Atmospheric
Chemistry
and Physics
Discussions

1 Input-adaptive linear mixed-effects model for estimating alveolar
2 Lung Deposited Surface Area (LDSA) using multipollutant datasets

3 Pak Lun Fung[1,2], Martha A. Zaidan[1,3], Jarkko V. Niemi[4], Erkka Saukko[5], Hilkka Timonen[6], Anu Kousa[4],
4 Joel Kuula[6], Topi Rönkkö[7], Ari Karppinen[6], Sasu Tarkoma[8], Markku Kulmala[1,3], Tuukka Petäjä[1,3] and
5 Tareq Hussein[1,9]

6 [1]Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Finland
7 [2]Helsinki Institute of Sustainability Science, Faculty of Science, University of Helsinki, Finland
8 [3]Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing
9 University, Nanjing 210023, China
10 [4]Helsinki Region Environmental Services Authority (HSY), P.O. Box 100, FI-00066 Helsinki, Finland
11 [5]Pegasor Ltd., FI-33100 Tampere, Finland
12 [6]Atmospheric Composition Research, Finnish Meteorological Institute, FI-00560 Helsinki, Finland
13 [7]Aerosol Physics Laboratory, Physics Unit, Tampere University, FI-33720 Tampere, Finland
14 [8]Department of Computer Science, Faculty of Science, University of Helsinki, Finland
15 [9]Department of Physics, the University of Jordan, Amman 11942, Jordan

16 *Correspondence to*: Pak Lun Fung (pak.fung@helsinki.fi), Tareq Hussein (tareq.hussein@helsinki.fi)

17 **Abstract.** Lung deposited surface area (LDSA) has been considered to be a better metric to explain nanoparticle toxicity
18 instead of the commonly used particulate mass concentration. LDSA concentrations can be obtained either by direct
19 measurements or by calculation based on the empirical lung deposition model and measurements of particle size distribution.
20 However, the LDSA or size distribution measurements are neither compulsory nor regulated by the government. As a result,
21 LDSA data are often scarce spatially and temporally. In light of this, we develop a novel statistical model, named input-
22 adaptive mixed-effects (IAME) model, to estimate LDSA based on other already existing measurements of air pollutant
23 variables and meteorological conditions. During the measurement period in 2017–2018, we retrieved LDSA data measured by
24 Pegasor AQ Urban and other variables at a street canyon (SC, average LDSA = 19.7±11.3 µm^2 cm^{-3}) site and an urban
25 background (UB, average LDSA = 11.2±7.1 µm^2 cm^{-3}) site in Helsinki, Finland. For the continuous estimation of LDSA,
26 IAME model is automatised to select the best combination of input variables, including a maximum of three fixed effect
27 variables and three time indictors as random effect variables. Altogether, 696 sub-models were generated and ranked by the
28 coefficient of determination (R^2), mean absolute error (MAE) and centred root-mean-square differences ($cRMSD$) in order. At
29 the SC site, the LDSA concentrations were best estimated by mass concentration of particle of diameters smaller than 2.5 µm
30 (PM$_{2.5}$), total particle number concentration (PNC) and black carbon (BC), all of which are closely connected with the vehicular
31 emissions. At the UB site the LDSA concentrations were found to be correlated with PM$_{2.5}$, BC and carbon monoxide (CO).
32 The accuracy of the overall model was better at the SC site (R^2 = 0.80, MAE = 3.7 µm^2 cm^{-3}) than at the UB site (R^2 = 0.77,
33 MAE = 2.3 µm^2 cm^{-3}) plausibly because the LDSA source was more tightly controlled by the close-by vehicular emission
34 source. The results also demonstrate that the additional adjustment by taking random effects into account improves the sensitivity
35 and the accuracy of the fixed effect model. Due to its adaptive input selection and inclusion of random effects,
36 IAME could fill up missing data or even serve as a network of virtual sensors to complement the measurements at reference
37 stations.

38 1 Introduction

39 Particulate matter is one of the key components determining urban air pollution. Particulate matter can be described by a
40 combination of varying concentration (number, surface area and mass) and chemical composition. The mass concentrations of
41 particulate matter are dominated by large particles whereas the number concentrations are governed by sub-micron particles
42 (particle diameter (d$_p$) <1 µm), particularly ultrafine particles (UFP, d$_p$< 0.1 µm) (e.g. Petäjä et al., 2007; Rönkkö et al., 2017;

Atmospheric
Chemistry
and Physics
Discussions
Open Access
EGU

43 Zhou et al., 2020). Particulate matter of varying sizes, carrying various harmful substances, have been known for playing a

44 major role in adverse health effects (Dockery et al., 1993; Oberdorster, 2012; Shiraiwa et al., 2017) in particular to respiratory

45 system. A particle could be deposited in lung airways upon inhalation (Oberdörster et al., 2005) through three main

46 mechanisms: inertial impaction, gravitational sedimentation and Brownian diffusion. Interception, and electrostatic forces are

47 to a lesser extent. An airborne particle might be inhaled either through nasal or oral passage and enter the respiratory tract.

48 Coarser particles (5–30 μm) are usually partly deposited in the head airway by the inertial impaction mechanism because they

49 cannot follow the air streamline. Some finer particles (1–5 μm) are deposited in the tracheobronchial region, mainly through

50 gravitational sedimentation while some are removed by mucociliary clearance (Gupta and Xie, 2018). The remaining sub-

51 micron particles diffuse by Brownian motion and penetrate deeply into the alveolar region, which is considered to be the most

52 vulnerable section in lungs because removal mechanisms might be insufficient (Gupta and Xie, 2018). Inhaled particulate

53 matter could also function as a carrier, or as a transport vector, for many viruses, including the SARS-CoV-2 virus (COVID-

54 19, Prather et al., 2020), which is responsible for the pandemic recently declared by the World Health Organization (WHO).

55 Particulate matter may, therefore, increase the effectiveness of the virus spread in the aerosol as it creates a microenvironment

56 suitable for its persistence (Liu et al., 2018a) . Regular exposure to particulate matter increases the chance to suffer from acute

57 and chronic diseases (Brown et al., 2001; Oberdörster et al., 2005), and the susceptibility and severity of the COVID-19

58 patients' symptoms (Fennelly, 2020). In light of this, besides commonly monitored particulate matter number concentration

59 and mass concentration, the surface area of a particle is also an important factor when considering the harmfulness of

60 particulate matter (Duffin et al., 2002). In particular, the total surface area of particles which are deposited in alveolar section

61 of human lungs, known as Lung Deposited Surface Area (LDSA), is of the greatest concern because in vitro nanoparticle

62 toxicity has been demonstrated to be better explained when the lung burden was expressed as total particle surface area instead

63 of atmospheric particulate matter mass (e.g. Brown et al., 2001; Oberdorster, 2012; Schmid and Stoeger, 2016).

64

65 LDSA can be considered as an intermediary parameter between particle mass and particle number concentration as it cannot

66 be simply inferred from either of those parameters. Moreover, due to the various deposition efficiency with respect to particle

67 sizes, the quantification of LDSA is not simple. Conventionally, LDSA concentrations can be retrieved by (1) derivation from

68 particle size distribution with a deposition model or (2) direct measurements.

69

70 By fitting experimental lung deposition data on human beings, empirical deposition models are developed with the use of the

71 lung deposition model modified by Yeh and Schum (1980). Examples include the International Commission on Radiological

72 Protection (ICRP) Human Respiratory Tract Model (ICRP, 1994), the NCRP model (NCRP, 1997) and Multiple Path Particle

73 Dosimetry (MPPD) model (Anjilvel and Asgharian, 1995). Different conceptual particle deposition models vary primarily

74 with respect to lung morphometry and mathematical modelling techniques, rather than by using different deposition equations.

75 The three whole lung deposition models define regions of the human lungs (head airway, tracheobronchial and alveolar) for

76 any combination of particle size and breathing pattern (Hofmann, 2009). Among all models, single-path models, such as ICRP

77 model, are often used over multiple-path models due to their simplicity and their applicability to an average path without

78 requiring detailed knowledge of the branching structure of lungs. Owing to a higher potential health risk, LDSA in alveolar

79 region is often of greatest concern and it can be calculated by summing up the products of the surface concentration across

80 particle size spectrum and their corresponding deposition efficiency based on the selected deposition model.

81

82 Apart from numerical computation method, LDSA could also be measured by accredited instruments. LDSA concentration in

83 many urban environments is mainly driven by the particles smaller than 400 nm (Asbach et al., 2009; Kuuluvainen et al.,

84 2016), generated vastly by anthropogenic sources such as vehicular exhaust emissions (Karjalainen et al., 2016) and residential

85 wood combustion (Tissari, 2008) which typically produce large amount of small particles. The impact of larger particles (>400

86   nm) might be significant due to regional background in very polluted cities (e.g. Delhi, Salo et al., 2021a) or very low-quality

87   residential burning in detached housing areas (e.g. HMA, Pirjola et al., 2017). These small particles cannot be measured

88   precisely with methods relying solely on optical detection (e.g. no artificial growing of particles) as the light scattering intensity

89   of these particles is weak (Kulkarni et al., 2011). Hence alternative approaches are required. One approach is filter sampling

90   of aerosolised material followed by gas adsorption method (e.g. Lebouf et al., 2011). Another more common approach is using

91   a diffusion charging based technique where particles are charged with a unipolar corona charger (Fissan et al., 2006). This

92   method enables measurement of ultrafine particles and, more specifically, the LDSA concentration with good accuracy (Todea

93   et al., 2015) and stable performance in long term measurements (Rostedt et al., 2014). Nanoparticle Surface Area Monitor

94   (NSAM) has been used for decades (e.g. Asbach et al., 2009; Hama et al., 2017; Kiriya et al., 2017; Hennig et al., 2018), and

95   several other instruments and sensors, including DiSCmini, Testo Inc. (e.g. Eeftens et al., 2016; Habre et al., 2018) and

96   Partector, Naneos Ltd. (e.g. Cheristanidis et al., 2020), and Pegasor AQ Urban, Pegasor Ltd. (e.g. Kuuluvainen et al., 2018;

97   Kuula et al., 2020), using similar measuring techniques, are developed later on. Recently, this diffusion charging based LDSA

98   measurement has been combined with electrical cascade impactor method, which enables high time resolution measurements

99   of particle LDSA size distributions (Lepistö et al., 2020). Using these instruments in campaigns and continuous measurements,

100   LDSA concentrations and size distribution measurements in various environments have been reported across the globe in the

101   past decade (**Table 1**). When comparing LDSA concentrations measured by different instruments, it should be noted that the

102   instruments' limitation should be taken into account in experimental LDSA studies, which will be further discussion in Sect.

103   2.2.

104

105   Although each of these methods is capable of measuring aerosol surface area concentrations, the corresponding uncertainties

106   (Asbach et al., 2017) and cost hinder the widespread use in monitoring networks. Even though the instruments are available,

107   missing data often takes place due to instruments maintenance and data corruption. Kuula et al. (2020) demonstrated high

108   correlations of measured LDSA concentrations with black carbon (BC) and nitrogen oxide (NO_x) under certain circumstances.

109   Traffic activities have been observed to be significant source contribution to the LDSA concentrations (Järvinen et al., 2015).

110   A clear correlation was also found between the emission factors of exhaust plume BC and LDSA in on-road studies for city

111   buses (e.g. Järvinen et al., 2019). These highly correlating relationships provide good grounds for estimating LDSA

112   concentrations and short-term trends by the other pollutants measured at the same site with the use of data mining-based

113   approach as statistical models. Data mining-based approach exploits statistical or machine learning techniques to detect

114   patterns between predictors and dependent variables in the time series data. They do not demand in-depth understanding of air

115   pollutant dynamics, but evaluation by experts is still required to determine whether the models work properly. Simple yet

116   apprehensible models, such as multiple linear regression (MLR, e.g. Fernández-Guisuraga et al., 2016) and generalized

117   additive models (GAM, e.g. Chen et al., 2019), are commonly utilised as white-box models in air pollutant proxy studies.

