ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Atmospheric Environment ▮ (▮▮▮▮) ▮▮▮–▮▮▮

ATMOSPHERIC ENVIRONMENT

www.elsevier.com/locate/atmosenv

Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in Finland

Jaakko Kukkonen[a,*], Ranjeet Sokhi[b], Lakhu Luhana[b], Jari Härkönen[a], Timo Salmi[a], Mikhail Sofiev[a], Ari Karppinen[a]

[a] *Finnish Meteorological Institute, Erik Palmenin aukio 1, P.O. Box 503, FI-00101 Helsinki, Finland*
[b] *University of Hertfordshire, Science and Technology Research Institute, College Lane, Hatfield AL10 9AB, UK*

Received 10 July 2006; received in revised form 15 February 2007; accepted 20 February 2007

Abstract

We have developed a simple statistical model for the estimation of the long-range transport (LRT) contribution to $PM_{2.5}$ concentrations. The modelling is based on linear regressions of the 'Co-operative programme for monitoring and evaluating the long-range transmission of air pollutants in Europe' (EMEP) background ionic components (SO_4^{2-}, NO_3^-, NH_4^+) with the monitored $PM_{2.5}$ concentrations. We present an evaluation and application of the model against data measured in the United Kingdom (UK) and in Finland. We have studied the correlation of ion sum values with the $PM_{2.5}$ data measured at two EMEP stations in Finland (i.e., a comparison of ionic sum and $PM_{2.5}$ at one station). The statistical correlations of the $PM_{2.5}$ concentrations with the ion sum values were very high (R^2 varied from 0.77 to 0.83) at both of the stations considered; this provides confidence that the ion sum is a good proxy variable for the LRT $PM_{2.5}$. The comparison of different modelling options using the data measured in the UK showed that the regression model gave systematically substantially better results than the model using merely sulphate concentrations. Similarly, using the distance weighted ion sum based on data from two EMEP stations gave better correlation, compared with the option of using only one EMEP station. The evaluated average LRT contribution accounted for 35–37% of the regional air $PM_{2.5}$ concentrations in UK from 1998 to 2000. The corresponding contributions at two urban stations in London were 24–31%. We conclude that the model is a useful and simple tool for the assessment of LRT $PM_{2.5}$ that is applicable within a fairly good accuracy.

Keywords: Long-range transport; $PM_{2.5}$; Model; EMEP; Ionic component

1. Introduction

Aerosol particle concentrations in urban areas originate from several source categories, such as local vehicular traffic, industrial emissions and long-range transport (LRT). Airborne particulate matter has both a primary component, which is emitted directly from pollution sources, and a secondary component, which is formed in the atmosphere by chemical reactions of gases, most notably sulphur dioxide, oxides of nitrogen and volatile organic compounds (e.g., APEG (Airborne Particles Expert

---

[*] Corresponding author. Tel.: +358 9 1929 5412; fax: +358 9 1929 5403.
E-mail address: jaakko.kukkonen@fmi.fi (J. Kukkonen).

1352-2310/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2007.02.036

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

Group), 1999). Secondary sulphate and nitrate particles have long residence times in the atmosphere due to their physical size (e.g., Deacon et al., 1997). Particles are formed via gas-to-particle conversion, and are constantly transformed and depleted during atmospheric transport by various physical and chemical processes, including coagulation, condensation and evaporation, chemical transformation, and dry and wet deposition (e.g., Pohjola et al., 2003; Kukkonen et al., 2003; Ketzel and Berkowicz, 2004).

Source apportionment studies of the United Kingdom (UK) atmosphere have shown that there are three predominant contributors to PM_{10} mass (APEG (Airborne Particles Expert Group), 1999). These are (i) road traffic, (ii) secondary particles and (iii) coarse particles arising from a number of sources including suspension of surface soils and dusts, sea spray and construction activity. Secondary particles are much more uniform spatially across the UK than primary particles, although there is a general reduction in concentrations moving from a maximum in the south-east to a minimum in the north-west of the British Isles.

Turnbull and Harrison (2000) studied the physical and chemical characteristics of PM_{10}, including chloride, nitrate, sulphate, and black smoke, at four coastal, rural, and urban sites across the UK. They performed statistical analysis on the measured data to quantify the contribution of major components of PM_{10} in the UK atmosphere. They found out that the secondary particles contributed 28–35% of site-mean PM_{10}, primary combustion particles 20–57%, sodium chloride 11–34% and other, mainly crustal particles, accounted for 3–21%. The results also showed that the absolute contribution of secondary component and NaCl are quite comparable for all but one site. In case of a coastal site, the contribution of NaCl was significantly higher, 34%, than in the other sites.

Malcolm et al. (2000) used dispersion modelling to investigate LRT particulate pollution in the UK, with special emphasis on two PM_{10} episodes. They concluded that as expected, the concentrations were highest when the air masses were transported from the European source regions. By considering a number of measurement sites in the UK over the whole of 1996 they showed that about 75% of the yearly average contribution of PM_{10} was from the UK with the other 25% being due to other European emissions. However, for the sites which were located in the south and east of the UK, these proportions were approximately 60% from UK and 40% other European.

Abdalmogith and Harrison (2005) used trajectory cluster analysis to examine the LRT of secondary inorganic aerosol in the UK. The highest nitrate and sulphate levels were associated with south-easterly and easterly trajectory clusters at the stations of Belfast and Harwell, respectively, and highest chloride concentrations were associated with fast moving maritime trajectory categories at both sites. Nitrate showed greater episodicity than sulphate. Ratios of sulphate/nitrate varied by a factor of more than two according to the trajectory due to the different precursor source areas.

In Nordic countries, atmospheric LRT constitutes a substantial part of the total urban background $PM_{2.5}$ concentrations (e.g., Johansson et al., 1999; Tiitta et al., 2002; Karppinen et al., 2004). As in the case of the UK, the measured fine particulate matter concentrations ($PM_{2.5}$) only show moderate spatial and temporal variation (e.g., Pohjola et al., 2002; Karppinen et al., 2005). A more substantial variation is found for the urban concentrations of PM_{10} (Kukkonen et al., 2001) as a result of local variations in the coarse fraction emissions.

