R ($PPM_{2.5}$ resuspended by road traffic—secondary emission)	$PM_{2.5}=PPM_{2.5}+PPM_{2.5}\_R+SO_4^{=}+NO_3^{-}$
BaP (benzo[*a*]pyrene)	
Ni (nickel)	
Cd (cadmium)	
Pb (lead)	

predominant share of the point sources in the heating system; the high stacks of plants in the district central heating system affect mainly the very distant receptors (often outside the domain), while the small point sources of individual heating systems are mainly active in suburban districts. Moreover, external inflow of sulfur dioxide is a substantial part of the aggregate SO_2 pollution.

Other details related to the issue of uncertainty are explained further in Fig. 4 which contains pairs of maps showing spatial distribution of the mean concentration (left) and the relative uncertainty level (right), for NO_x, PM_{10}, Pb, $PM_{2.5}$, and SO_2, respectively. As can be observed, the typical traffic-related pollutants, like NO_x, PM_{10}, and Pb, show high uncertainty in the vicinity of the main roads, with local peaks near intersections. These local uncertainty peaks are predetermined by the share of emission sources affecting a given receptor.

The general similarity observed in the respective map types (Fig. 4) is related mainly to the movement of the induced

Table 2 Comparison of modeled and measured air pollution concentrations

No.	Monitoring station	PM_{10} [μg/m^3]			NO_x [μg/m^3]			SO_2 [μg/m^3]		
		Modeled	Measured	Error [%]	Modeled	Measured	Error [%]	Modeled	Measured	Error [%]
1	Białobrzeska	33	29.5	11.9	–	–	–	–	–	–
2	Bednarska	42	34.1	23.2	46.7	56.8	−17.8	–	–	–
3	Komunikacyjna	40	51.7	−22.6	51.2	76.2	−32.8	13.9	9.8	41.8
4	Żelazna	34	32.9	3.3	45.6	36.3	25.6	15.6	9	73.3
5	Krucza	43.2	41.7	3.6	31.3	35.7	−12.3	9.8	9.3	5.4
6	Ursynów	32	32.8	−2.4	27.5	19.2	43.2	9.3	8.8	5.7
7	Nowoursynowska	33.5	42.2	−20.6	23.1	25.6	−9.8	9.5	10.5	−9.5
8	Tołstoja	22	37.2	−40.9	35.2	43.1	−18.3	12.1	11.8	2.5
9	Targówek	31.6	31.9	−0.9	–	–	–	–	–	–
10	Anieli Krzywoń	24	31.3	−23.3	–	–	–	–	–	–
11	Bernardyńska	33	21.2	55.7	–	–	–	10.4	8.7	19.5
12	Bora-Komorowskiego	40.5	34.9	16.0	–	–	–	–	–	–
13	Żegańska	27	39.2	−31.1	–	–	–	–	–	–
14	Puszczy Solskiej	–	–	–	26.4	34.1	−22.6	10.6	12.9	−17.8
15	Porajów	–	–	–	19.6	24.1	−18.7	–	–	–
16	Lazurowa	–	–	–	–	–	–	10.1	11.2	−9.8
No.	Monitoring station	Pb [ng/m^3]			Ni [ng/m^3]			Cd [ng/m^3]		
		Modeled	Measured	Error [%]	Modeled	Measured	Error [%]	Modeled	Measured	Error [%]
1	Bernardyńska	22.7	12	89.2	2.3	6.1	−62.3	0.73	0.45	62.2
2	Żelazna	24.5	34	−27.9	1.8	1.5	20.0	0.65	0.7	−7.1
3	Żegańska	20.2	41	−50.7	3.2	2.8	14.3	1.10	0.9	22.2
4	Anieli Krzywoń	18.3	47	−61.1	–	–	–	–	–	–