118   Furthermore, more sophisticated machine learning black-box models, such as artificial neural network (ANN, e.g. Cabaneros

119   et al., 2019; Zaidan et al., 2019), nonlinear autoregressive network with exogenous inputs (NARX, Zaidan et al., 2020) and

120   support vector regression (SVR, e.g. Fung et al., 2021), have been intensively investigated in recent years. They work better

121   in terms of accuracy; however, they provide limited transparency and accountability regarding the outcomes (Rudin, 2019;

122   Fung et al., 2021).

123

124   Apart from model structures, the criteria of selecting variables in multipollutant datasets for model development have received

125   considerable attention over the years, and a large number of methods have been proposed (Miller, 2002). Traditional methods,

126   like stepwise procedures, which is a combination of forward selection and backward elimination (e.g. Liu et al., 2018b; Chen

127   et al., 2019), can be unstable because it uses restricted search through the space of potential models, which eventually causes

128   inherent problem of multiple hypothesis testing (Breiman, 1996; Faraway, 2014). Another approach named regularization has

129   emerged as a successful method to reduce the data dimension in an automated way, yet deal poorly with multi-collinear

130   variables, for example Least Absolute Shrinkage and Selection Operator (LASSO, e.g. Fung et al., 2021; Šimić et al., 2020),

131   ridge regression (e.g. Chen et al., 2019) and ELASTINET (e.g. Chen et al., 2019). Criterion-based procedures, which choose

132   the best predictor variables according to some criteria (e.g. coefficient of determination, residual, etc), are sensitive to outliers

133   and influential points, but involve a wider search and compare models in a preferable manner. Examples are best subset

134   regression (e.g. Chen et al., 2019), input adaptive proxy (e.g. Fung et al., 2020; Fung et al., 2021), etc. Hastie et al. (2020)

135   compared some of the models using the three approaches and concluded that no single feature selection method uniformly

136   outweighs the others. Despite the extensive research of feature selection methods, the inclusion of random effects together

137   with the fixed effects as linear mixed-effects (LME) model has received little attention (e.g. Font et al., 2019; Tong et al.,

138   2020) in air pollution research, let alone LDSA study in particular. This inclusion of random effects could acknowledge a

139   possible effect coming from a factor where specific and fixed values are not of interest.

140

141   In this study, we combine the use of criterion-based feature selection method and the inclusion of random effects, and develop

142   a novel input-adaptive mixed effects (IAME) model to estimate alveolar LDSA concentrations, which is the first study of this

143   context to our best knowledge. The description of LDSA measurements and the techniques of IAME model are outlined in

144   Sect. 2 and 3, respectively. Section 4 presents the characteristics of alveolar LDSA, including its seasonal variability, weekend

145   effect and diurnal pattern, in four types of environments. We also aim to investigate the correlation with other air pollutants.

146   In Sect. 5, we evaluate the performance of the IAME proxy ($LDSA_{IAME}$) with the measured alveolar LDSA by Pegasor AQ

147   Urban ($LDSA_{Pegasor}$), ICRP lung deposition model derived LDSA ($LDSA_{ICRP}$) and another modelled alveolar LDSA by IAP

148   ($LDSA_{IAP}$) as well as the benefits and implication of this alveolar LDSA model. It should be noted that this study discusses

149   LDSA in alveolar region, unless stated otherwise.

2 Measurement description

2.1 Measurement sites

152   We retrieved aerosol, gaseous and meteorological data from two types of measurement sites, i.e., street canyon (SC, 2017–

153   2018) and urban background (UB, 2017–May 2018), in Helsinki Metropolitan Area (HMA) described in more detail below.

154   Data from detached housing (DH, 2017) and regional background (RB, 2017) sites were also included in the study to provide

155   comparison and data from the background concentrations. Situated on a relatively flat land at the coast of Gulf of Finland,

156   HMA has land area of 715 km^2 and population of about 1.13 million inhabitants. Helsinki can be classified as continental or

157   marine climate depending on the air flows and the pressure system. Figure S1 and Table S1 show the detailed site description.

158   **Street canyon site (SC)**: Mäkelänkatu urban supersite is operated by the Helsinki Region Environmental Services Authority

159   (HSY, Kuuluvainen et al., 2018). The station is located at 3 km from the city centre in a street canyon in the immediate vicinity

160   to one of the main roads leading to downtown Helsinki. The street, with speed limit of 50 km h^{-1}, consists of six lanes and two

161   tramlines. The annual mean traffic volume in 2018 per workday was 28 100 vehicles, 11% of which were recorded as the

162   heavy duty vehicles. The traffic loads are especially high during rush hours at 8 a.m. and 5 p.m. (Figure S2). The street canyon

163   of width of 42 m is surrounded by rows of buildings of 17 m high, which weaken the dispersion process of the direct vehicular

164   emissions. All the inlets for the measuring devices are positioned approximately at a height of 4 m from the ground level.

165   **Urban background site (UB)**: The Station for Measuring Ecosystem-Atmosphere Relations III (SMEAR III, Järvi et al.,

166   2009) in Kumpula, situated on a rocky hill at 26 m above sea level, is about 4 km northeast from the Helsinki centre. The

167   surroundings of this urban background station are heterogeneous, constituting of residential buildings, small roads, parking

168   lots, patchy forest and low vegetation from different direction. One main road (45 000 vehicles per workday) is located at the

169   distance of 150 m east from the site. Trace gases and meteorological conditions are measured at a height of 4 m and 32 m,

170 respectively, at a triangular lattice tower while aerosol measurements are conducted inside a container approximately 4 m

171 above the ground. The site is co-operated by Finnish Meteorological Institute (FMI) and the University of Helsinki (UHEL).

172 **Detached housing site (DH)**: Three measurement stations, Rekola (DH1), Itä-Hakkila (DH2) and Hiekkaharju (DH3), were

173 chosen since they represent a sub-urban residential area surrounded by detached houses. These sites are mainly affected by the

174 wood combustion emissions from residential activities, especially in cold weather conditions. Emissions from traffic source

175 also account for a small portion of the whole pollution. It is estimated that 90 % of the households burn wood to warm up

176 houses and saunas, less than 2 % of which use wood burning as the main heating source in detached houses in HMA (Hellén

177 et al., 2017).

178 **Regional background site (RB)**: The RB site is located about 23 km away from the Helsinki city centre at Luukki, surrounded

179 by a wooded outdoor recreational area right at the edge of the Greater Helsinki golf course. The measuring station is in an

180 open place away from busy traffic routes and large point sources. As a result, this site can represent background concentration

181 levels outside the urban area without any main local sources.

182 **2.2 Instruments**

183 **LDSA measurements:** The sensor unit and the core of the Pegasor AQ Urban is practically another instrument called a Pegasor

184 PPS-M sensor (Pegasor Ltd., Finland) originally designed for automotive exhaust emission measurements (e.g. Maricq, 2013;

185 Amanatidis et al., 2017). The operation of the sensor is based on diffusion charging of particles and the measurement of electric

186 current without the collection of particles. The diffusion charging of particles is carried out by a corona-ionized flow that is

187 mixed with the ambient sample air in an ejector diluter inside the sensor. The sampling lines and the sensor unit are heated

188 40°C above the ambient temperature (1) to dry the aerosol sample, (2) to prevent interference from humidity, and (3) to prevent

189 any water condensation inside the sensor. The performance of the Pegasor PPS-M sensors for long-term ambient measurements

190 has been improved after they were tested in Helsinki (Järvinen et al., 2015) and Beijing (Dal Maso et al., 2016). The suggestions

191 have been considered for the design of the current form of the Pegasor AQ Urban in this study.

192 The Pegasor AQ Urban (dimension: 320 mm×250 mm×1000 mm), which consists of a weatherproof cover, clean air supply,

193 and the abovementioned Pegasor PPS-M sensor, has been designed such that its response to LDSA is not to be subjected to

194 meteorological fluctuation for outdoor operation. Kuuluvainen et al. (2016) used two Pegasor AQ Urban devices during a 2

195 week period at an urban street canyon and an urban background measurement station in Helsinki, Finland whereas Kuula et

196 al. (2019) later used instrument in a 3 month long campaign at the same urban street canyon station. These studies demonstrated

197 that the output signal of the Pegasor AQ Urban correlated well with other devices measuring LDSA concentrations such as the

198 Partector and DiSCmini. Kuula et al. (2020) further validated the accuracy and stability of Pegasor AQ Urban at the street

199 canyon station by comparing the measured values of one full year with DMPS reference instruments ($R^2 = 0.90$, RMSE = 4.1

200 $\mu m^2 \, cm^{-3}$). The instrument is optimized to measure the alveolar LDSA concentrations of particles in ~10–400 nm size range.

201 Pegasor AQ Urban tends to underestimate LDSA of particle larger than about 400 nm. In typical urban environments, most of

202 the particles from local combustion sources are in the size below the threshold (Kuuluvainen et al., 2016; Pirjola et al., 2017).

203 However, the impact of larger particles (>400 nm) to alveolar LDSA might be significant, for example a recent study on LDSA

204 concentrations in polluted urban environment in India observed high LDSA contribution from relatively large accumulation

205 mode particles although the experiment was conducted in close proximity of traffic (Salo et al., 2021a) and in mining

206 environment the mineral dust and other pollutants being typically in larger particle sizes can also contribute to the LDSA

207 concentrations (Salo et al., 2021b). In HMA, the impact of >400 nm might also be significant during $PM_{2.5}$ long-range transport

208 episodes or when there are many particles from very low-quality residential burning in detached housing areas (Pirjola et al.,

209 2017).

210 **Aerosol measurements:** Differential mobility particle sizer (DMPS) in combination of a differential mobility analyser (DMA)

211 and a condensation particle counter (CPC) measures aerosol size distribution (Kulkarni et al., 2011). Vienna DMA and

212   Airmodus A20 CPC (measurements of particle size range 6–800 nm) are used at the SC site while a twin DMPS (Hauke-type

213   DMA and TSI Model 3025 CPC + Hauke-type DMA and TSI Model 3010 CPC, merged particle size range 3–1000 nm) are

214   used at the UB site. Both instruments make use of the bipolar charging of aerosol particles, followed by classification of

215   particles into size classes according to their electrical equivalent mobility. In addition to particle size distribution, total particle

216   number concentration (PNC, in cm^{-3}) is calculated by summation. Particle mass concentration of diameter less than 2.5 µm

217   ($PM_{2.5}$, in $\mu g\ m^{-3}$) and less than 10 µm (PM_{10}, in $\mu g\ m^{-3}$) are measured continuously with ambient particulate monitor TEOM

218   1405 at the SC site and TEOM 1405-D at the UB site. Black carbon (BC, in $\mu g\ m^{-3}$) mass concentration is measured by a

219   multi-angle absorption photometer (MAAP) Thermo Scientific 5012 with a PM_1 inlet. The measured absorbance is converted

220   to BC mass concentration by using a fixed 6.6 $m^2\ g^{-1}$ mass absorption coefficient at wavelength of 637 nm. $PM_{2.5}$, PM_{10} and

221   BC are recorded in $\mu g\ m^{-3}$.

222   **Ancillary measurements:** Trace gas concentrations (in ppb), including nitrogen oxide (NO), nitrogen dioxide (NO_2), their

223   sum nitrogen oxide (NO_x), ozone (O_3) and carbon monoxide (CO) are determined with a suite of gas analysers. In addition,

224   supporting meteorological variables, including air temperature (Temp), relative humidity (RH), air pressure (P), wind speed

225   (WS), wind direction (WD) and photosynthetically active radiation (PAR), are measured at SC and UB. Figure S3 show the

226   meteorological conditions during the measurement period. A list of variables collected is shown in Table S2.