Tsyro (2002) has evaluated the relative contributions of different aerosol components to the total PM_{10} mass in 1999 in Europe using the EMEP ('Co-operative programme for monitoring and evaluating the long-range transmission of air pollutants in Europe') Unified Aerosol model. According to the computations, SO_4 was clearly the dominating aerosol component, accounting for 30–50% of PM_{10} concentrations in western, central and northern Europe, and for 50–70% in Eastern Europe and parts of Russia. NO_3 was the next important component to the ambient aerosol; its contribution was 20–30% in central Europe and decreased to 5–15% in the Eastern Europe and Russia, and to below 5% in the Northern Europe. NH_4 contributed 5–20%, and carbonaceous particles from 3% to 15% for primary organic carbon and from 1% to 7% for elemental carbon.

However, there are still major uncertainties in conducting European-wide and urban scale aerosol computations. In particular, in many major European cities no reliable PM emission inventories are available; especially the formation of vehicle non-exhaust emissions and suspension that originate from brakes, tires, etc. is poorly understood. It is therefore worthwhile to develop and evaluate also simpler, semi-empirical and statistical methods for

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

analysing the contributions of different emission sources to urban air PM concentrations.

Continuous measurements of LRT contributions to $PM_{2.5}$, however, are only available for a minor fraction of the European cities. The spatial representativity of such measurements may also not be known in detail, and it can be limited in many cases, e.g., due to the influence of regional or local scale emissions. However, the data from a measurement network of the EMEP are available for Europe. Using this data we have, therefore, developed a simple statistical model for estimating the LRT contribution to urban air $PM_{2.5}$ concentrations. The first version of the model was presented by Karppinen et al. (2004) and applied by Tiitta et al. (2002). The results by Karppinen et al. (2004) showed that there was a strong association between the ion sums interpolated from the EMEP data and the $PM_{2.5}$ concentrations measured at urban sites in Helsinki in 1998–2000.

The previous work used a limited dataset to test the modelling approach. This current study aims to employ a substantially more extensive and versatile (in terms of the diagnostic evaluation of the model) dataset to evaluate its accuracy and performance for the UK and Finland. Evaluation of the model in two European countries also provides insight on whether it would be a useful tool of assessment in other European regions.

2. Methodology

2.1. Experimental data

The locations of the stations that were utilized in this study have been presented in Figs. 1(a) and (b). These include EMEP stations, supplementary rural stations and urban stations.

2.1.1. The selected rural monitoring stations in UK and in Finland

We selected the EMEP stations and the time periods based on two criteria. First, the stations (i) should be measuring all of the three ionic components that are required in the model computations (sulphate, nitrate and ammonium), and (ii) whenever possible, they should also simultaneously be measuring $PM_{2.5}$ mass concentrations along with the ions. The stations of Virolahti and Utö in Finland have satisfied both conditions since July and October of 2003, respectively, while the stations of Eskdalemuir and High Muffles in UK fullfilled the first condition during the selected years of

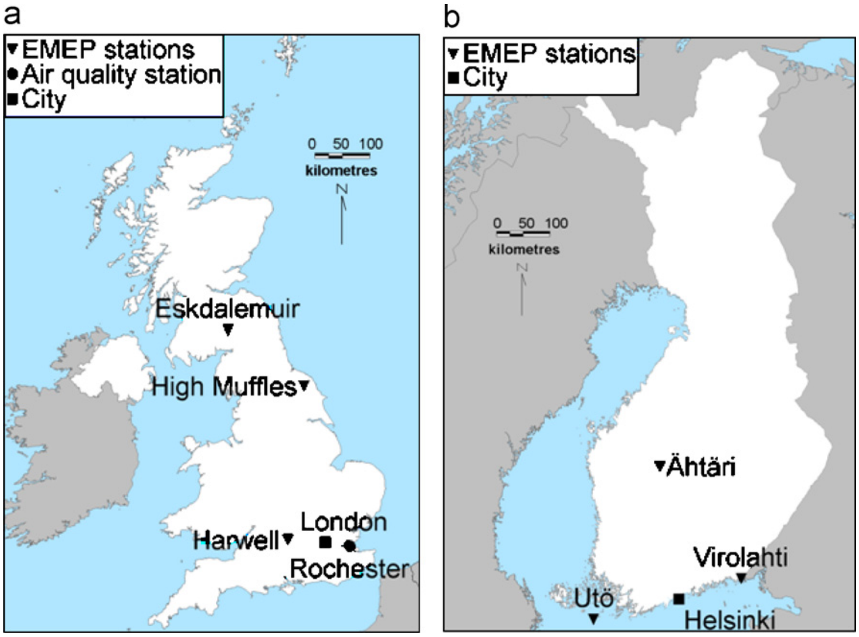

Fig. 1. (a) Locations of the regional background monitoring stations that were utilized in this study, and (b) the cities addressed in the United Kingdom and Finland.

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

assessment, from 1998 to 2000. We selected the period from 1998 to 2000, as the required measurements of the ion sums ended at High Muffles and Eskdalemuir in March 1999 and November 2000, respectively.

The station of Eskdalemuir is located in an open rural area with trees at a distance of 150 m as the nearest obstruction. The station of High Muffles has been established in a cleared forest plantation in a remote, open area. The station of Utö is located on a small island on the Baltic Sea, Ähtäri is situated in a forested regional background area in Central Finland, and Virolahti is located near the eastern border of Finland, 150 km W–NW from St. Petersburg. Although no new data from the station of Ähtäri is presented in this study, it is presented here for completeness (we utilise the previous results presented by Karppinen et al., 2004 based on data measured at that station).

The following ion concentrations are measured daily at these EMEP stations: (i) sulphate (SO_4^2), (ii) the sum of nitrate (NO_3^-) and nitrogen acid (HNO_3), and (iii) the sum of ammonium (NH_4^+) and ammonia (NH_3). The sulphate, nitrate and ammonium ions are in particulate form, while nitrogen acid and ammonia are gaseous compounds in atmospheric conditions. The sampling and reporting period of the daily EMEP values starts at 06:00 Coordinated Universal Time (UTC) (in accordance with meteorological conventions). A detailed description of the sampling and analysis methods, and the QA/QC procedures in the EMEP programme has been published elsewhere (e.g., EMEP, 2001; Hayman, 2005).

We also utilised supplementary $PM_{2.5}$ data measured at the EMEP station of Harwell and at the rural background station of Rochester, as both of these are located in the vicinity of London in Southern England. The $PM_{2.5}$ measurements were started at Harwell and Rochester in April 1998. However, the required ionic data is not available at the station of Harwell. The measurement sites of Harwell and Rochester are part of the Automatic Urban and Rural Network (AURN). A rigorous QA/QC and data ratification protocol is used by National Environmental Technology Centre (NETCEN) at the sites of the network (Eaton and Stacey, 2006).