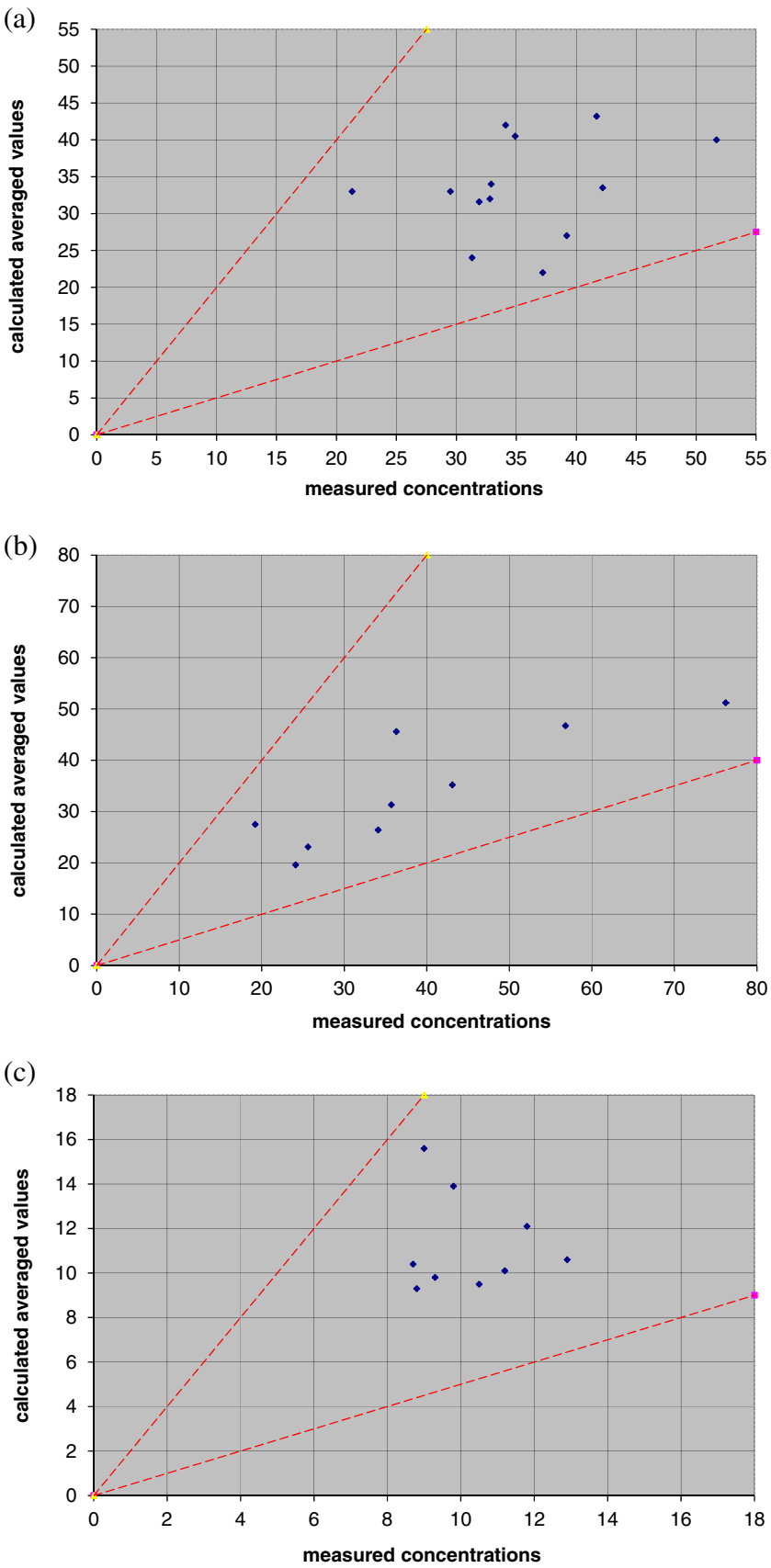


Fig. 2 Calculated versus measured concentrations [μg/m^3]: **a** PM_{10}, **b** NO_x, and **c** SO_2. The *dotted lines* represent factor 2 uncertainty ranges

Table 3 Assumed input uncertainty range depending on emission category (according to [18])

Pollutant	Emission sources			
	High point (%)	Other point (%)	Area (%)	Linear (%)
SO_2	±15	±20	±30	±30
NO_x	±20	±30	±40	±40
PPM_{10}	±25	±40	±40	±40
$PPM_{2.5}$	±25	±40	±40	±40
PPM_{10}_R	–	–	–	±40
$PPM_{2.5}$_R	–	–	–	±40
BaP	±30	±40	±50	±50
Ni	±30	±40	±50	±50
Cd	±30	±40	±50	±50
Pb	±30	±40	±50	±50

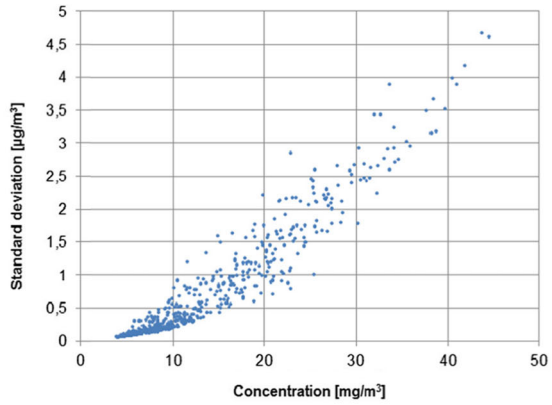
pollution, such as particular matter (PM_{10}), nitrogen oxides, and lead. Their high spatial diversification, with local peaks in the city center zones and near the main traffic arteries, can be observed in Fig. 4 (left). The spatial diversity is much more apparent on uncertainty maps (Fig. 4 right) where the high values reflect well the structure of the road network and the main intersections. This effect can be observed mainly for NO_x, PM_{10}, and Pb. In these cases, high uncertainties are correlated to some extent with the concentration values, but in fact, they depend also on the location of the receptors. Such location determines the relative share of the contributing emission categories and the quantity of the individual emission sources which affect a given receptor point. A specific coincidence of these factors leads to extreme values of the overall uncertainties at some receptor locations.

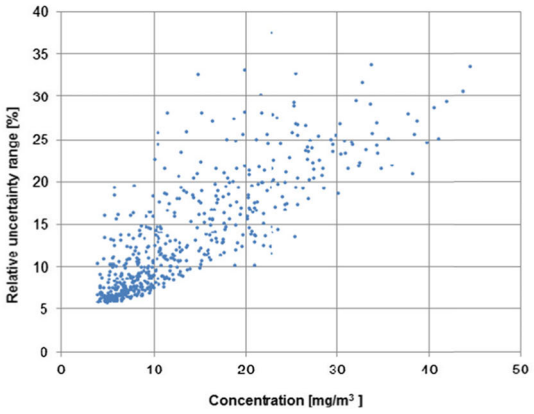
On the other hand, the pollutants which are not under the predominant impact of the transport sources (such as SO_2) are more distributed across the domain in a more uniform manner and their uncertainty levels are relatively low. In this case, the high levels of uncertainty as reported at some receptor points do not coincide with the high concentration values and rather are caused by certain other factors as further explained below.