227   **3. Method**

228   **3.1 Data pre-processing**

229   The collected data was quality checked by the corresponding operating organisation, HSY, FMI and UHEL. No additional

230   pre-processing was done for general analysis. For proxy development, outliers were detected using the interquartile range

231   (IQR) rule, which is applicable for non-Gaussian distribution sample. We calculated the cut-off for outliers as 2 times the IQR,

232   subtracted this cut-off from the 25th percentile and added it to the 75th percentile to give the actual limits on the data. We

233   applied a natural logarithm transformation to all the skewed-distributed aerosol and trace gases measurements in order to keep

234   the distribution of each parameter following a normal distribution. Since wind direction is a circular variable, it is resolved

235   into North-South (WD–N) and East-West (WD–E) vector components by trigonometric functions.

236   **3.2 Size-fractionated lung deposited surface area ($LDSA_{ICRP}$)**

237   Alveolar deposition fraction (DF_{AL}) as a function of particle size with the unit density is determined with the ICRP Human

238   Respiratory Tract Model by the following equation (ICRP, 1994).

$$DF_{AL} = \left(\frac{0.0155}{d_p}\right)\left(\exp\left(-0.416\left(\ln d_p + 2.84\right)^2\right) + 19.11 \exp\left(-0.482\left(\ln d_p - 1.362\right)^2\right)\right)$$ **(1)**,

239   where d_p is the aerodynamic diameter (µm) of spherical particles with the unit density (1 $g\ cm^{-3}$). The equation is determined

240   in two parts with respect to the two different peaks in the deposition curve in Figure 1. The peak near the size of 20 nm can be

241   approximated to represent the Brownian deposition, whereas the peak between 1 µm and 2 µm represents the inertial

242   deposition. From the particle number size distribution, we calculated the particle surface area distribution assuming each

243   particle is monodisperse sphere of standard density at standard conditions. By Eq. (1), a deposition factor for each particle size

244   bin (26 size bins at SC and 49 at UB) were calculated. Size-fractionated LDSA was then computed by multiplying the surface

245   area concentration with DF_{AL} in the corresponding size class. Total LDSA calculated by the ICRP lung model ($LDSA_{ICRP}$) can

246   be obtained by summing up the all the size-fractionated LDSA values. In this study, the alveolar $LDSA_{ICRP}$ was calculated

247   based on DMPS measurements in SC and UB. Thus, while the alveolar LDSA measured by Pegasor ($LDSA_{Pegasor}$) represent

248   the ~10–400 nm size range, the alveolar $LDSA_{ICRP}$ represent 6–800 nm and 3–1000 nm size range in SC and UB, respectively.

249 **3.3 Novel Input-adaptive mixed-effects (IAME) model**

250 Input-adaptive mixed-effects (IAME) model is a combination of input-adaptive proxy (IAP) and linear mixed-effects (LME)

251 model. IAP was first introduced by Fung et al. (2020) and has been demonstrated reliable and flexible to fill up missing values

252 by taking input variables adaptively with robust ordinary least square regression models. IAP has been able to estimate BC

253 concentration by other air quality indicators with a satisfactory performance in two different categorised urban environments,

254 street canyon (adjusted $R^2 = 0.86$–0.94) and urban background (adjusted $R^2 = 0.74$–0.91). Some models outperformed IAP in

255 accuracy performance, but its transparent model structure and ability to impute missing values still make it a preferred option

256 as a virtual sensor (Fung et al., 2021).

257

258 In this study, we primarily stick to the strength to select input variables adaptively with the introduction of mixed effects. The

259 mixed effect approach is a generalization of the linear model that can incorporate both fixed (i.e. causing a main

260 effect/interaction) and random effects (i.e. causing variance/variability in responses), allowing the account of several sources

261 of variations (Chudnovsky et al., 2012). As seen in Figure 2, We picked the direct air pollutant measurement from the station

262 (variables of high correlation: $PM_{2.5}$, BC and NO_2 and other supporting variables: PM_{10}, O_3, NO_x, NO, CO and PNC) and

263 meteorological data of higher correlation (Temp, RH, P, PAR, WS, WD–N, WD–E) as the fixed variables because the air

264 pollutants can indicate the sources of LDSA which largely come from combustion and meteorological data could influence the

265 dispersion and dilution of LDSA. They are the most direct factors to the fluctuation of LDSA concentrations. Due to the strong

266 seasonal variation, weekend effects and diurnal pattern in urban air pollutant concentrations (Fung et al., 2020), the variance

267 in responses might depend on the time indicators that are not the primary cause of the concentration variability, but they

268 indirectly alter human-induced activities, such as traffic amounts. To take them into account, we created three time hierarchical

269 sub-groups (12 months of year, 7 days of week and 24 hours of day) as the inputs of random effect variables.

270

271 The regression equation of IAME is similar to the equation of IAP, except that IAME includes additional intercepts term for

272 random effects as below:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \sum_{j=1}^{q} b_{ij} + e_i$$

(2),

273 where y_i is the ith estimated LDSA concentration. The first term on the right β_0 indicates the fixed intercept of the equation.

274 The second term represents the total contribution by the direct measurement of variable x as fixed effects with a slope β at

275 each data point i. A maximum of three inputs from the total 16 fixed variables are selected to from 696 sub-models (Figure 2).

276 The inputs for random effects are indicated by b as intercepts of the corresponding three hierarchical sub-groups. A Gaussian

277 error term is indicated by e. The explanation of Eq (2), is visualised in Figure 2.

278

279 One of the assumptions of LME models is that the random effects, together with the error term, have the following prior

280 distribution:

$$b \sim N\left(0, \sigma^2 D(\theta)\right)$$

(3),

281 where D is a q-by-q symmetric and positive semidefinite matrix, parameterized by a variance component vector θ, q is the

282 number of variables in the random-effects term, and σ^2 is the observation error variance. We use an optimiser, restricted

283 maximum likelihood, commonly known as ReML, with the value 1×10^{-6} as the relative tolerance on gradient of objective

284 function and 1×10^{-12} as absolute tolerance on step size. The use of ReML over the conventional ML could produce unbiased

285 estimates of variance and covariance parameters (Lindstrom and Bates, 1988).

286

287 After the sub-model formation, the dataset is randomly divided into five portions. 80% of the data are allocated for 4-fold cross
288 validation to remove variance of accuracy. The results of all the folds are averaged and the sub-models are ranked by several
289 evaluation metrics, which are further demonstrated in Figure 2 and described in Sect. 3.4. Some of the sub-models are subject
290 to rejection under two conditions: (1) strong multi-collinearity among the fixed parameters (variance inflation factor (VIF)
291 exceeding a threshold of 5) and (2) violation of the normality assumption of residuals also known as heteroscedasticity (fail in
292 Kolmogorov-Smirnov (K-S) test, $p < 0.05$). Based on the situation of missing data, the automatised IAME model will search
293 for the best sub-model option from the ranking chart. Hence, each data point might be estimated differently depending on the
294 available data. The number of data points being estimated by each sub-model is reported to show their frequency of usage.

295 **3.4 Evaluation metrics**

296 In order to evaluate the model performance quantitatively, we use the following metrics:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4},$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{5},$$

$$cRMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((y_i - \bar{y}) - (\hat{y}_i - \tilde{y}))^2} \tag{6},$$

$$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \tilde{y})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \tilde{y})^2}} \tag{7},$$

$$NSD = \frac{SD_{predicted}}{SD_{reference}} = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{y}_i - \tilde{y})^2}}{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2}} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - \tilde{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{8},$$

297 where $\hat{y}_i$ and $\hat{y}_i$ are i^{th} measured data point and estimated variable by the model, respectively. $\bar{y}$ and $\tilde{y}$ are the expected value
298 of the measured and modelled dataset, respectively. N is the number of complete data input to the model. Coefficient of
299 determination (R^2) is a measure of how close the data lie to the fitted regression line. It, however, does not consider the biases
300 in the estimation. Therefore, we further validated the models with mean absolute error (MAE) and centred root-mean-square
301 differences ($cRMSD$), where MAE measures the arithmetic mean of the absolute differences between the members of each
302 pair, whilst $cRMSD$ calculates the square root of the average squared difference between the forecast and the observation pairs.
303 $cRMSD$ is more sensitive to larger errors than MAE. Furthermore, together with $cRMSD$, Pearson correlation coefficient (r)
304 and normalised standard deviation (NSD) of the modelled data set are also studied. r describes the correlation between the
305 measured and modelled data whereas NSD measures the relative spread of the data. Due to their unique mathematical
306 relationship, the three metrics can be portrayed on Taylor's diagram, which has been used for sub-model selection purpose.
307 We ranked our sub-models first by R^2, followed by MAE and $cRMSD$. r and NSD serve as additional evidence when we
308 explain the model performance.

309 **3.5 Two-sample t-tests**

310 We assessed the temporal and spatial impact on the IAME model by comparing the means of absolute differences between the
311 hourly measured and modelled LDSA in different time windows at both stations. Two-sample t-tests were performed on the
312 two populations of absolute differences abovementioned to determine whether the difference between these is statistically
313 significant. A significance level α of 5% is chosen as the probability of rejecting the null hypothesis when it is true, denoted
314 as p.

8

Atmospheric
Chemistry
and Physics
Discussions

4 LDSA measurement characterization

4.1 General characteristics of LDSA$_{Pegasor}$ in Helsinki metropolitan area

The annual mean alveolar LDSA concentrations at four station types SC (2017–2018), UB (2017–May 2018), DH (2018) and RB (2018) are 19.7±11.3 µm^2 cm^{-3}, 11.2±7.1 µg m^{-3}, 11.7±8.6 µm^2 cm^{-3} and 7.6±5.4 µm^2 cm^{-3}, respectively (Table 2). The DH and RB site are included to give more substantial interpretation of data because the LDSA concentrations at RB can be viewed as background measurements and the local LDSA increments in HMA can be represented by the LDSA at the hotspot measurement site subtracted by the LDSA at the RB site. The timeseries of LDSA concentrations at the SC and the UB site are presented in Figure 3 and Fig. S4, where the missing data of LDSA for the whole measurement period is 3% and 30%, respectively. When comparing with the same site type in other cities around the globe, LDSA concentrations detected in HMA are the lowest among the European cities with reported values, and about one-fifth that in Japan (**Table 1**). Some literatures also report LDSA at tracheobronchial region but most just consider LDSA at alveolar which is considered to bring most harm to human's lungs.

The diurnal pattern of LDSA at RB is not observable on workdays or over weekends (Figure 4, upper panel). The relatively low variability can be explained by the scarcity of human activities. We can then regard the LDSA at RB as the background concentrations mainly influenced by the regionally and long-range transported aerosol and meteorological variation. As the concentrations at RB is stable throughout the different hours of day; therefore, the diurnal pattern of LDSA concentration is apparently indistinguishable between the measured concentration and the local increments. At the UB and DH site, the magnitudes and the patterns of the average hourly LDSA concentrations at workdays are comparable, and both show bimodal curves, one peak at 6−9 a.m., the other at 9−11 p.m.. The former has a larger peak during the morning peak hour because of the vehicular emissions (Timonen et al., 2013; Teinilä et al., 2019) while the latter has a larger peak in the evening attributed mainly by the residential burning (Hellén et al., 2017; Helin et al., 2018; Luoma et al., 2021). Over weekends, the peaks in the morning are not identifiable and the evening peaks are amplified due to enhanced human activities. Similar diurnal variation at residential area was observed for BC emitted by residential combustion by Helin et al. (2018). At the SC site, the morning peak on weekends is not obvious because of the lack of work-related traffic. It appears that a similar bimodal curve can be seen during workdays, but the evening peak is seen during the evening traffic rush hour around 4−6 p.m.. The reason is that the main contributor of LDSA at the SC site is traffic and combustion processes and the diurnal variability mainly depends on the citizen's movement by vehicles in the city. Over weekends, the average hourly LDSA concentrations are the minimum at 5 a.m. and they increase and remain at a high level at 2 p.m. until the late night. The level of LDSA concentrations at DH is comparable with that at UB site. However, the amplitudes of the evening peak is higher than that of the morning peak both on workdays and weekends due to elevated residential combustion.