The NETCEN instruments are TEOM analysers; the instruments at the Virolahti and Utö stations are of the type Eberline FH 62 IR. All TEOM analysers measuring $PM_{2.5}$ are set up with the USEPA default adjustment factor (TEOM reading × $1.03 + 3\,\mu g\,m^{-3}$; AQEG (Air Quality Experts Group), 2005).

2.1.2. Measured urban $PM_{2.5}$ concentration data in London

We utilised the hourly time-series datasets of $PM_{2.5}$ concentrations measured at the urban centre station of Bloomsbury and the urban kerbside station of Marylebone Road for 1998–2000. Both of these measurement stations are located in central London and are part of the monitoring network of the Department of Environment, Food and Rural Affairs (DEFRA; 〈http://www.airquality.co.uk/archive/index.php〉). The measurement site of Bloomsbury is also part of the above-mentioned AURN Network. The station of Marylebone Road side is part of London Air Quality Network (LAQN). The instruments are TEOM analysers; set up with the USEPA default adjustment factor (as described above).

The station of Bloomsbury is located in the north-east corner of central London gardens, and is surrounded by mature trees. There are busy roads at all sides of the station; the distance of the nearest road is approximately 25 m. The monitoring station cabin at Marylebone Road is located approximately at a distance of 1 m away from the kerbside. The daily traffic flows on Marylebone Road are in excess of 80,000 vehicles and the road is frequently congested. There are major buildings on both sides of the road.

2.2. Model

The model equations were presented briefly by Karppinen et al. (2004); however, we present a more refined version of the model here.

In the vicinity of the source, the fraction of sulphates of the total aerosol mass concentration is commonly lower than 20% and that of nitrates is negligible (e.g., Simpson et al., 2003); their concentrations are therefore not necessarily initially correlated with the total aerosol mass concentration. On the other hand, the anthropogenic emissions of sulphur and nitrogen oxides are highly correlated with those of black carbon and, also somewhat correlated to the emissions of organic carbon (EEA, 2006). This is to be expected, as all of these pollutants partly originate from the combustion of fossil fuels.

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

During atmospheric transport, SO_2 and NO_2 are gradually oxidised. The characteristic timescales of these processes vary from hours to days; this depends on the concentrations of strong oxidants, such as OH radicals and ozone, and on environmental conditions (e.g., Jacobson, 1999). Simultaneously, coarse and ultrafine primary particles are relatively efficiently scavenged, and secondary organic particulate matter is formed. These processes commonly lead to the formation of an aerosol mixture within the long-range distance transported size range 0.1–1 μm that contains a substantial fraction of sulphates, nitrates and secondary organic aerosol. For sufficiently aged plumes, the sulphur and nitrogen aerosol are therefore expected to be highly correlated with the total aerosol mass.

2.2.1. *Definition of the ion sum*

The EMEP measurements are reported as equivalent masses of sulphur and nitrogen. We have defined the so-called ion sum as follows:

$$C_{\mathrm{ion}} = 3.0[\mathrm{SO}_4^{2-}]_{\mathrm{S}} + 4.4[(\mathrm{NO}_3^- + \mathrm{HNO}_3)]_{\mathrm{N}} + 1.3[(\mathrm{NH}_4^+ + \mathrm{NH}_3)]_{\mathrm{N}}, \qquad (1)$$

where the subscripts S and N denote that the mass has been given as the equivalent mass of sulphur or nitrogen. These values were converted to equivalent masses of the ions SO_4^{2-}, NO_3^- and NH_4^+, using the conversion factors (CF) 3.0, 4.4 and 1.3, respectively. This conversion was necessary in order to treat the particulate concentration variables in a comparable manner. The CFs are computed simply based on the ratio of the relative approximate atomic (for H, N, O and S, these are 1, 7, 8 and 16, respectively) and molecular masses, as follows: $\mathrm{CF}\ (\mathrm{SO}_4^2)_{\mathrm{S}} = (16+4\times8)/16 = 3.0$, $\mathrm{CF}\ (\mathrm{NO}_3^-)_{\mathrm{N}} = (7+3\times8)/7 = 4.4$, and $\mathrm{CF}\ (\mathrm{NH}_4^+)_{\mathrm{N}} = (7+4)/7 = 1.3$.

The C_{ion} variable was assumed to be a suitable proxy variable for LRT in the model. However, it should be noted that this variable contained in part measurements of gaseous substances (HNO_3, NH_3), and on the other hand, that LRT contains also other compounds, such as primary particular matter, and elemental and organic carbon.

In fact each of the three ion terms in Eq. (1) could be treated as a suitable proxy variable for LRT. However, it was convenient to define a combined variable that was based on all three measurements of ion concentrations. The main reason for this was that combined variable represented a larger fraction of the $PM_{2.5}$ mass than any of its components, which made it a better candidate for predicting the temporal variations.

2.2.2. *Utilisation of data from several EMEP stations*

One of the main criteria in the selection of sites for the EMEP stations has been that these would represent regional background concentrations, and not be directly influenced by local emission sources. However, this may not be strictly correct in all cases, due to reasons such as changes in land use and location of major roads nearby. In the present study, we utilised a combination of data from two or three nearest EMEP stations to the urban area considered, in order to smooth out any possible disturbances caused by local emission sources. This procedure also removed the difficulties that might have arisen from missing data at any individual station.

Whenever possible, we utilised an interpolated value of the ion sum, defined as

$$C_{\mathrm{ion}} = \sum_{i=1}^{n} \chi_i C_{\mathrm{ion},i}, \qquad (2)$$

where the subscript i referred to the included EMEP stations, n the total number of stations, χ_i the weight coefficient and $C_{\mathrm{ion},i}$ the ion sum at each included station. The weight coefficient was defined, on the basis of the EMEP stations with available data, as a normalised inverse value of the distance between the urban measurement location and the EMEP station. The normalisation implied simply that the sum of all χ_i values was required to be one. The purpose of using the inverse values of the distances was to give larger weight coefficients for the EMEP stations that were closer to the city considered. For instance, if we assume that the data from the three EMEP stations around London were to be utilised, let the distances between each EMEP station and the London urban station be d_1, d_2 and d_3. Denoting the sum of the un-normalised weight coefficients as $S = d_1^{-1} + d_2^{-1} + d_3^{-1}$, the normalised weight coefficients are $1/(Sd_1)$, $1/(\mathrm{S}d_2)$ and $1/(Sd_3)$, respectively.