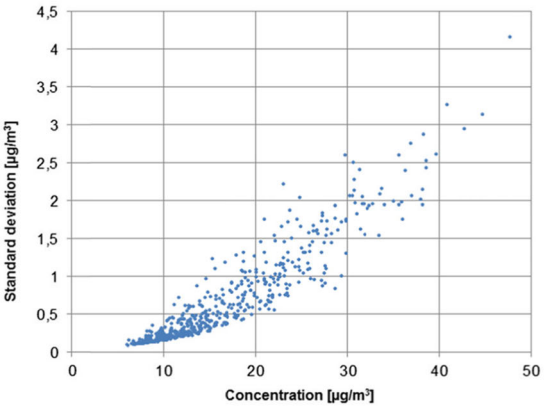
To better illustrate this issue, the concentrations and the corresponding levels of uncertainty are analyzed for two selected receptors, 136 and 156 (compare Fig. 1). The first one is a characteristic for a typical traffic-affected spot (intersection of main roads), while the other one is affected mainly by the local area sources. As can be observed (again refer to Fig. 4, right), the level of uncertainty for the typical urban transport pollutants, like NO_x, PM_{10}, and Pb, and to some extent $PM_{2.5}$, is high near the intersection (receptor 136). According to Fig. 5, the linear sources contribute in a predominant way. However, in addition to the emission categories, another important factor influencing the final level of uncertainty is the quantity of individual sources which substantially contribute to the pollutant concentration at this spot. Generally, due to the averaging effect, the rising number of such emission sources leads to lower aggregate level of the relative uncertainty (the standard deviation for n equally contributing sources is reduced proportionally to $1/\sqrt{n}$). So, the fewer sources contribute to the pollution level, the higher level of the relative uncertainty may be expected. At the same time, an unbalanced contribution of the individual sources (for example, a very predominant and uncertain emission source) generally increases the aggregate level uncertainty for the forecasted pollution.

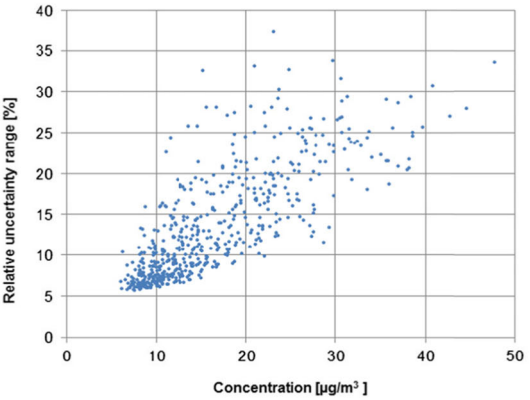
The last property can be quantified, as well. It is assessed below and illustrated qualitatively using sample results taken from Fig. 5 which lists the predominant individual sources accounting for approximately 60 % of the overall pollution (concentration) assigned to receptor 136. As can be observed, for the highly uncertain pollutants, NO_x (~35 %), Pb (~42 %), and $PM_{2.5}$ (~25 %), there are only four contributing sources that have approximately equal level of input uncertainty, with one of them accountable for the predominant share. Similarly, for PM_{10} (~30 %), the number of contributing sources is higher but the predominant share is clear with respect to only one of them. The level uncertainty resulting from the above is at the level of the input uncertainty related to the pollutant emissions. On the other hand, in the case of SO_2 (Fig. 5, bottom), a more balanced contribution between the four emission categories can be observed, with about 15 individual emission sources in total and most of them having a meaningful influence on the final pollution. The aggregate level of relative uncertainty for SO_2 forecast is considerably lower for the receptor considered (~12 %).

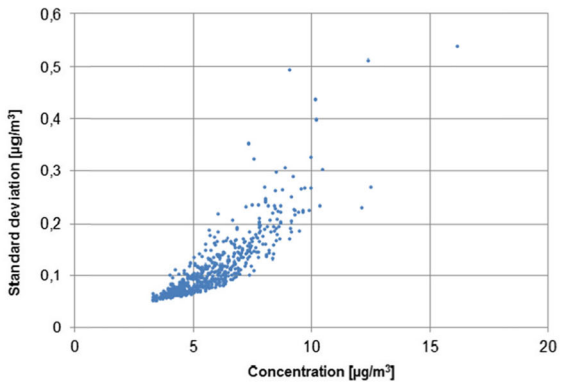
The situation of receptor 156 (compare its location in Fig. 1) is very different. As can be observed in Fig. 4 (right), the levels of uncertainty related to the typical traffic-related pollutants (NO_x, PM_{10}, Pb) are relatively low in the case of receptor 156, while the peak values (across the entire domain) are connected with SO_2 and $PM_{2.5}$. This again can be explained by reference to Fig. 6 which shows the relative share of emission categories as well as the quantity of the individual sources which are the main contributors to the aggregate concentration level as measured at this receptor. Here, the area sources predominate (with the exception of NO_x) because this zone comprises residential areas with a number of small houses equipped with local coal-based heating/cooking systems. On the other hand, a substantial share of the linear emission sources can be observed in the case of all NO_x, PM_{10}, and Pb. This contribution is low for sulfur dioxide and medium for $PM_{2.5}$ because resuspended fraction (secondary $PM_{2.5}$ emission) has a much lower impact than that of PM_{10}. It can be observed in Fig. 6 (right) that only five individual sources contribute to SO_2 and $PM_{2.5}$ pollution, including two cases of the highest values of relative uncertainty reported. The level of uncertainty for Pb is also noticeably high because of the relatively small quantity of the contributing sources.