However, the monthly variability of background measurements at the RB site is stronger compared to the diurnal pattern and the calculation of local increment is necessary. With no intense point sources, the variations at RB are probably due to horizontal dispersion and advection of aerosol particles and vertical dilution controlled by the boundary layer dynamics. In the summer, when solar radiation is persistently stronger, the boundary layer becomes elevated due to surface heating and associated thermal turbulence. This turbulence would dilute the concentration of pollutants at the surface. Another plausible reason could be the higher regional and long-range transported LDSA in the summer, as demonstrated by Kuula et al. (2020) and Barreira et al. (2021). The lower panel in Figure 4 shows the LDSA local increments after subtraction of the LDSA at the RB site. For instance, the local LDSA increments at DH are the highest in the winter probably due to local small-scale wood combustion (and traffic). However, without subtracting the background concentrations, the LDSA concentrations at DH are higher in the summer than in the winter (due to high regional background concentrations in summer), as was observed also by

357 Kuula et al. (2020). This piece of evidence can help in the source apportionment. The variation of diurnal and seasonal LDSA

358 for all sites are visualised in Fig. S5.

4.2 The connection between LDSA and other parameters

360 Alveolar LDSA concentration, as a single number, comprises particles across the whole particle size spectrum measured (e.g.

361 Pegasor AQ Urban ~10–400 nm). InHMA, the two local main sources of particles contributing to LDSA are vehicular

362 combustion and residential wood combustion emissions. Upon the two combustion processes, particles of different sizes and

363 different gaseous pollutants are emitted. A study by Lamberg et al. (2011) has shown that the geometric mean diameter of

364 residential wood combustion is typically 70–150 nm whereas Barreira et al. (2021) presented that the typical particle size for

365 vehicular combustion can be smaller than 50 nm. By calculating the proportion of LDSA with respect to different pollutant

366 parameters BC, NO_x, PNC (dominated by UFP), and $PM_{2.5}$, we could identify the contribution of LDSA across the hour of day

367 (Fig. S6 for workdays and Fig. S7 for weekends). Since the vehicular combustion emits smaller particles which elevate the

368 LDSA concentration but meanwhile do not substantially influence the value of $PM_{2.5}$ (e.g. Salo et al., 2021a); therefore,

369 $LDSA/PM_{2.5}$ has a diurnal pattern similar to the LDSA concentrations which peaks in the morning rush hour during workdays.

370 Conversely, LDSA/BC, LDSA/PNC and $LDSA/NO_x$ have a higher value before the morning rush hour and they plunged in

371 the morning rush hour. This can be explained by the fact that vehicular combustion emits high concentration of BC, PNC and

372 NO_x (Reche et al., 2015) compared to its contribution to LDSA concentration. In other words, the role of regional background

373 is higher for LDSA compared to those of NO_x, BC and PNC. At the UB site, the average LDSA/BC at all hours remain at a

374 constant level in the winter while the variability of the ratio is much higher in the summer. The general LDSA/PNC ratio at

375 UB is steadily 2−3 times higher than that at all hours in all seasons because the proportion of larger particles at UB is usually

376 higher than SC. This large variability again validate the heterogeneity of source of LDSA.

377

378 The integrated alveolar LDSA with a various size ranges was calculated to explore the correlation of size-fractionated LDSA

379 and other parameters in our multipollutant dataset. No single fractionated LDSA correlates well with meteorological

380 parameters at both sites (Figure 5). Out of all fractions, alveolar LDSA of the whole spectrum ($LDSA_{6-800}$) and $LDSA_{250-400}$,

381 which explains majority of LDSA, correlates best with other air pollutants. In general, alveolar LDSA has high correlation

382 with BC. BC correlates the best with $LDSA_{100-250}$, which is in alignment with the reported values from previous literature

383 (Gramsch et al., 2014; Ding et al., 2016). As expected, $PM_{2.5}$ show better correlation with the LDSA of larger particles because

384 larger particles contributes more to $PM_{2.5}$ mass concentration values. In the meanwhile, PM_{10} has fair correlation with all

385 selected size bins. NO_2 correlates highly with LDSA of smaller particles, indicating the dominant role of local traffic exhausts.

386 CO has a higher correlation with LDSA of 400−800 nm since CO concentrations are more affected by regionally transported

387 pollutants. O_3 has a fair correlation with LDSA of all sections because the formation of O_3 is mostly secondary and the chemical

388 interactions with pollutants are more complicated than the other compounds. In general, the correlation of LDSA with other

389 air pollutant parameters is higher at the SC site than that at the UB site (Fig. S8). The high correlation of LDSA with BC, $PM_{2.5}$

390 and NO_2, which agrees with the results by Kuula et al. (2020), proves the possibility of developing a model to estimate LDSA

391 concentrations.

5 Model evaluation

5.1 Sub-model diagnostics

394 Following the evaluation attributes described in Sect. 3.4,

395   Table 3 depicts the descriptive statistics of the overall model evaluation on its testing set. The overall model at the SC site is

396   able to explain 80% of the variability of the testing set of the measured data. The R^2 in the winter is 0.86 being the highest

397   while the worst R^2 is shown in the summer, i.e., 0.70. The MAE and $cRMSD$ are the smallest during weekend with R^2 not

398   particularly high ($R^2 = 0.72$) probably because the LDSA concentration itself is relatively low in that period. The overall

399   performance is generally worse in UB in terms of R^2, except during weekends that R^2 is 10% higher.

400

401   For individual sub-models, their performance could be seen on the Taylor's diagram in Figure 6 (Taylor, 2001). Each marker

402   represents one sub-model, the contribution of which to the outcome of the final model is displayed in various colours. The

403   sub-model performance can be evaluated by the distance of the sub-model marker and the red point, which represents the

404   reference station, i.e., the perfect model. The location of each marker indicates its individual performance in terms of r,

405    $cRMSD$ and NSD. At the SC site, the narrow distribution of the sub-models on the Taylor's diagram gives a clue that they are

406   very similar in terms of model performance of LDSA estimation. The five mostly used sub-models are concentrated within

407   the region where r is 0.85–0.87, $cRMSD$ is 5.67−5.77 µm^2 cm^{-3} and NSD is 0.75−0.79 (Table 4). The values of their

408   evaluation metrics are close to each other where R^2 and MAE differ in the narrow range of 10% ($R^2 = 0.72$–0.74, $MAE = 3.8$

409   µm^2 cm^{-3}). It infers that if one metric is prioritised over another, the rank of the sub-models can be greatly different. Although

410   no individual sub-models show r greater than 0.9, the overall model comprising the outcomes by all the sub-models remains

411   high ($R^2 = 0.80$, $MAE = 3.8$ µm^2 cm^{-3}). The best sub-model is also the most used one, which accounts for 81% of the total

412   data points while the two succeeding sub-models constitute another 16%. This also indicates that the input adaptivity function

413   of the suggested method supplement 19% of estimates which would be a missing estimate if a single model with fixed predictor

414   variables is used. Four out of the five most used sub-models contain BC as an input predictor with the combination of other

415   two air pollutants or meteorological parameters. In case BC is missing at a certain time stamp, the sub-model without BC as

416   an input could be used. It further supports the input adaptive function.

417

418   At the UB site, the sub-model performance is more scattered on the Taylor's diagram (Figure 6). The five most used sub-

419   models have varying metrics ($r = 0.77−0.92$, $cRMSD = 2.5−3.9$ µm^2 cm^{-3} and $NSD = 0.63−0.89$, see Table 5). Although some

420   show exceptionally good performance, the overall model has a slightly worse performance than that in street canyon. The best

421   sub-model estimates 49% of the total measurement, followed by 17%. The third and fourth most used sub-models, which form

422   up to 30% of the estimates, have rather moderate performance ($R^2 = 0.58$ and 0.69). Considering all possible outcomes, the

423   overall model is still able to explain 77% of the total variance. CO and PNC dominate in the top five used sub-models. BC,

424   NO$_x$ and meteorological parameters, like RH and WD-N are also involved in the final LDSA estimation.

425

426   By checking the variance inflation factor (VIF) of all 696 sub-models, 4.6% and 2.2% are rejected respectively. The higher

427   rejection rate at SC can be explained by the fact that some of the predictor variables are highly correlating to each other and

428   the inclusion of them would result in an inflation of multi-collinearity of the sub-model, from which biases arise. At UB, since

429   the source of LDSA is more varied and the correlation of LDSA with other pollutants is generally lower, the probability of the

430   VIF of the individual sub-models exceeding the threshold is lower.

431   **5.2 Temporal difference in comparison with other models**

432   Figure 7 presents the comparison of measured LDSA (LDSA$_{Pegasor}$), deposition model derived LDSA (LDSA$_{ICRP}$) and the

433   LDSA modelled by IAP and IAME (LDSA$_{IAP}$ and LDSA$_{IAME}$) as a timeseries plot between 14 and 28 February 2017. This

434   particular time window is selected because it suffers the least in data missing for all the respective instruments at both sites.

435   This figure during this period can also showcase the difference in magnitudes of the diurnal shape over workdays and weekends

436   (shaded regions in Figure 7). At the SC site, the estimates by both LDSA$_{IAP}$ and LDSA$_{IAME}$ could generally catch up with the

Atmospheric
Chemistry
and Physics
Discussions

437   diurnal cycle of the measured data. However, the models underestimate the peak if the change of the measured LDSA

438   concentration is sudden and relatively large. Despite the small difference observed in the figure, the blue dotted line

439   representing $LDSA_{IAME}$ often stays closer to the measured LDSA concentration (black line). When we smooth out all the

440   estimates at each hour, the ability for IAME to catch the morning peak on workdays is much better. At the UB site, IAME

441   underestimates the LDSA concentration by almost 50% and 25% in the morning on 15 and 23 February 2017, respectively.

442   The overestimation reaches 100% during the midnight between 26 and 17 February 2017.

443

444   A more generalised diurnal cycle can be found in Figure 8. The error bars of the modelled $LDSA_{IAP}$ and $LDSA_{IAME}$ are
445   consistently smaller than that of $LDSA_{Pegasor}$ and $LDSA_{ICRP}$. It might be due to the reason that the model fails to catch the
446   extreme values although it manages to catch the general diurnal cycle. Since outliers are removed in the pre-processing stage
447   and the model penalises the extreme values, the model tends to give a more centralised estimate. It is a trade-off between the
448   option with better coefficients of determination but stronger extreme errors and that with better estimations at tails but
449   derivation of averaged estimation. This circumstance is more apparent on workdays than weekends. Furthermore, $LDSA_{IAME}$
450   could follow the diurnal cycle of $LDSA_{Pegasor}$ much better than $LDSA_{IAP}$, especially during the start of the peak hours over
451   workdays at the SC site where the LDSA concentrations jump to a high level. $LDSA_{IAME}$ can explain 80% and 77% of the
452   variability of the reference measurements at SC and UB, respectively (Table 6

453   Table 6), and compared to LDSA$_{IAP}$'s 77% and 66%, LDSA$_{IAME}$ perform better in terms of accuracy. In addition, the slightly

454   smaller MAE and the closer to 1 NSD of the LDSA$_{IAME}$ suggest that the mean absolute error is improved and the spread of the

455   estimation distribution is closer to the reference measurement by taking random effects into account.

456

457   Furthermore, we assessed the temporal and spatial impact on the IAME model by comparing the means of absolute differences

458   between the hourly LDSA$_{Pegasor}$ and LDSA$_{IAME}$ in different time windows at both stations. A descriptive statistic is presented

459   in Table 7. We used two-sample t-tests to assess whether the distribution of absolute differences were statistically significant.