2.2.3. *Statistical correlation of ion sum and $PM_{2.5}$ concentrations*

The measured urban air $PM_{2.5}$ concentrations (i.e., 24 h average values corresponding to the EMEP network sampling times starting at 06:00 UTC) were associated with the ion sum values

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

as follows:

$$\mathrm{PM}_{2.5}(\mathrm{measured}) = k_1 C_{\mathrm{ion}} + k_{\mathrm{o}}, \tag{3}$$

where k_1 and k_{o} were regression coefficients determined statistically. The statistical estimation of k_1 indicated the strength of temporal association of the background C_{ion} concentration with the urban air $\mathrm{PM}_{2.5}$ concentration, while k_{o} indicated the uncorrelated portion of the $\mathrm{PM}_{2.5}$ concentration. Consequently, the terms $k_1 C_{\mathrm{ion}}$ and k_{o} were interpreted to represent contributions from LRT and all local sources, respectively. In a regional background station, it is expected that $k_{\mathrm{o}} \ll k_1 C_{\mathrm{ion}}$.

In principle, a part of the statistical correlation between the measured $\mathrm{PM}_{2.5}$ and the ion sum values could be caused by micro- and meso-scale meteorological conditions that would influence substantially both concentrations. Consequently, the simple regression model, as written in Eq. (3), may overestimate the LRT contribution to the urban air $\mathrm{PM}_{2.5}$ concentrations.

Clearly, a more detailed method for evaluation of the LRT contribution could include additional factors in Eq. (3) that would be dependent on relevant meteorological parameters. However, these parameters should correspond to weather conditions along the trajectories of air masses, instead of merely local or regional meteorological measurements. This would drastically increase the complexity of the model.

For the analyses required in determining the two regression coefficients in Eq. (3), we utilised the statistical software tool SPSS for Windows (Release 11.5.0, SPSS Inc, USA, ⟨http://www.spss.com/⟩).

3. Results

In the previous study by Karppinen et al. (2004), we applied $\mathrm{PM}_{2.5}$ measurement data from two urban locations in Helsinki (Vallila, Kallio) and statistically correlated these data with daily ion concentration data obtained from three EMEP monitoring stations located in Southern Finland (Utö, Virolahti and Ähtäri), in 1998–2000. The present study extends this evaluation in two respects: (i) by considering also the correlation of ion sum values with the $\mathrm{PM}_{2.5}$ and PM_{10} data measured at two EMEP stations in Finland and one regional background station in UK, and (ii) by analysing the corresponding data measured in the UK. The former analysis is useful for methodological reasons, in order to evaluate model accuracy and reliability, and the latter analysis provides a better insight on the applicability of the method.

3.1. Evaluation of the model against the $\mathrm{PM}_{2.5}$ data measured at two Finnish EMEP stations

Hourly measurements of $\mathrm{PM}_{2.5}$ began at the two EMEP stations, Virolahti and Utö, on 8 July and 9 September 2003, respectively. Measurements of PM_{10} have also been conducted at the EMEP station of Virolahti since 8 July 2002. We have computed the ion sum values based on Eq. (1) for both of these stations, using solely the daily averaged ion concentration data measured at the same EMEP station as the corresponding $\mathrm{PM}_{2.5}$ or PM_{10} data, i.e., assumed $n = 1$ in Eq. (2). The linear regression correlations between the daily concentrations of $\mathrm{PM}_{2.5}$ and PM_{10} against the corresponding ion sum values at these stations have been presented in Figs. 2(a) and (b), as in Eq. (3). The time period considered was approximately 4 months (the number of data $N = 84$) and 6 months ($N = 177$) in 2003 for the stations of Utö and Virolahti, respectively. In 2004, $N = 366$ for both stations.

The statistical correlations of the $\mathrm{PM}_{2.5}$ concentrations with the ion sum values are high (R^2 varies from 0.77 to 0.83). Strong correlation between PM_{10} concentrations and the ion sum values at the station of Virolahti was also observed in 2003 ($R^2 = 0.61$; data not shown here). These correlations can be compared with those previously obtained for two urban locations in Helsinki and three EMEP monitoring stations located in Southern Finland, in 1998–2000 (Karppinen et al., 2004). The latter R^2 values for the daily average $\mathrm{PM}_{2.5}$ concentrations with the ion sum ranged from 0.59 to 0.61, computed separately for each year. As expected, the statistical correlations are substantially higher in the case of considering the $\mathrm{PM}_{2.5}$ data measured at the same location (i.e., the EMEP station).

The intercept k_{o} is a measure of the purely local contribution to the total particular matter concentration. As expected, the intercepts are small compared with the average concentrations at both stations.

3.2. Analysis and evaluation of the model against the data measured in UK

For model evaluation purposes, we have first computed the ion sum values, based on Eqs. (1) and (2), for the stations of Harwell and Rochester, using

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $\mathrm{PM}_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