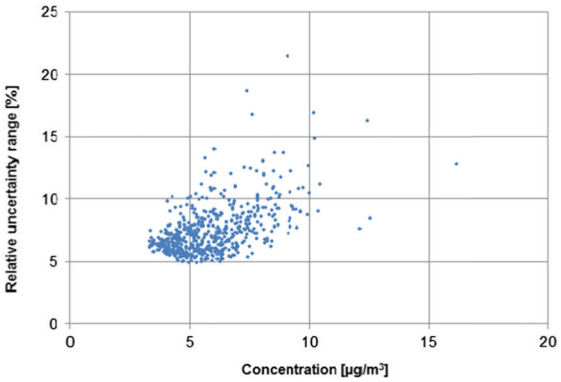

Standard deviation [µg/m³]
Concentration [mg/m³]

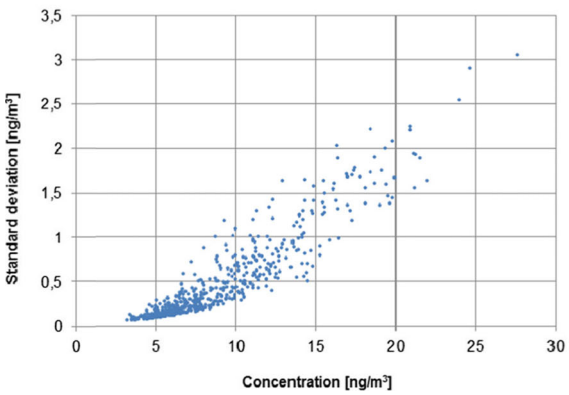

Relative uncertainty range [%]
Concentration [mg/m³]

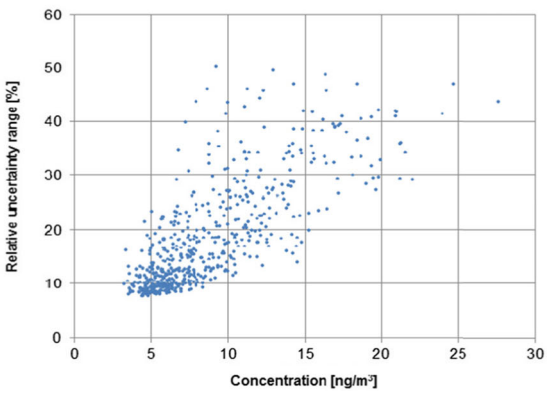

Standard deviation [µg/m³]
Concentration [µg/m³]


Relative uncertainty range [%]
Concentration [µg/m³]


Standard deviation [µg/m³]
Concentration [µg/m³]


Relative uncertainty range [%]
Concentration [µg/m³]


Standard deviation [ng/m³]
Concentration [ng/m³]


Relative uncertainty range [%]
Concentration [ng/m³]

◀ **Fig. 3** Standard deviation (*left*) and relative uncertainty range (*right*) versus concentration level in 563 receptors. From top to bottom: NO_x, PM_{10}, SO_2, and Pb, respectively

Meanwhile, NO_x and PM_{10} concentrations in receptor 156 result from superposition of a large quantity of the contributing sources, which are mostly area sources, but also the neighboring linear sources. This again implies relatively low levels of uncertainty for both pollutants due to the averaging effect as mentioned above.

4 Discussion

This study presents the results of the computational modeling and uncertainty analysis of air pollution dispersion in the Warsaw metropolitan area. The analysis deals with the main types of urban pollutants and relies on the real meteorological data and emission field inventory for 2005. For computational purpose and the detailed uncertainty assessment, the emission field has been split down into four categories: (a) high point sources (power plants), (b) other point sources (industry), (c) area sources (residential sector), and (d) linear sources (transportation).

The main forecasting tool used in the air pollution dispersion simulations is the regional scale transport model CALPUFF [19]. All benefits of the model's linear structure were leveraged in order to calculate the concentration values of the pollutants and implement parallel computation for Monte Carlo method and the uncertainty analysis. One of the crucial steps in quantifying the level of uncertainty is the determination of source contributions to the concentrations in receptor sites (compare Figs. 5 and 6). Here again, thanks to the linearity of CALPUFF structure, the unit emission-concentration transfer matrix can be preprocessed for any individual emission source. Next, the exhaustive computations allow calculating the probability distributions for annual concentrations of the pollutants, both primary and secondary, which are characteristic for the atmospheric environment in urban areas.

The results discussed in the paper cover mainly the sulfur and nitrogen oxide pollutants, lead, and particulate matter, PM_{10} and $PM_{2.5}$. Secondary pollutants of $SO_4^{=}$ and NO_3^{-} were taken into account in the process of forming sulfate and nitrate aerosols, which are then integrated in particulate matters, as stated in Table 1. Also, the annual mean concentrations of Ni, Cd, and BaP were calculated, including their uncertainty, however this study skips their analysis because of the missing representative set of reference data for the period considered.

The annual mean concentrations of sulfur dioxide in the urban domain are relatively low and do not exceed air quality limits (critical level [29] 20 $\mu g/m^3$). The four main categories of emission sources contribute to this type of pollution, with a noticeable share of power/heating plants which are equipped with very high stacks (pollutants are released mainly outside the urban area) and desulfurization installations. On the other hand, concentrations of particulate matter and nitrogen oxides are substantial and have a negative impact on the urban environment. NO_x and PM_{10} concentrations are strongly diversified in spatial terms and can reach very high values locally (especially in the city center zones and near the traffic arteries), often above the limits (the critical levels [29] are 30 and 40 $\mu g/m^3$ for NO_x and PM_{10}, respectively). The main source of such adverse environmental impact is the transport system (including the transit traffic). New ring roads (beltlines), which are now under construction, and the anticipated displacement of the transit and truck transport out of the city center zones, should improve the situation.

Validation of the model (Table 2 and Fig. 2) shows satisfactory accuracy of the forecasted concentrations. The ratio of the modeled versus observed concentrations of the main pollutants meet the standard accuracy index $0.5 \leq FAC2 \leq 2$ [27]. These results confirm the general opinion [20, 25, 26] about the reliable properties of the CALMET/CALPUFF system, especially in relation to the annual mean concentrations.