460   At SC, the p value of the t-tests at all selected windows are below 0.05, which demonstrate that the performance at different

461   seasons, days of week and hours of day of absolute differences between the measured and modelled LDSA were significantly

462   different at the confidential level of 95%. At the UB site, the difference between the two selected hour periods is not statistically

463   significant. The same applies to the difference between winter and spring. There are no statistically sufficient evidence to

464   validate the difference among the rest of the selected time period. In other words, with the use of random effects of time

465   constraint, the overall models still perform differently at different time windows most of the time. This indicates that IAME

466   still needs improvements on minimising temporal differences.

467   **6 Conclusion**

468   In this study, we develop a novel input-adaptive mixed-effects (IAME) proxy, to estimate alveolar LDSA by other already

469   existing air pollutant variables and meteorological conditions in Helsinki Metropolitan Area. During the measurement period

470   2017–2018, we retrieved LDSA measurements measured by Pegasor AQ Urban (alveolar LDSA in the ~10−400 size range)

471   and other variables in a street canyon (SC, average LDSA = 19.7±11.3 µm^2 cm^{-3}) site and an urban background (UB, average

472   LDSA = 11.2±7.1 µm^2 cm^{-3}) site in Helsinki, Finland. Furthermore, three detached housing sites (DH, average LDSA =

473   11.7±8.6 µm^2 cm^{-3}) and a regional background site (RB, average LDSA = 7.6±5.4 µm^2 cm^{-3}) are also included as reference

474   and background source estimation, respectively. At the SC site, LDSA concentrations are closely correlated with traffic

475   emission. The ratio to black carbon (LDSA/BC), to particle number concentration (LDSA/PNC), and to nitrogen oxide

476   (LDSA/NO$_x$) have a higher value before the morning peak and it reaches its minimum during the morning peak since the role

477   of regional background is higher for LDSA compared to those of NO$_x$, BC and PNC. However, the ratio of LDSA to mass

478   concentration of particles of diameter smaller than 2.5 µm (LDSA/PM$_{2.5}$) perform differently since the freshly vehicular

479   emitted particles are smaller than 50 nm, which do not contribute much to PM$_{2.5}$ mass concentration.

480

481   For the continuous estimation of LDSA, IAME is automatised to select the best combination of input variables, including a

482   maximum of three fixed effect variables and three time indictors as random effect variables. Altogether, 696 sub-models are

483   generated and ranked by the coefficient of determination (R^2), mean absolute error (MAE) and centred root-mean-square

484   differences ($cRMSD$) in order. At the SC site, LDSA concentrations can be best estimated by PM$_{2.5}$, PNC and BC, all of which

485   are closely connected with the vehicular emissions, while they are found correlating with PM$_{2.5}$, BC and carbon monoxide

486   (CO) the best at the UB site. At both sites, PM$_{2.5}$ also indicates the regionally and long-range transported pollutants, which is

487   a significant source of LDSA concentrations. The accuracy of the overall model is higher at the SC site ($R^2 = 0.80$, $MAE =$

488   3.7 µm^2 cm^{-3}) than at the UB site ($R^2 = 0.77$, $MAE = 2.3$ µm^2 cm^{-3}) plausibly because the LDSA source was more tightly

489   controlled by the close-by vehicular emission source. This model could catch the temporal pattern of LDSA; however, the

490   two-sample t-tests of the residuals at all selected time windows show that their distributions are different. This indicates that

491   the model still performs differently at different time windows. Despite this, the novel IMAE model works better in explaining

492   the variability of the measurements than the previously suggested IAP model as indicted by a higher R^2 and lower MAE in

Atmospheric
Chemistry
and Physics
Discussions

493   both sites. This adjustment by taking random effects into account improves the sensitivity and the accuracy of the fixed effect

494   model IAP.

495

496   The models alone cannot replace the need for reference measurements. However, the IAME proxy could serve as virtual

497   sensors to complement the measurements at reference stations in case of missing data. The two measurement sites in this study

498   serve as a pilot of the proxy development, and the next step is to extend the work to the existing network of several measurement

499   stations within the Helsinki metropolitan region. With similar configurations, we could fill up the voids with the information

500   from the other stations after conscientious calibration. For example, in this paper, the two measurement sites are characterised

501   as street canyon and urban background. In a different setup, we may assume the similarity of the same type of environment

502   and utilise the measurements as replacement.

503

504   Furthermore, this continuous LDSA estimation could be useful in updating some of the current air quality application, for

505   instance GreenPaths application which searches for the best route to wished destination with the least exposure to air pollution

506   (Poom et al., 2020) and ENFUSER air quality model which provide accurate spatio-temporal estimation for air pollutants in

507   Helsinki (Johansson et al., 2015).

508

Data availability

510   The air quality data and meteorological data are available from HSY website (https://www.hsy.fi/avoindata) and through

511   SmartSMEAR online tool (https://smear.avaa.csc.fi/).

Author contributions

513   PLF performed formal analysis and writing – original draft of the manuscript. PLF, MAZ, TP and TH conceptualized and

514   designed the methodology of this work. MAZ, ST, MK, TP and TH provided supervision in this research activity. ES (Pegasor

515   Ltd.), JVN and AKo (HSY), and HT, JK and AKa (FMI) provided instruments and data for the campaign. All the co-authors

516   (MAZ, JVN, ES, HT, AKo, JK, TR, Aka, ST, MK, TP and TH) reviewed and commented on the manuscript.

Competing interests

518   Prof. Markku Kulmala and Prof. Tuukka Petäjä are members of the editorial board of the journal Atmospheric Chemistry and

519   Physics. Dr. Erkka Saukko works in Pegasor Ltd. which is the manufacturer of Pegasor AQ Urban.

Atmospheric
Chemistry
and Physics
Discussions

Open Access

531   **References**

532   Albuquerque, P. C., Gomes, J. F., and Bordado, J. C.: Assessment of exposure to airborne ultrafine particles in the urban
533   environment of Lisbon, Portugal, J. Air Waste Manag. Assoc., 62, 373-380, https://doi.org/10.1080/10962247.2012.658957,
534   2012.

535   Amanatidis, S., Maricq, M. M., Ntziachristos, L., and Samaras, Z.: Application of the dual Pegasor Particle Sensor to real-
536   time measurement of motor vehicle exhaust PM, J. Aerosol Sci., 103, 93-104, https://doi.org/10.1016/j.jaerosci.2016.10.005,
537   2017.

538   Anjilvel, S., and Asgharian, B.: A multiple-path model of particle deposition in the rat lung, Fund. Appl. Toxicol., 28, 41-50,
539   https://doi.org/10.1006/faat.1995.1144, 1995.

540   Asbach, C., Fissan, H., Stahlmecke, B., Kuhlbusch, T., and Pui, D.: Conceptual limitations and extensions of lung-deposited
541   Nanoparticle Surface Area Monitor (NSAM), J. Nanopart. Res., 11, 101-109, https://doi.org/10.1007/s11051-008-9479-8,
542   2009.

543   Asbach, C., Alexander, C., Clavaguera, S., Dahmann, D., Dozol, H., Faure, B., Fierz, M., Fontana, L., Iavicoli, I., Kaminski,
544   H., MacCalman, L., Meyer-Plath, A., Simonow, B., van Tongeren, M., and Todea, A. M.: Review of measurement techniques
545   and methods for assessing personal exposure to airborne nanomaterials in workplaces, Sci. Total Environ., 603, 793-806,
546   https://doi.org/10.1016/j.scitotenv.2017.03.049, 2017.

547   Barreira, L. M. F., Helin, A., Aurela, M., Teinilä, K., Friman, M., Kangas, L., Niemi, J. V., Portin, H., Kousa, A., Pirjola, L.,
548   Rönkkö, T., Saarikoski, S., and Timonen, H.: In-depth characterization of submicron particulate matter inter-annual variations
549   at a street canyon site in northern Europe, Atmos. Chem. Phys., 21, 6297-6314, https://doi.org/10.5194/acp-21-6297-2021,
550   2021.

551   Breiman, L.: Heuristics of instability and stabilization in model selection, Ann. Stat., 24, 2350-2383,
552   https://doi.org/10.1214/aos/1032181158, 1996.

553   Brown, D. M., Wilson, M. R., MacNee, W., Stone, V., and Donaldson, K.: Size-dependent proinflammatory effects of ultrafine
554   polystyrene particles: a role for surface area and oxidative stress in the enhanced activity of ultrafines, Toxicol. Appl. Pharm.,
555   175, 191-199, https://doi.org/10.1006/taap.2001.9240, 2001.

556   Buonanno, G., Marini, S., Morawska, L., and Fuoco, F. C.: Individual dose and exposure of Italian children to ultrafine
557   particles, Sci. Total Environ., 438, 271-277, https://doi.org/10.1016/j.scitotenv.2012.08.074, 2012.

558   Cabaneros, S. M., Calautit, J. K., and Hughes, B. R.: A review of artificial neural network models for ambient air pollution
559   prediction, Environ. Modell. Softw., 119, 285-304, https://doi.org/10.1016/j.envsoft.2019.06.014, 2019.

560   Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.
561   A., Katsouyanni, K., Janssen, N. A. H., Martin, R. V., Samoli, E., Schwartz, P. E., Stafoggia, M., Bellander, T., Strak, M.,
562   Wolf, K., Vienneau, D., Vermeulen, R., Brunekreef, B., and Hoek, G.: A comparison of linear regression, regularization, and
563   machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide, Environ. Int., 130,
564   104934, https://doi.org/10.1016/j.envint.2019.104934, 2019.

565   Cheristanidis, S., Grivas, G., and Chaloulakou, A.: Determination of total and lung-deposited particle surface area
566   concentrations, in central Athens, Greece, Environ. Monit. Assess., 192, 627, https://doi.org/10.1007/s10661-020-08569-8,
567   2020.

568   Chudnovsky, A. A., Lee, H. J., Kostinski, A., Kotlov, T., and Koutrakis, P.: Prediction of daily fine particulate matter
569   concentrations using aerosol optical depth retrievals from the Geostationary Operational Environmental Satellite (GOES), J.
570   Air Waste Manag. Assoc., 62, 1022-1031, https://doi.org/10.1080/10962247.2012.695321, 2012.

571   Dal Maso, M., Gao, J., Järvinen, A., Li, H., Luo, D., Janka, K., and Rönkkö, T.: Improving urban air quality measurements by
572   a diffusion charger based electrical particle sensors-A field study in Beijing, China, Aerosol Air Qual. Res., 16, 3001-3011,
573   https://doi.org/10.4209/aaqr.2015.09.0546, 2016.

574   Ding, A., Huang, X., Nie, W., Sun, J., Kerminen, V. M., Petäjä, T., Su, H., Cheng, Y., Yang, X. Q., and Wang, M.: Enhanced
575   haze pollution by black carbon in megacities in China, Geophys. Res. Lett., 43, 2873-2879,
576   https://doi.org/10.1002/2016GL067745, 2016.

577   Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E.: An association
578   between air pollution and mortality in six US cities, New Engl. J. Med., 329, 1753-1759,
579   https://doi.org/10.1056/NEJM199312093292401, 1993.

580   Duffin, R., Tran, C., Clouter, A., Brown, D., MacNee, W., Stone, V., and Donaldson, K.: The importance of surface area and
581   specific reactivity in the acute pulmonary inflammatory response to particles, Ann. Occup. Hyg., 46, 242-245,
582   https://doi.org/10.1093/annhyg/mef684, 2002.