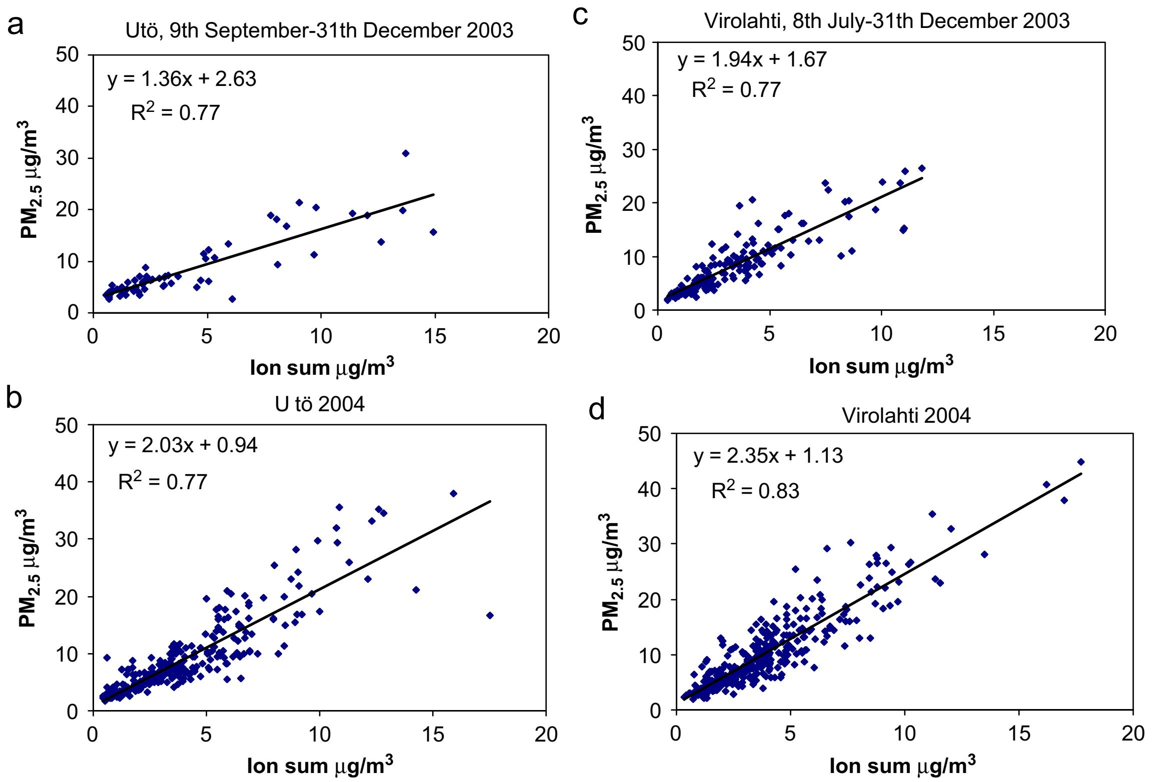

Fig. 2. The daily means of the $PM_{2.5}$ ($\mu g\,m^{-3}$) concentrations against the ion sum values ($\mu g\,m^{-3}$) at the stations of Utö (a) and (b) and Virolahti (c) and (d) in Finland in 2003 and 2004, respectively. The time period considered was approximately 4 and 6 months in 2003 for the stations of Utö and Virolahti, respectively. The linear statistical fits have also been shown. R^2 = correlation coefficient squared.

the ion concentration data measured at the EMEP stations of Eskdalemuir and High Muffles. The correlations between the daily means of the concentrations of $PM_{2.5}$ against the corresponding ion sum values at these stations have been presented in Figs. 3(a) and (b). The statistical correlations of the $PM_{2.5}$ concentrations with the ion sum values are fairly high ($R^2 = 0.44$ and 0.41), although as expected, these are lower than those observed for the Finnish stations mentioned above, where the $PM_{2.5}$ and ion sums from the same station were employed. The slopes and intercepts are similar at the stations of Harwell and Rochester.

The correlations between the daily means of the concentrations of $PM_{2.5}$ against the corresponding ion sum values at two urban stations, Bloomsbury and Marylebone Road in London have been presented in Figs. 4(a) and (b). The statistical correlations are significant ($R^2 = 0.40$ and 0.32), although as expected, these are lower than the corresponding correlations at the regional background stations. The slopes are similar at the two stations; however, the intercept is larger for the station of Marylebone Road. This is caused by the larger local particular matter contribution at the densely trafficked kerbside station of Marylebone Road, compared with that at the urban centre station of Bloomsbury.

In Figs. 4(a) and (b), there are a number of days, when high measured $PM_{2.5}$ concentrations correlated poorly with the relatively lower ion sum values. These events could be caused by (i) higher than average contributions from local primary emissions, e.g., due to exceptionally severe traffic congestion, or airborne dust suspended from road and street surfaces, and also by (ii) specific meteorological conditions on a meso-scale (such as local stable, low wind speed or calm conditions).

For analysis and evaluation of the model, we have also computed the corresponding correlations, using only the data of one EMEP station for computing the ion sum values, and using only the sulphate (SO_4^{2-}) concentration as the proxy variable for the LRT concentration of $PM_{2.5}$. The results of

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

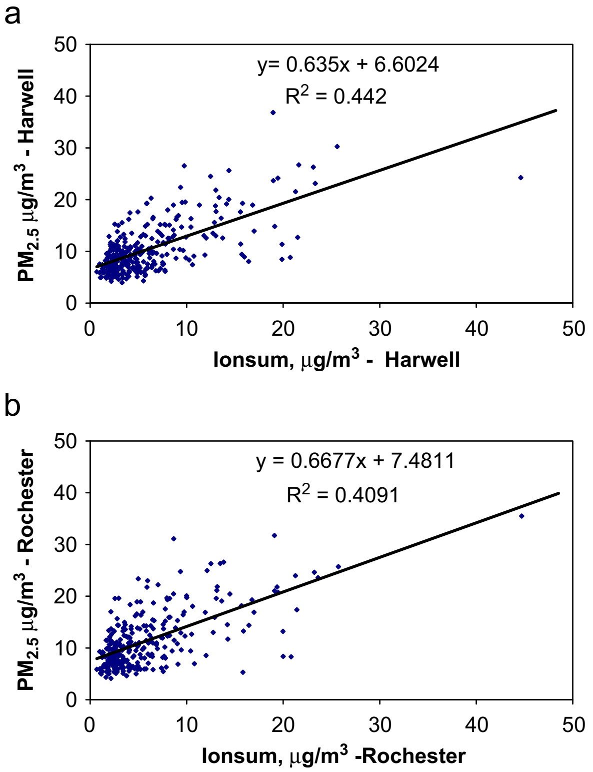

Fig. 3. The daily means of the $PM_{2.5}$ ($\mu g\,m^{-3}$) concentrations against the ion sum values ($\mu g\,m^{-3}$) at the regional background stations of Harwell (a) and Rochester (b) in UK, from 1998 to 2000.

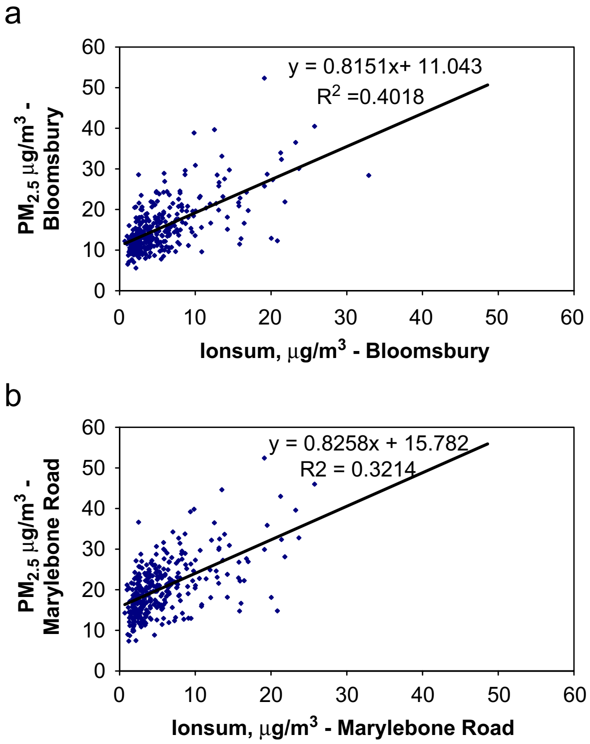

Fig. 4. The daily means of the $PM_{2.5}$ ($\mu g\,m^{-3}$) concentrations against the ion sum values ($\mu g\,m^{-3}$) at the urban stations of Bloomsbury (a) and Marylebone Road in London (b), from 1998 to 2000.

the linear regression analyses for all different modelling options and monitoring stations have been summarised in Table 1.