The sample results shown in Figs. 4, 5 and 6 show that the main factor determining the final level of uncertainty in the model forecast at any receptor point is the share of the predominant emission sources, including the quantity of the main contributing sources. This effect is apparent mainly in the case of a single source that predominates strongly. This general conclusion is illustrated in the two selected receptor points, namely receptor 136: the intersection where the high level of uncertainty occurs for traffic-related pollutants (NO_x, Pb) and receptor 156: the residential area in a peripheral district where SO_2 and fine particulates $PM_{2.5}$ predominate in the emission field and, respectively, the high level of uncertainty is related to the concentrations of both pollutants. On the other hand, in such cases, the impact of the uncertainty related to the input emission as assumed in Table 3 becomes less important.

The main assumptions used in the uncertainty analysis need an elaboration. The first issue is the choice of the typical year (2005) from the meteorological point of view. The year 2005 was analyzed as providing a representative set of meteorological data. The *typical year* method is often used in atmospheric modeling because it greatly reduces the load of computational work, and that issue has been very important in this case where Monte Carlo simulation was utilized. Although the use of a typical year requires some approximations, any changes in the importance of the point sources contributing to specific receptors in other years are not extremely different and stay in a range of ±5–10 %.

Uncertainty estimates are necessary for carrying out risk analysis in decision-making. The Monte Carlo simulation is considered to be the best method of assessing uncertainty distributions of the pollution concentrations and therefore

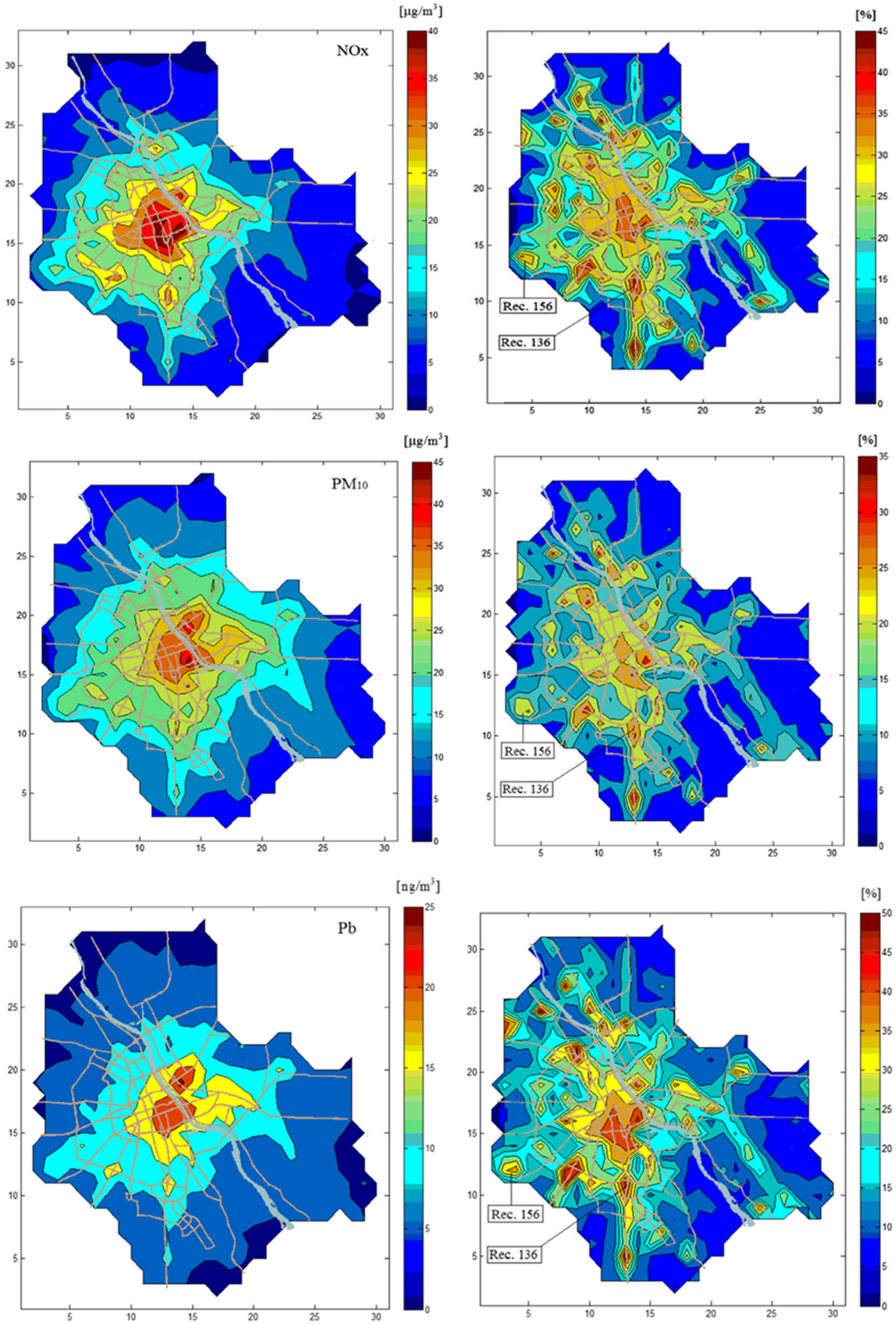


Fig. 4 Concentration (*left*) and relative uncertainty (*right*) for NO_x, PM_{10}, and Pb. Concentration (*left*) and relative uncertainty (*right*) for $PM_{2.5}$ and SO_2