583   Eeftens, M., Meier, R., Schindler, C., Aguilera, I., Phuleria, H., Ineichen, A., Davey, M., Ducret-Stich, R., Keidel, D., Probst-
584   Hensch, N., Kunzli, N., and Tsai, M. Y.: Development of land use regression models for nitrogen dioxide, ultrafine particles,

585   lung deposited surface area, and four other markers of particulate matter pollution in the Swiss SAPALDIA regions, Environ.
586   Health, 15, 53, https://doi.org/10.1186/s12940-016-0137-9, 2016.
587   Faraway, J. J.: Linear models with R, CRC press, 2014.
588   Fennelly, K. P.: Particle sizes of infectious aerosols: implications for infection control, Lancet Resp. Med.,
589   https://doi.org/10.1016/S2213-2600(20)30323-4, 2020.
590   Fernández-Guisuraga, J. M., Castro, A., Alves, C., Calvo, A., Alonso-Blanco, E., Blanco-Alegre, C., Rocha, A., and Fraile,
591   R.: Nitrogen oxides and ozone in Portugal: trends and ozone estimation in an urban and a rural site, Environ. Sci. Pollut. R.,
592   23, 17171-17182, https://doi.org/10.1007/s11356-016-6888-6, 2016.
593   Fissan, H., Neumann, S., Trampe, A., Pui, D., and Shin, W.: Rationale and principle of an instrument measuring lung deposited
594   nanoparticle surface area, J. Nanopart. Res., 53-59, https://doi.org/10.1007/s11051-006-9156-8, 2006.
595   Font, A., Guiseppin, L., Blangiardo, M., Ghersi, V., and Fuller, G. W.: A tale of two cities: is air pollution improving in Paris
596   and London?, Environ. Pollut., 249, 1-12, https://doi.org/10.1016/j.envpol.2019.01.040, 2019.
597   Fung, P. L., Zaidan, M. A., Sillanpaa, S., Kousa, A., Niemi, J. V., Timonen, H., Kuula, J., Saukko, E., Luoma, K., Petaja, T.,
598   Tarkoma, S., Kulmala, M., and Hussein, T.: Input-Adaptive Proxy for Black Carbon as a Virtual Sensor, Sensors (Basel), 20,
599   https://doi.org/10.3390/s20010182, 2020.
600   Fung, P. L., Zaidan, M. A., Timonen, H., Niemi, J. V., Kousa, A., Kuula, J., Luoma, K., Tarkoma, S., Petäjä, T., Kulmala, M.,
601   and Hussein, T.: Evaluation of white-box versus black-box machine learning models in estimating ambient black carbon
602   concentration, J. Aerosol Sci., https://doi.org/10.1016/j.jaerosci.2020.105694, 2021.
603   Gramsch, E., Reyes, F., Oyola, P., Rubio, M., López, G., Pérez, P., and Martínez, R.: Particle size distribution and its
604   relationship to black carbon in two urban and one rural site in Santiago de Chile, J. Air Waste Manag. Assoc., 64, 785-796,
605   https://doi.org/10.1080/10962247.2014.890141, 2014.
606   Gupta, R., and Xie, H.: Nanoparticles in daily life: applications, toxicity and regulations, J. Environ. Pathol. Tox., 37,
607   https://doi.org/10.1615/JEnvironPatholToxicolOncol.2018026009, 2018.
608   Habre, R., Zhou, H., Eckel, S. P., Enebish, T., Fruin, S., Bastain, T., Rappaport, E., and Gilliland, F.: Short-term effects of
609   airport-associated ultrafine particle exposure on lung function and inflammation in adults with asthma, Environ. Int., 118, 48-
610   59, https://doi.org/10.1016/j.envint.2018.05.031, 2018.
611   Hama, S. M. L., Ma, N., Cordell, R. L., Kos, G. P. A., Wiedensohler, A., and Monks, P. S.: Lung deposited surface area in
612   Leicester urban background site/UK: Sources and contribution of new particle formation, Atmos. Envrion., 151, 94-107,
613   https://doi.org/10.1016/j.atmosenv.2016.12.002, 2017.
614   Hastie, T., Tibshirani, R., and Tibshirani, R.: Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based
615   on Extensive Comparisons, Stat. Sci., 35, 579-592, https://doi.org/10.1214/19-STS733, 2020.
616   Helin, A., Niemi, J. V., Virkkula, A., Pirjola, L., Teinilä, K., Backman, J., Aurela, M., Saarikoski, S., Rönkkö, T., Asmi, E.,
617   and Timonen, H.: Characteristics and source apportionment of black carbon in the Helsinki metropolitan area, Finland, Atmos.
618   Envrion., 190, 87-98, https://doi.org/10.1016/j.atmosenv.2018.07.022, 2018.
619   Hellén, H., Kangas, L., Kousa, A., Vestenius, M., Teinilä, K., Karppinen, A., Kukkonen, J., and Niemi, J. V.: Evaluation of
620   the impact of wood combustion on benzo [a] pyrene (BaP) concentrations; ambient measurements and dispersion modeling in
621   Helsinki, Finland, Atmos. Chem. Phys., 17, 3475-3487, https://doi.org/10.5194/acp-17-3475-2017, 2017.
622   Hennig, F., Quass, U., Hellack, B., Kupper, M., Kuhlbusch, T. A. J., Stafoggia, M., and Hoffmann, B.: Ultrafine and Fine
623   Particle Number and Surface Area Concentrations and Daily Cause-Specific Mortality in the Ruhr Area, Germany, 2009-2014,
624   Environ. Health Persp., 126, 027008, https://doi.org/10.1289/EHP2054, 2018.
625   Hofmann, W.: Modelling particle deposition in human lungs: modelling concepts and comparison with experimental data,
626   Biomarkers, 14, 59-62, https://doi.org/10.1080/13547500902965120, 2009.
627   ICRP: PUBLICATION 66: Human Respiratory Tract Model for Radiological Protection, Pergamon Press, New York, 1994.
628   Järvi, L., Hannuniemi, H., Hussein, T., Junninen, H., Aalto, P. P., Hillamo, R., Mäkelä, T., Keronen, P., Siivola, E., and Vesala,
629   T.: The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface–atmosphere interactions
630   in Helsinki, Finland, Boreal Environ. Res., 19, 86-109, 2009.
631   Järvinen, A., Kuuluvainen, H., Niemi, J. V., Saari, S., Dal Maso, M., Pirjola, L., Hillamo, R., Janka, K., Keskinen, J., and
632   Rönkkö, T.: Monitoring urban air quality with a diffusion charger based electrical particle sensor, Urban Clim., 14, 441-456,
633   https://doi.org/10.1016/j.uclim.2014.10.002, 2015.
634   Järvinen, A., Timonen, H., Karjalainen, P., Bloss, M., Simonen, P., Saarikoski, S., Kuuluvainen, H., Kalliokoski, J., Dal Maso,
635   M., Niemi, J. V., Keskinen, J., and Rönkkö, T.: Particle emissions of Euro VI, EEV and retrofitted EEV city buses in real
636   traffic, Environ. Pollut., 250, 708-716, https://doi.org/10.1016/j.envpol.2019.04.033, 2019.
637   Johansson, L., Epitropou, V., Karatzas, K., Karppinen, A., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., and
638   Kompatsiaris, I.: Fusion of meteorological and air quality data extracted from the web for personalized environmental
639   information services, Environ. Modell. Softw., 64, 143-155, https://doi.org/10.1016/j.envsoft.2014.11.021, 2015.
640   Karjalainen, P., Timonen, H., Saukko, E., Kuuluvainen, H., Saarikoski, S., Aakko-Saksa, P., Murtonen, T., Bloss, M., Maso,
641   M. D., Simonen, P., Ahlberg, E., Svenningsson, B., Brune, W. H., Hillamo, R., Keskinen, J., and Rönkkö, T.: Time-resolved
642   characterization of primary particle emissions and secondary particle formation from a modern gasoline passenger car, Atmos.
643   Chem. Phys., 16, 8559-8570, https://doi.org/10.5194/acp-16-8559-2016, 2016.
644   Kiriya, M., Okuda, T., Yamazaki, H., Hatoya, K., Kaneyasu, N., Uno, I., Nishita, C., Hara, K., Hayashi, M., Funato, K., Inoue,
645   K., Yamamoto, S., Yoshino, A., and Takami, A.: Monthly and Diurnal Variation of the Concentrations of Aerosol Surface
646   Area in Fukuoka, Japan, Measured by Diffusion Charging Method, Atmosphere (Basel), 8,
647   https://doi.org/10.3390/atmos8070114, 2017.

648   Kulkarni, P., Baron, P. A., and Willeke, K.: Aerosol measurement: principles, techniques, and applications, John Wiley &
649   Sons, 2011.
650   Kuula, J., Kuuluvainen, H., Rönkkö, T., Niemi, J. V., Saukko, E., Portin, H., Aurela, M., Saarikoski, S., Rostedt, A., Hillamo,
651   R., and Timonen, H.: Applicability of Optical and Diffusion Charging-Based Particulate Matter Sensors to Urban Air Quality
652   Measurements, Aerosol Air Qual. Res., 19, 1024-1039, https://doi.org/10.4209/aaqr.2018.04.0143, 2019.
653   Kuula, J., Kuuluvainen, H., Niemi, J. V., Saukko, E., Portin, H., Kousa, A., Aurela, M., Rönkkö, T., and Timonen, H.: Long-
654   term sensor measurements of lung deposited surface area of particulate matter emitted from local vehicular and residential
655   wood combustion sources, Aerosol Sci. Tech., 54, 190-202, https://doi.org/10.1080/02786826.2019.1668909, 2020.
656   Kuuluvainen, H., Rönkkö, T., Järvinen, A., Saari, S., Karjalainen, P., Lähde, T., Pirjola, L., Niemi, J. V., Hillamo, R., and
657   Keskinen, J.: Lung deposited surface area size distributions of particulate matter in different urban areas, Atmos. Envrion.,
658   136, 105-113, https://doi.org/10.1016/j.atmosenv.2016.04.019, 2016.
659   Kuuluvainen, H., Poikkimaki, M., Jarvinen, A., Kuula, J., Irjala, M., Dal Maso, M., Keskinen, J., Timonen, H., Niemi, J. V.,
660   and Ronkko, T.: Vertical profiles of lung deposited surface area concentration of particulate matter measured with a drone in
661   a street canyon, Environ. Pollut., 241, 96-105, https://doi.org/10.1016/j.envpol.2018.04.100, 2018.
662   Lamberg, H., Nuutinen, K., Tissari, J., Ruusunen, J., Yli-Pirilä, P., Sippula, O., Tapanainen, M., Jalava, P., Makkonen, U.,
663   Teinilä, K., Saarnio, K., Hillamo, R., Hirvonen, M.-R., and Jokiniemi, J.: Physicochemical characterization of fine particles
664   from small-scale wood combustion, Atmos. Envrion., 45, 7635-7643, https://doi.org/10.1016/j.atmosenv.2011.02.072, 2011.
665   Lebouf, R. F., Stefaniak, A. B., Chen, B. T., Frazer, D. G., and Virji, M. A.: Measurement of airborne nanoparticle surface
666   area using a filter-based gas adsorption method for inhalation toxicology experiments, Nanotoxicology, 5, 687-699,
667   https://doi.org/10.3109/17435390.2010.546951, 2011.
668   Lepistö, T., Kuuluvainen, H., Juuti, P., Järvinen, A., Arffman, A., and Rönkkö, T.: Measurement of the human respiratory tract
669   deposited surface area of particles with an electrical low pressure impactor, Aerosol Sci. Tech., 54, 958-971,
670   https://doi.org/10.1080/02786826.2020.1745141, 2020.
671   Lindstrom, M. J., and Bates, D. M.: Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-
672   measures data, J. Am. Stat. Assoc., 83, 1014-1022, https://doi.org/10.2307/2290128, 1988.
673   Liu, H., Zhang, X., Zhang, H., Yao, X., Zhou, M., Wang, J., He, Z., Zhang, H., Lou, L., Mao, W., Zheng, P., and Hu, B.: Effect
674   of air pollution on the total bacteria and pathogenic bacteria in different sizes of particulate matter, Environ. Pollut., 233, 483-
675   493, https://doi.org/10.1016/j.envpol.2017.10.070, 2018a.
676   Liu, Y., Wu, J., Yu, D., and Hao, R.: Understanding the patterns and drivers of air pollution on multiple time scales: the case
677   of northern China, Environ. Manage., 61, 1048-1061, https://doi.org/10.1007/s00267-018-1026-5, 2018b.
678   Luoma, K., Niemi, J. V., Aurela, M., Fung, P. L., Helin, A., Hussein, T., Kangas, L., Kousa, A., Rönkkö, T., Timonen, H.,
679   Virkkula, A., and Petäjä, T.: Spatiotemporal variation and trends in equivalent black carbon in the Helsinki metropolitan area
680   in Finland, Atmos. Chem. Phys., 21, 1173-1189, https://doi.org/10.5194/acp-21-1173-2021, 2021.
681   Maricq, M. M.: Monitoring Motor Vehicle PM Emissions: An Evaluation of Three Portable Low-Cost Aerosol Instruments,
682   Aerosol Sci. Tech., 47, 564-573, https://doi.org/10.1080/02786826.2013.773394, 2013.
683   Miller, A.: Subset selection in regression, CRC Press, 2002.
684   NCRP: Report No. 125: Deposition, Retention and Dosimetry of Inhaled Radioactive Substances, National Council on
685   Radiation Protection and Measurements, 1997.
686   Oberdorster, G.: Nanotoxicology: in vitro-in vivo dosimetry, Environ. Health Persp., 120, A13; author reply A13,
687   https://doi.org/10.1289/ehp.1104320, 2012.
688   Oberdörster, G., Maynard, A., Donaldson, K., Castranova, V., Fitzpatrick, J., Ausman, K., Carter, J., Karn, B., Kreyling, W.,
689   Lai, D., Olin, S., Monteiro-Riviere, N., Warheit, D., Yang, H., and A report from the ILSI Research Foundation/Risk Science
690   Institute Nanomaterial Toxicity Screening Working Group: Principles for characterizing the potential human health effects
691   from exposure to nanomaterials: elements of a screening strategy, Part. Fibre Toxicol., 2, 1-35, https://doi.org/10.1186/1743-
692   8977-2-8, 2005.
693   Pacitto, A., Stabile, L., Russo, S., and Buonanno, G.: Exposure to Submicron Particles and Estimation of the Dose Received
694   by Children in School and Non-School Environments, Atmosphere (Basel), 11, https://doi.org/10.3390/atmos11050485, 2020.
695   Petäjä, T., Kerminen, V.-M., Maso, M. D., Junninen, H., Koponen, I., Hussein, T., Aalto, P. P., Andronopoulos, S., Robin, D.,
696   Hämeri, K., Bartzis, J. G., and Kulmala, M.: Sub-micron atmospheric aerosols in the surroundings of Marseille and Athens:
697   physical characterization and new particle formation, Atmos. Chem. Phys., 7, 2705-2720, https://doi.org/10.5194/acp-7-2705-
698   2007, 2007.
699   Pirjola, L., Niemi, J. V., Saarikoski, S., Aurela, M., Enroth, J., Carbone, S., Saarnio, K., Kuuluvainen, H., Kousa, A., Rönkkö,
700   T., and Hillamo, R.: Physical and chemical characterization of urban winter-time aerosols by mobile measurements in Helsinki,
701   Finland, Atmos. Environ., 158, 60-75, https://doi.org/10.1016/j.atmosenv.2017.03.028, 2017.
702   Prather, K. A., Wang, C. C., and Schooley, R. T.: Reducing transmission of SARS-CoV-2, Science, 368, 1422-1424,
703   https://doi.org/10.1126/science.abc6197, 2020.
704   Reche, C., Viana, M., Brines, M., Perez, N., Beddows, D., Alastuey, A., and Querol, X.: Determinants of aerosol lung-
705   deposited surface area variation in an urban environment, Sci. Total Environ., 517, 38-47,
706   https://doi.org/10.1016/j.scitotenv.2015.02.049, 2015.
707   Rönkkö, T., Kuuluvainen, H., Karjalainen, P., Keskinen, J., Hillamo, R., Niemi, J. V., Pirjola, L., Timonen, H. J., Saarikoski,
708   S., Saukko, E., Järvinen, A., Silvennoinen, H., Rostedt, A., Olin, M., Yli-Ojanperä, J., Nousiainen, P., Kousa, A., and Dal
709   Maso, M.: Traffic is a major source of atmospheric nanocluster aerosol, Proc. Natl. Acad. Sci. U.S.A., 114, 7549-7554,
710   https://doi.org/10.1073/pnas.1700830114, 2017.