Let us consider first the correlations obtained using the ion sum as a proxy variable using (i) the data of both EMEP sites considered (Eskdalemuir and High Muffles), compared with (ii) only the data of either of these stations. In all cases examined, the utilisation of two stations provides better correlations. The same result was obtained previously by Karppinen et al. (2004), regarding three Finnish EMEP stations and two urban stations in Helsinki. The reasons for this are that the utilisation of two or three stations instead of one tends to smooth out the influences caused by local emission sources, and the results are also better representative on the average of the LRT air masses.

Using the ion sum instead of the sulphate concentration as a proxy variable resulted in substantially better correlations; this confirmed the initial study by Karppinen et al. (2004) which had used a different dataset. The ion sum variable represents a much larger fraction of the $PM_{2.5}$ mass compared with using merely the sulphate content.

As expected, the intercept values k_0 are systematically highest for the urban stations. The slope values k_1 are proportional to the LRT $PM_{2.5}$; their variation in Table 1 is therefore much smaller, compared with that of the intercepts.

The correlations obtained for the stations in UK are systematically lower than the corresponding values obtained previously for Finnish stations (Karppinen et al., 2004). There are several reasons for this. The EMEP stations in UK are at larger distances from the centre of London (approximately 315 and 460 km for High Muffles and Eskdalemuir, respectively), compared with the corresponding distances of the three selected Finnish EMEP stations from Helsinki (from 155 to 260 km). The Finnish EMEP sites are probably more representative of regional conditions for Helsinki, i.e., in the northerly, easterly and south-westerly directions, compared with the corresponding locations available in the UK with respect to London. It is recognised that there are more local and regional-scale

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

Table 1
A summary of the results of the statistical analyses regarding the stations in the UK, during 1998–2000

The station considered and its classification		The proxy variable and the station(s) used for its determination		The results of the linear regression analysis			
Station	Classification	Proxy variable	Stations	N	k_1	k_o	R^2
Harwell	Regional BG	Ion sum	Both sites	315	0.634	6.632	0.438
Rochester	Regional BG	Ion sum	Both sites	292	0.667	7.521	0.407
Bloomsbury	Urban centre	Ion sum	Both sites	318	0.813	11.163	0.399
Marylebone Road	Urban kerbside	Ion sum	Both sites	323	0.826	15.827	0.320
Harwell	Regional BG	Ion sum	Eskdalemuir	886	0.701	7.433	0.307
Rochester	Regional BG	Ion sum	Eskdalemuir	839	0.697	8.707	0.242
Bloomsbury	Urban centre	Ion sum	Eskdalemuir	842	0.900	11.705	0.272
Marylebone Road	Urban kerbside	Ion sum	Eskdalemuir	882	0.697	20.280	0.090
Harwell	Regional BG	Ion sum	High Muffles	343	0.488	6.760	0.353
Rochester	Regional BG	Ion sum	High Muffles	319	0.574	7.273	0.390
Bloomsbury	Urban centre	Ion sum	High Muffles	346	0.640	11.373	0.334
Marylebone Road	Urban kerbside	Ion sum	High Muffles	351	0.700	15.757	0.293
Harwell	Regional BG	Sulphate	Eskdalemuir	933	1.696	7.409	0.279
Rochester	Regional BG	Sulphate	Eskdalemuir	883	1.596	8.854	0.196
Bloomsbury	Urban centre	Sulphate	Eskdalemuir	889	2.006	11.934	0.226
Marylebone Road	Urban kerbside	Sulphate	Eskdalemuir	929	1.446	20.691	0.066
Harwell	Regional BG	Sulphate	High Muffles	969	1.611	6.930	0.260
Rochester	Regional BG	Sulphate	High Muffles	921	1.853	7.809	0.270
Bloomsbury	Urban centre	Sulphate	High Muffles	926	1.938	11.279	0.232
Marylebone Road	Urban kerbside	Sulphate	High Muffles	965	1.998	19.102	0.132

BG = background, N = number of daily measurements, k_1 = slope, k_o = intercept and R^2 = correlation coefficient squared.

pollution sources in UK that influence the concentration data than there are in Finland. The relatively closer vicinity of major $PM_{2.5}$ source areas in Central Europe for UK, compared with Northern Europe, most likely tend to cause more shorter-term temporal variations in the data.

3.3. Evaluation of the LRT contributions in UK and in Finland

The evaluated local and LRT contributions are summarised in Table 2. These contributions have been evaluated using the ratios of the terms k_0 and $k_1\langle C_{\text{ion}}\rangle$ (the brackets denote an average value over the period considered) and their sum. Using this definition implies that the computed LRT contribution actually depends on the temporal and spatial scales of the transformation of precursors to ionic components. According to this analysis, on average the LRT contribution accounted for 35–37% of the regional air $PM_{2.5}$ concentrations at the stations of Harwell and Rochester from 1998 to 2000. As expected, the corresponding contributions at the selected urban stations in London were smaller, 24–31%.

Redington and Derwent (2002) calculated sulphate and nitrate aerosol concentrations over Europe using the Lagrangian dispersion model NAME. They concluded that in 1996, on average 30% of UK aerosol was imported from the rest of Europe (excluding UK) in the summer and 25% in winter months. These numbers are not in disagreement with the above-mentioned LRT contributions; however, these are not directly comparable for several reasons. The years examined are different, the stations of Harwell and Rochester are representative merely for Southern UK (where the LRT contributions tend to be higher, compared with the whole of the UK), and the LRT contribution determined by the ion sum method is not necessarily equal to the contribution from the rest of Europe.