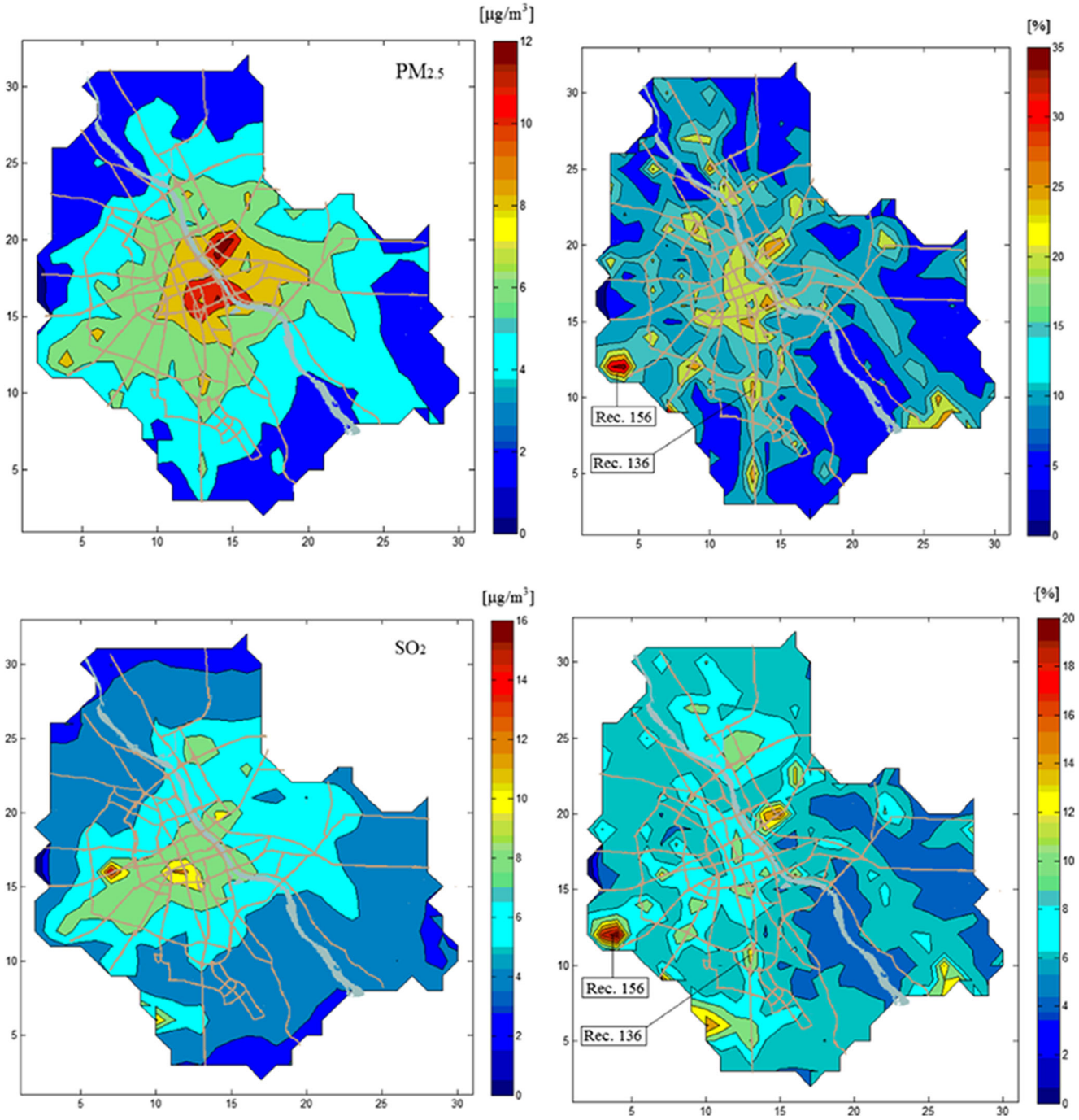


Fig. 4 (continued)

often used for this purpose [30, 31]. Error propagation method is another way to approach this issue. It is computationally more efficient and gives similar results for the Gaussian distributions. However, it is less accurate for non-Gaussian distributions.

The second issue pertains to the assumed ranges of uncertainty as stated in Table 3. Those ranges were adopted from an expert opinion (Warchałowski 2012, Personal communication, Warsaw University of Technology), as no better sources were available at that time. The ranges apply to the conditions in Poland in the recent years. Other ranges assessment can be found in the literature [17]. Quite recently, this issue was addressed in a report prepared by a Polish agency (Warchałowski 2012, Personal communication, Warsaw University of Technology), where many factors were adopted as international default values. In spite of this, similar uncertainty ranges have been obtained.

It should be stressed that the assumptions related to the input emission uncertainty do not influence the main conclusion of these investigations: the range of uncertainty in

Fig. 5 Share of emission categories (*left*) and attributes of sources (*right*) influencing different pollution types for receptor 136, with clearly predominant share of the linear sources