711 Rostedt, A., Arffman, A., Janka, K., Yli-Ojanperä, J., and Keskinen, J.: Characterization and Response Model of the PPS-M
712 Aerosol Sensor, Aerosol Sci. Tech., 48, 1022-1030, https://doi.org/10.1080/02786826.2014.951023, 2014.
713 Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,
714 Nat. Mach. Intell., 1, 206-215, https://doi.org/10.1038/s42256-019-0048-x, 2019.
715 Salo, L., Hyvärinen, A., Jalava, P., Teinilä, K., Hooda, R. K., Datta, A., Saarikoski, S., Lintusaari, H., Lepistö, T., Martikainen,
716 S., Rostedt, A., Sharma, V. P., Rahman, M. H., Subudhi, S., Asmi, E., Niemi, J. V., Lihavainen, H., Lal, B., Keskinen, J.,
717 Kuuluvainen, H., Timonen, H., and Rönkkö, T.: The characteristics and size of lung-depositing particles vary significantly
718 between high and low pollution traffic environments, Atmos. Environ., 118421,
719 https://doi.org/10.1016/j.atmosenv.2021.118421, 2021a.
720 Salo, L., Rönkkö, T., Saarikoski, S., Teinilä, K., Kuula, J., Alanen, J., Arffman, A., Timonen, H., and Keskinen, J.:
721 Concentrations and Size Distributions of Particle Lung-deposited Surface Area (LDSA) in an Underground Mine, Aerosol Air
722 Qual. Res., 21, 200660-200660, https://doi.org/10.4209/aaqr.200660, 2021b.
723 Schmid, O., and Stoeger, T.: Surface area is the biologically most effective dose metric for acute nanoparticle toxicity in the
724 lung, J. Aerosol Sci., 99, 133-143, https://doi.org/10.1016/j.jaerosci.2015.12.006, 2016.
725 Shiraiwa, M., Ueda, K., Pozzer, A., Lammel, G., Kampf, C. J., Fushimi, A., Enami, S., Arangio, A. M., Fröhlich-Nowoisky,
726 J., Fujitani, Y., Furuyama, A., Lakey, P. S. J., Lelieveld, J., Lucas, K., Morino, Y., Pöschl, U., Takahama, S., Takami, A.,
727 Tong, H., Weber, B., Yoshino, A., and Sato, K.: Aerosol health effects from molecular to global scales, Environ. Sci. Technol.,
728 51, 13545-13567, https://doi.org/10.1021/acs.est.7b04417, 2017.
729 Šimić, I., Lovrić, M., Godec, R., Kröll, M., and Bešlić, I.: Applying machine learning methods to better understand, model
730 and estimate mass concentrations of traffic-related pollutants at a typical street canyon, Environ. Pollut., 263, 114587,
731 https://doi.org/10.1016/j.envpol.2020.114587, 2020.
732 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmos., 106, 7183-
733 7192, https://doi.org/10.1029/2000JD900719, 2001.
734 Teinilä, K., Aurela, M., Niemi, J. V., Kousa, A., Petäjä, T., Järvi, L., Hillamo, R., Kangas, L., Saarikoski, S., and Timonen,
735 H.: Concentration variation of gaseous and particulate pollutants in the Helsinki city centre—Observations from a two-year
736 campaign from 2013–2015, Boreal Environ. Res., 2019.
737 Timonen, H., Carbone, S., Aurela, M., Saarnio, K., Saarikoski, S., Ng, N. L., Canagaratna, M. R., Kulmala, M., Kerminen, V.-
738 M., Worsnop, D. R., and Hillamo, R.: Characteristics, sources and water-solubility of ambient submicron organic aerosol in
739 springtime in Helsinki, Finland, J. Aerosol Sci., 56, 61-77, https://doi.org/10.1016/j.jaerosci.2012.06.005, 2013.
740 Tissari, J.: Fine particle emissions from residential wood combustion (Puun pienpolton pienhiukkaspäästöt), University of
741 Kuopio, Finland, 63 pp., 2008.
742 Todea, A. M., Beckmann, S., Kaminski, H., and Asbach, C.: Accuracy of electrical aerosol sensors measuring lung deposited
743 surface area concentrations, J. Aerosol Sci., 89, 96-109, https://doi.org/10.1016/j.jaerosci.2015.07.003, 2015.
744 Tong, X., Ho, J. M. W., Li, Z., Lui, K.-H., Kwok, T. C., Tsoi, K. K., and Ho, K.: Prediction model for air particulate matter
745 levels in the households of elderly individuals in Hong Kong, Sci. Total Environ., 717, 135323,
746 https://doi.org/10.1016/j.scitotenv.2019.135323, 2020.
747 Yeh, H.-C., and Schum, G.: Models of human lung airways and their application to inhaled particle deposition, B. Math. Biol.,
748 42, 461-480, https://doi.org/10.1016/S0092-8240(80)80060-7, 1980.
749 Zaidan, M. A., Wraith, D., Boor, B. E., and Hussein, T.: Bayesian proxy modelling for estimating black carbon concentrations
750 using white-box and black-box models, Appl. Sci., 9, 4976, https://doi.org/10.3390/app9224976, 2019.
751 Zaidan, M. A., Motlagh, N. H., Fung, P. L., Lu, D., Timonen, H., Kuula, J., Niemi, J. V., Tarkoma, S., Petäjä, T., and Kulmala,
752 M. J. I. S. J.: Intelligent calibration and virtual sensing for integrated low-cost air quality sensors, 20, 13638-13652, 2020.
753 Zhou, Y., Dada, L., Liu, Y., Fu, Y., Kangasluoma, J., Chan, T., Yan, C., Chu, B., Daellenbach, K. R., Bianchi, F., Kokkonen,
754 T. V., Liu, Y., Kujansuu, J., Kerminen, V.-M., Petäjä, T., Wang, L., Jiang, J., and Kulmala, M.: Variation of size-segregated
755 particle number concentrations in wintertime Beijing, Atmos. Chem. Phys., 20, 1201-1216, https://doi.org/10.5194/acp-20-
756 1201-2020, 2020.
757

758 **Table 1.** Ambient LDSA of alveolar region (in $\mu m^2\ cm^{-3}$, corrected to 2 significant figures) reported in the last decade in chronological
759 order of the measurement start. TS and RA represent traffic sites and residential area respectively. For the other acronyms, please see the
760 method section.

Site description	Location	Average (Mean, unless state otherwise)	Uncertainties (SD, unless state otherwise)	Period/Season	Instruments	Study
UB	Ruhr, Germany	median=36	IQR=21	Mar 2009–Dec 2014	NSAM	Hennig et al. (2018)
RB+UB+TS	Basel, Geneva, Lugano,	32	IQR=25	Jan 2011–Dec 2012	DiSCmini	Eeftens et al. (2016)

	Wald, Switerland					
City centre with heavy traffic	Lisbon, Portugal	35–89	4–8	Apr–May 2011	NSAM	Albuquerque et al. (2012)
UB	Cassino, Italy	88–240	-	Oct 2011– Mar 2012	NSAM	Buonanno et al. (2012)
RB		69				
UB with traffic influence	Barcelona, Spain	37	26	Nov 2011–May 2013	NSAM	Reche et al. (2015)
TS	Helsinki, Finland	65–94	-	Feb 2012	ELPI, NSAM	Kuuluvainen et al. (2016)
RA		15–31				
TS	Athens, Greece	65	21 4.8	Jul 2012	Partector Aerotrak 9000	Cheristanidis et al. (2020)
UB with traffic influence	Leichester, UK	30	25	Nov 2013–May 2015	NSAM	Hama et al. (2017)
		23	14	Warm months		
		38	33	Cold months		
Airport	Los Angeles	47	27	Nov–Dec 2014 and May–Jul 2015	DiSCmini	Habre et al. (2018)
UB	Fukuoka, Japan	127	62	Apr 2015–Mar 2016	NSAM	Kiriya et al. (2017)
TS	Helsinki, Finland	60 (ground level)		Nov 2016	Partector, ELPI, DiSCmini, Pegasor AQ Urban	Kuuluvainen et al. (2018)
		36-40 (below rooftop)				
		16-26 (above rooftop)				
SC	Helsinki, Finland	22	14	Feb 2017–Jan 2018	Pegasor AQ Urban	Kuula et al. (2020)
UB		9.4	6.9			
DH		12	10			
TS	Delhi, India	330	130	Nov–Dec 2018	ELPI	Salo et al. (2021a)
UB	Salerno	79	48	Nov 2018– May 2019	NanoTracer	Pacitto et al. (2020)
TS	Roma, Italy	110	57			
RB	Parma, Italy	17	10			

761

762

763

764 **Table 2.** Descriptive statistics of alveolar LDSA concentrations ($\mu m^2\ cm^{-3}$) at SC (2017–2018), UB (2017–May 2018), DH1–3 (2018) and
765 RB (2018) site. The mean (column 3), standard deviation (SD, column 4), 10th, 25th, 50th, 75th and 90th percentile (P10, P25, P50, P75 and
766 P90, column 5–9), geometric mean (Gmean, column 10) and geometric standard deviation (GSD, column 11) of the concentrations are
767 corrected to one decimal place. The percentage of valid data in the reported measurement period is shown in column 12.