Previously, Karppinen et al. (2004) evaluated LRT contributions using the same method; the annually averaged contributions of LRT to the measured urban air $PM_{2.5}$ concentration in Helsinki in 1998–2000 at urban background and roadside sites were estimated to vary from 64% to 76%. This result is not substantially different from the source apportionment study of Ojanen et al. (1998), in which the LRT contribution to the $PM_{2.5}$

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

Table 2
Comparison of the evaluated local and long-range transported proportions

The station considered and its classification		The results of the linear regression analysis			
Station	Classification	k_0 ($\mu g\, m^{-3}$)	$k_0/(k_0+k_1\langle C_{ion}\rangle)$ (%)	$k_1\langle C_{ion}\rangle$ ($\mu g\, m^{-3}$)	$k_1\langle C_{ion}\rangle/(k_0+k_1\langle C_{ion}\rangle)$ (%)
UK stations, 1998–2000					
Harwell	Rural, EMEP	6.6	64	3.8	37
Rochester	Rural	7.5	65	4.0	35
Bloomsbury	Urban centre	11.2	69	4.9	31
Marylebone Road	Urban kerbside	15.8	76	5.0	24
Finnish stations, 2003–2004					
Utö, 2003	Rural, EMEP	2.6	33	5.4	67
Utö, 2004	Rural, EMEP	0.94	11	7.58	89
Virolahti, 2003	Rural, EMEP	1.7	20	6.49	80
Virolahti, 2004	Rural, EMEP	1.1	11	8.77	89

The brackets denote an average value. The evaluated local and long-range transported proportions were assumed to be equal to k_0 and $k_1\langle C_{ion}\rangle$, respectively.

concentration at the same roadside site (Vallila) was evaluated to be approximately 60–63%.

The evaluated LRT contributions at the EMEP sites of Utö and Virolahti in 2003–2004 ranged from 67% to 89%. As expected, these values are on the average higher than the above mentioned corresponding LRT contributions at the urban sites in Helsinki, and substantially higher than the LRT contributions evaluated above for rural sites in Southern UK.

4. Conclusions

The present study describes a new regression model for the assessment of the contribution of LRT to urban and regional air $PM_{2.5}$ concentrations. We also discuss model evaluation and application in UK, especially applied for London and in Finland, especially for Helsinki. The overarching objective was to find out whether the model could possibly be a practical tool of assessment in various European regions and could provide reliable estimates of LRT contributions.

We have studied the correlation of ion sum values with the $PM_{2.5}$ data measured at two EMEP stations in Finland; this is a comparison of ionic sum and the fine particulate matter fraction at the same station. This analysis is useful for methodological reasons, in order to evaluate model accuracy and reliability. The statistical correlations of the $PM_{2.5}$ concentrations with the ion sum values were very high (R^2 varied from 0.77 to 0.83) at both of the stations considered. The analysis was complicated by the fact that one of the stations (Virolahti) is frequently influenced by severe particular matter pollution originating from Russia. This result provides confidence that the ion sum is a reliable proxy variable for the LRT $PM_{2.5}$.

The comparison of different modelling options using the data measured in the UK showed that the regression model gave systematically substantially better results than the model using merely sulphate (SO_4) concentrations as the proxy variable for LRT. The same result was obtained previously for the Finnish stations (Karppinen et al., 2004). This result is important, as previously the use of sulphate concentration alone has been suggested for this purpose (e.g., APEG (Airborne Particles Expert Group), 1999). Similarly, in all cases considered, using the distance weighted ion sum based on data from two EMEP stations gave better correlation, compared with the option of using only one EMEP station.

On average, the evaluated LRT contribution accounted for 35–37% of the regional air $PM_{2.5}$ concentrations at the stations of Harwell and Rochester from 1998 to 2000. As expected, the corresponding contributions at the selected urban stations in London were smaller, 24–31%. These values are qualitatively in agreement with those presented by Redington and Derwent (2002); they calculated that in 1996, on average 25–30% of UK

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

aerosol was imported from the rest of Europe (excluding UK) in the summer and 25% in winter months. The evaluated LRT contributions at the EMEP sites of Utö and Virolahti in Finland in 2003–2004 ranged from 67% to 89%.

It is known that the regional background concentration of $PM_{2.5}$ consists mainly of ammonium nitrate, ammonium sulphate, and carbon compounds (e.g., Turnbull and Harrison, 2000). The accuracy of the model presented depends on the chemical composition of $PM_{2.5}$, especially the content of carbonaceous species; however, measurements of these species at the EMEP stations have not been published. The ion sum parameter defined also contains in part the measurements of two gaseous substances (HNO_3 and NH_3). If their concentrations were high, compared to the concentrations of the corresponding compounds in particulate form, there could be substantial inaccuracies in the model predictions.

The model parameter values presented in this study are specific for two countries and two particular urban areas, and these values may also have temporal trends. The model cannot therefore be used as such for the evaluation of long-term future scenarios. In applications of the model for another urban area, it is recommended that the model parameters (k_1 and k_0) corresponding to that particular area have first to be determined.

The model presented is simple and cost-efficient, and requires as input only routine data from well-equipped national EMEP stations. As shown in this study, the accuracy of the model regarding daily values is fairly good. It is therefore envisaged that the model could potentially be used as a practical tool of assessment of LRT contributions in various European regions.

Acknowledgements

This study is a part of the EC-funded FP5 projects Source Apportionment of Airborne Particulate Matter and Polycyclic Aromatic Hydrocarbons in Urban Regions of Europe (SAPPHIRE); Project number: EVK4-2002-00089) and OSCAR (EVK4-CT-2002-00083). Both these projects are part of the Cluster of European Air Quality Research (CLEAR). We also wish to thank for the contribution and funding of the national projects KOPRA and the Nordic cooperative project NORPAC. We gratefully acknowledge DEFRA for providing the $PM_{2.5}$ data for the UK sites used in this study. Mr. Keith Vincent is thanked for providing information on the EMEP measurements.

References

Abdalmogith, S.S., Harrison, R.M., 2005. The use of trajectory cluster analysis to examine the long-range transport of secondary inorganic aerosol in the UK. Atmospheric Environment 39, 6686–6695.

APEG (Airborne Particles Expert Group), 1999. Source apportionment of airborne particulate matter in the United Kingdom. Prepared on behalf of the Department of the Environment, Transport and the Regions, the Welsh Office, the Scottish Office and the Department of the Environment (Northern Ireland), available on 〈http://www.environment.detr.gov.uk/airq/〉, ISBN:0-7058-1771-7.

AQEG (Air Quality Expert Group), 2005. Particulate matter in the United Kingdom. Department of Environment, Food and Rural Affairs (DEFRA), London, UK, available from 〈http://www.defra.gov.uk/environment/airquality/aqeg〉, 444pp.