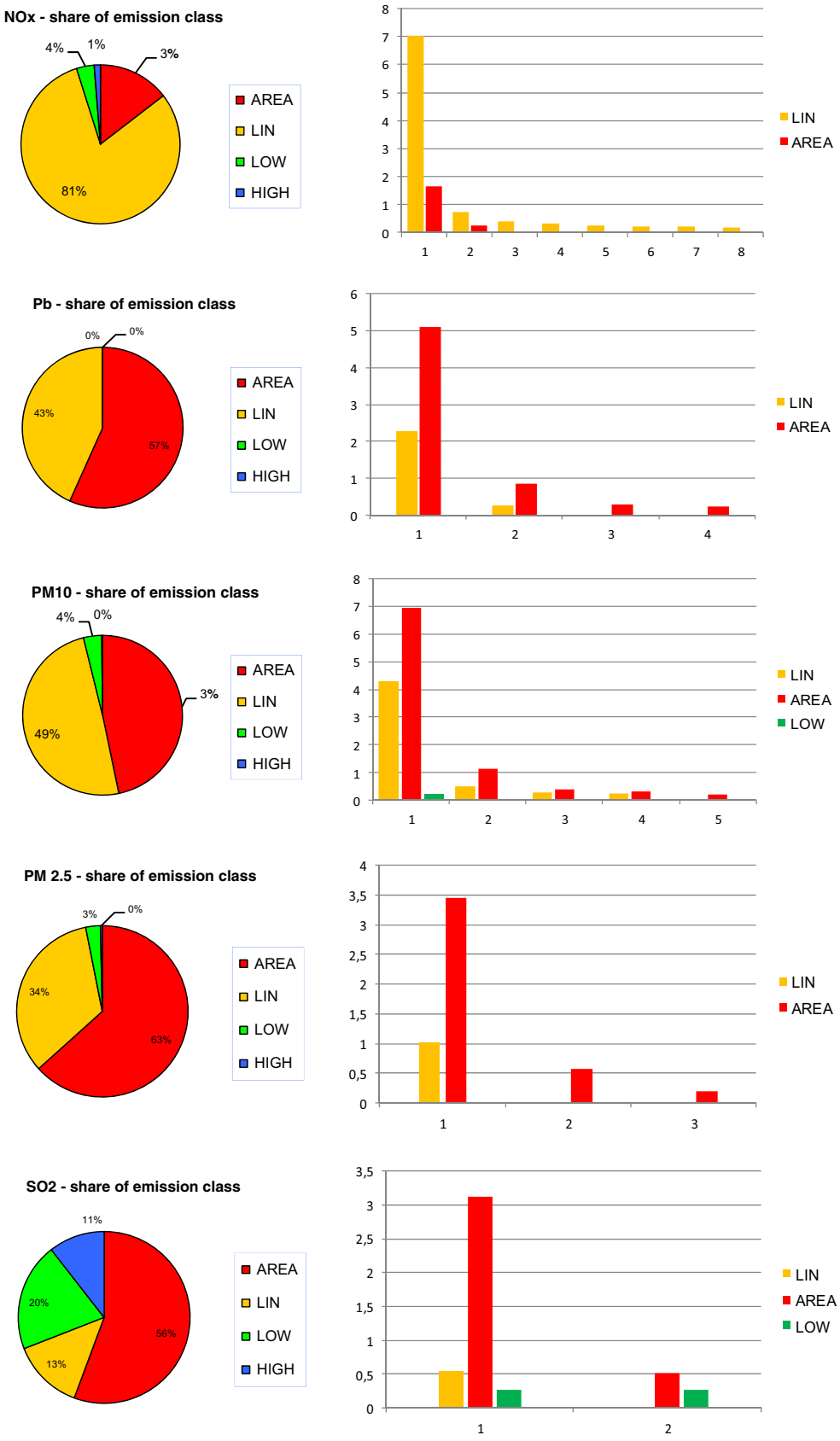


Fig. 6 Share of emission categories (*left*) and attributes of sources (*right*) influencing different pollution types for receptor 156, with clearly predominant contribution of the area sources

pollutant concentrations is lower when many emission sources of similar influence contribute to the concentration. It results from the averaging effect and happens particularly often in the urban environment with many concentrated emission sources. It causes a highly varying rate of uncertainty in air pollutant concentration, as shown in Fig. 4. This effect is important for analyses in which the level of uncertainty in air pollution concentration in cities may have a considerable impact on the final results, for example in the population health risk analysis. If a fixed percentage level of uncertainty was assumed for an entire city, it would cause higher discrepancies in the estimated uncertainty of the final indices being calculated.

5 Conclusions

The main problem of the paper is to analyze the spatial distribution of the uncertainty present in air quality forecasts as caused by the uncertainty present in the input emission data set. It is observed that the accuracy and the uncertainty of the air pollution forecast measured at any receptor point is directly related to the following three factors: (a) the type of pollutant analyzed, (b) the contributing and the predominant categories of emission sources and the assigned levels of their input uncertainty, and (c) the quantity of the individual emission sources having a substantial share in the aggregate pollution. The resulting level of uncertainty assigned to a given receptor point decreases as the quantity of the contributing emission sources increases (the averaging effect).

For the specific air pollutants analyzed in this paper, relatively homogeneous distribution and low uncertainty applies to the concentrations of SO_2, which depend mainly on relatively precise input emission from the point sources and also on the high number of the contributing sources (one exception is receptor 156 where high uncertainty is a result of the strong domination one area source—the local heating of a residential area). On the other hand, there is a very substantial level of uncertainty in NO_x, PM_{10}, and Pb forecasts which strongly depend on the structure of contributing sources, with the predominant impact of the urban transport system. Also, the paper discusses the very high spatial variability depending on the distribution of uncertainty, which is related to traffic-dependent pollutants. Besides the practical value of the above findings, the general objectives of the presented uncertainty results can be attributed to applications of air pollution models as the decision support tools in the air quality management. A complex analysis of the model performance, including uncertainty assessment, improves the credibility of the final policy decisions that allows obtaining confident measurable environmental gains. Moreover, a general aim of such an analysis is, as stated in [7], to bring scientific predictions closer to reality, increase decision-maker's confidence of scientific results, and improve the quality of decisions.

Acknowledgments This paper uses computations which were performed with the financial support from the Polish Ministry of Science and High Education, grant N N519 316735. The presented analysis and preparation of the final form of this paper received support from the statutory fund of the Systems Research Institute PAS. The authors wish to thank Mr. Wojciech Trapp and the EKOMETRIA team for their assistance in the preparation of the emission data.