Deacon, R., Derwent, R.G., Harrison, R.M., Middleton, D.R., Moorcoft, S., 1997. Analysis and interpretation of measurements of suspended particulate matter at urban background sites in the United Kingdom. Science of the Total Environment 203, 17–36.

Eaton, S., Stacey, B., 2006. QA/QC data ratification and intercalibration report for the automatic urban and rural network, July–September 2005. Report prepared for Department for Environment, Food and Rural Affairs, Scottish Executive, Welsh Assembly Government and the DoE in Northern Ireland. AEA Technology plc, National Environmental Technology Centre, Report number AEAT/ENV/R/2115, Harwell.

EEA, 2006. Air quality and ancillary benefits of climate change policies. EEA Technical report 4/2006, Office for Official Publications of the European Communities, Luxembourg, 2006, Copenhagen, available at 〈http://reports.eea.europa.eu/technical_report_2006_4/en/Air_quality_and_ancillary_report.pdf〉, 60pp.

EMEP, 2001. The Co-operative programme for monitoring and evaluating of the long-range transmission of air pollutants in Europe, EMEP Manual for sampling and chemical analysis. EMEP/CCC-Report 1/95, revised in November 2001, Norwegian Institute for Air Research, Kjeller.

Jacobson, M.Z., 1999. Fundamentals of Atmospheric Modelling. Cambridge University Press, Cambridge, UK, 656pp.

Johansson, C., Hadenius, A., Johansson, P.Å., Jonson, T., 1999. NO_2 and particulate matter in Stockholm—concentrations and population exposure. In: The Stockholm Study on Health Effects of Air Pollution and their Economic Consequences. Swedish National Road Administration. Borlänge, Sweden. Available at 〈http://www.slb.nu/lvf〉.

Hayman, G., 2005. Management and Operation of the UK Acid Deposition Monitoring Network: Data Summary for 2004, AEAT/ENV/R/2093. AEA Technology, Harwell, Oxfordshire, UK.

Karppinen, A., Härkönen, J., Kukkonen, J., Aarnio, P., Koskentalo, T., 2004. Statistical model for assessing the portion of fine particulate matter transported regionally and

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036

long range to urban air. Scandinavian Journal of Work Environment and Health 30 (2), 47–53.

Karppinen, A., Kukkonen, J., Kauhaniemi, M., Härkönen, J., Nikmo, J., Sokhi, R.S., Luhana, L., Kousa, A., Alaviippola, B., Koskentalo, T., Aarnio, P., 2005. Evaluation and application of a model for the urban and regional scale concentrations of $PM_{2.5}$. In: Sokhi, R.S., Millán, M.M., Moussiopoulos, N. (Eds.), Proceedings of the Fifth International Conference on Urban Air Quality, Valencia, 29–31 March 2005, University of Hertfordshire, UK, 2005, ISBN: 1-898543-92-5, 4pp. (CD).

Ketzel, M., Berkowicz, R., 2004. Modelling the fate of ultrafine particles from exhaust pipe to rural background: an analysis of time scales for dilution, coagulation, and deposition. Atmospheric Environment 38, 2639–2652.

Kukkonen, J., Härkönen, J., Karppinen, A., Pohjola, M., Pietarila, H., Koskentalo, T., 2001. A semi-empirical model for urban PM_{10} concentrations, and its evaluation against data from an urban measurement network. Atmospheric Environment 35, 4433–4442.

Kukkonen, J., Bozo, L., Palmgren, F., Sokhi, R.S., 2003. Particulate matter in urban air. In: Moussiopoulos, N. (Ed.), Air Quality in Cities. SATURN, EUROTRAC-2 Subproject Final Report. Springer, Berlin, Heidelberg, New York, pp. 91–120.

Malcolm, A.L., Derwent, R.G., Maryon, R.H., 2000. Modelling the long-range transport of secondary PM_{10} to the UK. Atmospheric Environment 34, 881–894.

Ojanen, C., Pakkanen, T., Aurela, M., Mäkelä, T., Meriläinen, J., Hillamo, R., et al., 1998. Size distribution, chemical composition and sources of inhalable particles in Helsinki Area. Helsinki Metropolitan Area Council, YTV, Publication series C, p. 7 (in Finnish).

Pohjola, M.A., Kousa, A., Kukkonen, J., Härkönen, J., Karppinen, A., Aarnio, P., Koskentalo, T., 2002. The spatial and temporal variation of measured urban PM_{10} and $PM_{2.5}$ concentrations in the Helsinki metropolitan area. International Journal on Water, Air and Soil Pollution: Focus 2 (5–6), 189–201.

Pohjola, M.A., Pirjola, L., Kukkonen, J., Kulmala, M., 2003. Modelling of the influence of aerosol processes for the dispersion of vehicular exhaust plumes in street environment. Atmospheric Environment 37 (3), 339–351.

Redington, A.L., Derwent, R.G., 2002. Calculation of sulphate and nitrate aerosol concentrations over Europe using a Lagrangian dispersion model. Atmospheric Environment 36, 4425–4439.

Simpson, D., Fagerli, H., Jonson, J.E., Tsyro, S., Wind, P., Tuovinen, J.P., 2003. Transboundary acidification, eutrophication and ground level ozone in Europe, PART I. Unified EMEP model description. EMEP Status Report 1/2003, Norwegian Meteorological Institute, also available at 〈http://manhaz.cyf.gov.pl/manhaz/links/MODELLING/Unified%20EMEP%20Model/emep_report_1_part1_2003.pdf〉.

Tiitta, P., Raunemaa, T., Tissari, J., Yli-Tuomi, T., Leskinen, A., Kukkonen, J., Härkönen, J., Karppinen, A., 2002. Measurements and modelling of $PM_{2.5}$ concentrations near a major road in Kuopio, Finland. Atmospheric Environment 36, 4057–4068.

Turnbull, A.B., Harrison, R.M., 2000. Major component contributions to PM_{10} composition in the UK atmosphere. Atmospheric Environment 34, 3129–3137.

Tsyro, S., 2002. First estimates of the effect of aerosol dynamics in the calculation of PM_{10} and $PM_{2.5}$. EMEP, Meteorological Synthesizing Center—West, Norwegian Meteorological Institute, Research Note No. 76, Blindern, Norway, 41pp.

Please cite this article as: Kukkonen, J., et al., Evaluation and application of a statistical model for assessment of long-range transported proportion of $PM_{2.5}$ in the United Kingdom and in..., Atmospheric Environment (2007), doi:10.1016/j.atmosenv.2007.02.036