References

1. ApSimon, H. M., Warren, R. F., & Kayin, S. (2002). Addressing uncertainty in environmental modeling: a case study of integrated assessment of strategies to combat long-range transboundary air pollution. *Atmospheric Environment, 36*, 5417–5426.
2. Buchholz, S., Krein, A., Junk, J., Heinemann, G., & Hoffmann, L. (2013). Simulation of urban-scale air pollution patterns in Luxembourg: Contributing sources and emission scenarios. *Environmental Modeling and Assessment, 18*, 271–283.
3. Carnevale, C., Finzi, G., Pisoni, E., Volta, M., Guariso, G., Gianfreda, R., Maffeis, G., Thunis, P., White, L., & Triacchini, G. (2012). An integrated assessment tool to define effective air quality policies at regional scale. *Environmental Modelling & Software, 38*, 306–315.
4. Mediavilla-Sahagún, A., & ApSimon, H. M. (2006). Urban scale integrated assessment for London: which emission reduction strategies are more effective in attaining prescribed PM10 air quality standards by 2005? *Environmental Modelling & Software, 21*, 501–513.
5. Pisoni, E., Carnevale, C., & Volta, M. (2010). Sensitivity to spatial resolution of modeling systems designing air quality control policies. *Environmental Modelling & Software, 25*, 66–73.
6. Calori, G., Clemente, M., De Maria, R., Finardi, S., Lollobrigida, F., & Tinarelli, G. (2006). Air quality integrated modelling in Turin urban area. *Environmental Modelling & Software, 21*, 468–476.
7. Maxim, L., & van der Sluijs, J. (2011). Quality in environmental science for policy: assessing uncertainty as a component of policy analysis. *Environmental Science & Policy, 14*, 482–492.
8. Colvile, R. N., Woodfield, N. K., Carruthers, D. J., Fisher, B. E. A., Rickard, A., Neville, S., & Hughes, A. (2002). Uncertainty in dispersion modeling and urban air quality mapping. *Environmental Science & Policy, 5*, 207–220.
9. Malherbe, L., Ung, A., Colette, A., Derby, E. (2011). Formulation and quantification of uncertainties in air quality mapping. *ETC/ACM Technical Paper* 2011/9.
10. Russel, A., & Dennis, D. (2000). NASTRO critical review of photochemical models and modeling. *Atmospheric Environment, 34*, 2283–2324.
11. Sportisse, B. (2007). A review of current issues in air pollution modeling and simulation. *Computational Geosciences, 11*, 159–181.
12. Park, S.-K., Cobb, C. E., Wade, K., Mulholland, J., Hu, Y., & Russel, A. G. (2006). Uncertainty in air quality model evaluation for particulate matter due to spatial variations in pollutant concentrations. *Atmospheric Environment, 40*, S563–S573.
13. Sax, T., & Isakov, V. (2003). A case study for assessing uncertainty in local-scale regulatory air quality modeling applications. *Atmospheric Environment, 37*, 3481–3489.

14. Lim, L. L., Hughes, S. J., & Hellawell, E. E. (2005). Integrated decision support system for urban air quality assessment. *Environmental Modelling & Software, 20*, 947–954.
15. Oxley, T., Valiantis, M., Elshkaki, A., & ApSimon, H. M. (2009). Background, road and urban transport modeling of air quality limit values (The BRUTAL model). *Environmental Modelling & Software, 24*, 1036–1050.
16. Holnicki, P., Nahorski, Z., Tainio, M. (2010). Uncertainty in air quality forecasts caused by emission uncertainty. In *HARMO 13th Conference on Harmonisation within Atmospheric Dispersion Modelling* (pp. 119–123).
17. Page, T., Whyatt, J. D., Beven, K. J., & Metcalfe, S. E. (2004). Uncertainty in modelled estimates of acid deposition across Wales: e GLUE approach. *Atmospheric Environment, 38*, 2079–2090.
18. Holnicki, P., & Nahorski, Z. (2013). Air quality modeling in Warsaw Metropolitan Area. *Journal of Theoretical and Applied Computer Science, 7*, 56–69.
19. Scire, J. S., Strimaitis, D. G., Yamartino, R. J. (2000). *A User's Guide for the CALPUFF Dispersion Model*. Earth Technology Inc.
20. Elbir, T. (2003). Comparison of model predictions with the data of an urban air quality monitoring network in Izmir, Turkey. *Atmospheric Environment, 37*, 2149–2157.
21. US EPA (1998). *A comparison of Calpuff modeling results to two tracer field experiments,* http://www.epa.gov/scram001/7thconf/calpuff/tracer.pdf.
22. Huber, A., Georgopulos, P., Giliam, R., Stenchikov, G., Wang, S., Kelly, B., Feingersh H. (2004). Modeling air pollution from the collapse of the World Trade Center and assessing the potential impacts on human exposures. *EM*, 35–40.
23. Trapp, W. (2010). The application of CALMET/CALPUFF models in air quality assessment system in Poland. *Archives of Environmental Protection, 36*, 63–79.
24. Villasenor, R., Lopez-Villegas, M. T., et al. (2003). A mesoscale modeling study of wind-blown dust on the Mexico City Basin. *Atmospheric Environment, 37*, 2451–2462.
25. Holmes, N. S., & Morawska, L. (2006). A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment, 40*, 5902–5928.
26. Tartakovsky, D., Broday, D. M., & Stern, E. (2013). Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. *Environmental Pollution, 179*, 138–145.
27. Juda-Rezler, K. (2010). New challenges in air quality and climate modeling. *Archives of Environmental Protection, 36*, 3–28.
28. UN ECE (2014). *Convention on long-range transboundary air pollution. Poland's Informative Inventory Report 2014. Appendix 5* "*Uncertainty analysis of emissions of selected air pollutants*".
29. EU (2008). *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.*
30. Hanna, S. R., Chang, J. C., & Fernau, M. E. (1998). Monte Carlo estimates of uncertainties in predictions by photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmospheric Environment, 32*, 3619–3628.
31. Moore, G. E., & Londergan, R. J. (2001). Sampled Monte Carlo uncertainty analysis for photochemical grid models. *Atmospheric Environment, 35*, 4863–4876.