		Mean	SD	P10	P25	P50	P75	P90	Gmean	GSD	%
SC	All	19.7	11.3	8.4	11.7	17.0	24.7	34.4	17.0	1.7	97
	Winter	19.4	12.2	7.6	10.7	16.1	24.7	35.3	16.3	1.8	98
	Spring	19.6	11.0	8.6	11.8	16.9	24.3	34.2	17.1	1.7	94
	Summer	20.8	10.4	10.5	13.5	18.4	25.5	34.2	18.6	1.6	98
	Autumn	18.4	11.7	7.1	10.0	15.0	23.8	34.6	15.3	1.8	96
	Workdays	21.4	12.3	8.6	12.5	18.8	27.7	37.6	18.4	1.8	97
	Weekends	15.9	7.5	8.1	10.7	14.4	19.4	25.2	14.4	1.6	97
UB	All	11.2	7.1	4.6	6.4	9.5	14.0	19.6	9.5	1.8	70
	Winter	12.4	9.1	4.8	6.3	10.0	15.4	22.5	10.1	1.9	89
	Spring	10.4	6.1	4.6	6.2	9.0	12.8	18.3	9.0	1.7	100
	Summer	12.8	5.8	6.7	8.5	11.4	15.8	20.7	11.6	1.6	57
	Autumn	7.7	4.7	3.2	4.5	6.7	9.7	13.2	6.7	1.7	56
	Workdays	11.5	7.3	4.8	6.7	9.7	14.1	20.3	9.8	1.8	70
	Weekends	10.4	6.6	4.1	5.8	8.8	13.6	18.3	8.8	1.8	70
DH1–3	All	11.7	8.6	4.2	6.3	9.7	14.5	21.1	9.5	1.9	94
	Winter	12.3	10.2	4.1	6.2	9.6	14.8	23.4	9.7	2.0	86
	Spring	12.8	8.2	5.3	7.4	10.8	15.9	23.1	10.7	1.8	98
	Summer	11.8	5.9	5.7	7.8	10.8	14.5	19.2	10.6	1.6	98
	Autumn	10.5	10.2	3.0	4.6	6.8	13.0	22.2	7.5	2.2	95
	Workdays	11.8	8.3	4.3	6.4	9.9	14.6	20.8	9.6	1.9	95
	Weekends	11.7	9.3	4.0	6.0	9.4	14.3	21.8	9.3	2.0	93
RB	All	7.6	5.4	2.4	4.0	6.5	10.2	14.0	6.1	2.0	99
	Winter	6.6	6.0	2.2	3.5	5.6	8.3	11.6	5.3	1.9	100
	Spring	9.1	6.4	3.5	5.1	7.4	11.0	16.6	7.5	1.9	99
	Summer	9.8	4.3	4.7	6.6	9.3	12.5	15.3	8.9	1.6	99
	Autumn	4.9	4.1	1.6	2.6	3.9	5.6	8.9	3.8	2.0	99
	Workdays	7.7	5.6	2.5	4.1	6.6	10.2	14.1	6.2	2.0	99
	Weekends	7.6	5.0	2.4	4.0	6.5	10.1	14.0	6.1	2.0	100

768

769

770

771 **Table 3.** The evaluation attributes by IAME model at the SC and the UB site, corrected to 2 significant figures.

	Street canyon					Urban background				
	R^2	MAE	$cRMSD$	r	NSD	R^2	MAE	$cRMSD$	r	NSD
All	0.80	3.7	5.6	0.87	0.78	0.77	2.3	3.7	0.86	0.80
Winter	0.86	3.4	5.3	0.92	0.74	0.81	2.5	4.6	0.89	0.68
Spring	0.75	3.9	5.9	0.85	0.79	0.61	2.4	3.3	0.84	0.85
Summer	0.70	4.1	5.9	0.83	0.84	0.61	2.7	3.7	0.79	0.95
Autumn	0.85	3.4	5.4	0.9	0.75	0.85	1.3	2.0	0.91	0.83
Workdays	0.81	4.1	6.1	0.87	0.77	0.75	2.4	3.8	0.86	0.77
Weekends	0.72	3.0	4.3	0.82	0.82	0.8	2.1	3.5	0.85	0.87

772

773

774 **Table 4.** Five most successful sub-models at the SC site. The table shows only the fixed predictors with their coefficient (β, all p<0.05) and
775 corresponding standard error (SE). The variance inflation factor (VIF) among the fixed predictors is also shown for the 5 sub-models. The
776 evaluation attributes of the sub-models are shown column 6–10. The percentage of the sub-model usage and the number of data points (n)
777 is shown in column 11 and 12. Natural logarithm is taken for parameters with asterisk (*).

	Fixed predictors	β	SE	VIF	R^2	MAE	cRMSD	r	NSD	%	n
	*PM2.5	0.119	0.005	1.54							
1	*PNC	0.313	0.005	2.89	0.74	3.7	5.7	0.87	0.79	81	2603
	*BC	0.223	0.004	2.17							
	*NOx	0.236	0.005	3.79							
2	*PNC	0.153	0.005	1.63	0.74	3.8	5.7	0.86	0.77	13	2629
	*BC	0.231	0.007	4.90							
	*PNC	−0.044	0.003	1.07							
3	*BC	0.375	0.004	2.20	0.74	3.8	5.8	0.86	0.78	4	6622
	WS	0.201	0.004	2.15							
	*NO_x	0.250	0.005	3.09							
4	*$PM_{2.5}$	0.243	0.004	1.17	0.74	3.8	5.7	0.87	0.78	<1	2596
	*PNC	0.184	0.005	3.02							
	*NOx	0.176	0.005	3.51							
5	*PM_{10}	0.070	0.004	1.3	0.72	3.8	5.8	0.85	0.75	<1	2713
	*BC	0.326	0.006	3.65							

778

779

780 **Table 5.** Five most successful sub-models at the UB site. The table shows only the fixed predictors with their coefficient (β, all $p<0.05$) and
781 corresponding standard error (SE). The variance inflation factor (VIF) among the fixed predictors is also shown for the 5 sub-models. The
782 evaluation attributes of the sub-models are shown column 6–10, corrected to 2 significant figures. The percentage of the sub-model usage
783 and the number of data points (n) is shown in column 11 and 12. Natural logarithm is taken for parameters with asterisk (*).

	Fixed predictors	β	SE	VIF	R^2	MAE	cRMSD	r	NSD	%	n
	*CO	0.072	0.027	1.72							
1	*PNC	0.400	0.006	2.08	0.84	1.7	2.5	0.92	0.87	49	941
	*BC	2.956	0.007	1.52							
	*PNC	–0.098	0.005	1.09							
2	*BC	0.398	0.004	1.44	0.82	1.9	2.9	0.91	0.89	17	6608
	WD-N	0.328	0.006	1.55							
	*NO$_2$	0.237	0.007	1.88							
3	*CO	0.520	0.024	1.10	0.69	2.4	3.4	0.84	0.73	17	941
	*PNC	0.341	0.010	2.00							
	*CO	0.009	0.000	1.08							
4	*PNC	0.348	0.025	1.07	0.58	2.7	3.9	0.77	0.63	11	9757
	RH	0.590	0.007	1.15							
	*NO$_x$	0.107	0.006	2.22							
5	*CO	0.182	0.032	1.72	0.81	1.9	3.0	0.90	0.85	2	7036
	*BC	0.455	0.007	2.56							

784
785
786
787

788 **Table 6.** Model evaluation comparison of deposition model derived LDSA (LDSA$_{ICRP}$), modelled LDSA by IAP (LDSA$_{IAP}$) and modelled
789 LDSA by IAME (LDSA$_{IAME}$) against reference measurements LDSA$_{Pegasor}$ at the SC and the UB site. Parameters with an asterisk represent
790 natural logarithm. The evaluation attributes of the three methods are corrected to 2 significant figures.

	Street canyon					Urban background				
	R^2	MAE	cRMSD	r	NSD	R^2	MAE	cRMSD	r	NSD
LDSA$_{ICRP}$	0.72	4.1	6.2	0.88	1.1	0.83	1.8	2.9	0.93	1.1
LDSA$_{IAP}$	0.77	4.0	6.0	0.85	0.78	0.66	2.8	3.9	0.84	0.81
LDSA$_{IAME}$	0.80	3.7	5.6	0.87	0.78	0.77	2.3	3.7	0.86	0.80

791

792

Atmospheric
Chemistry
and Physics
Discussions

793 **Table 7.** Statistics to show temporal difference. The number of data (n), mean and standard deviation (SD) of absolute error and the
794 corresponding *p*-values of t-tests at the selected time windows at both sites.

Street canyon (SC)	n	Mean	SD	t-test	*p*
Workdays	11658	4.1	4.8	Workdays vs Weekends	4.13×10^{-81}
Weekends	5322	3.0	3.2		
				Winter vs Spring	3.64×10^{-24}
Winter	4023	3.4	4.2	Winter vs Summer	5.89×10^{-5}
Spring	2297	4.0	4.5	Winter vs Autumn	7.07×10^{-7}
Summer	6457	4.2	4.4	Spring vs Summer	6.38×10^{-34}
Autumn	4320	3.4	4.3	Spring vs Autumn	1.02×10^{-4}
				Summer vs Autumn	2.69×10^{-15}
Hour 4–10 a.m.	4953	4.8	5.6	Hour 4–10 a.m. vs	2.58×10^{-40}
Hour 4–10 p.m.	4981	3.5	3.6	4–10 p.m.	

795

Urban background (UB)	n	Mean	SD	t-test	*p*
Workdays	8473	2.3	2.6	Workdays vs Weekends	5.08×10^{-8}
Weekends	3852	2.1	2.6		
				Winter vs Spring	1.96×10^{-7}
Winter	2539	2.5	3.2	Winter vs Summer	0.39***
Spring	1101	1.9	3.1	Winter vs Autumn	1.90×10^{-2}
Summer	1628	2.6	2.4	Spring vs Summer	2.75×10^{-9}
Autumn	812	2.3	2.1	Spring vs Autumn	2.20×10^{-3}
				Summer vs Autumn	1.40×10^{-3}
Hour 4–10 a.m.	3620	2.3	2.7	Hour 4–10 a.m. vs	0.86***
Hour 4–10 p.m.	3591	2.3	2.7	4–10 p.m.	

796

	n	Mean	SD	t-test	*p*
Street canyon (SC)	11940	3.9	4.6	SC vs UB	8.21×10^{-246}
Urban background (UB)		2.3	2.6	(in same time period)	

797 *** *p*>0.05 the null hypothesis of different distribution is rejected

798

799
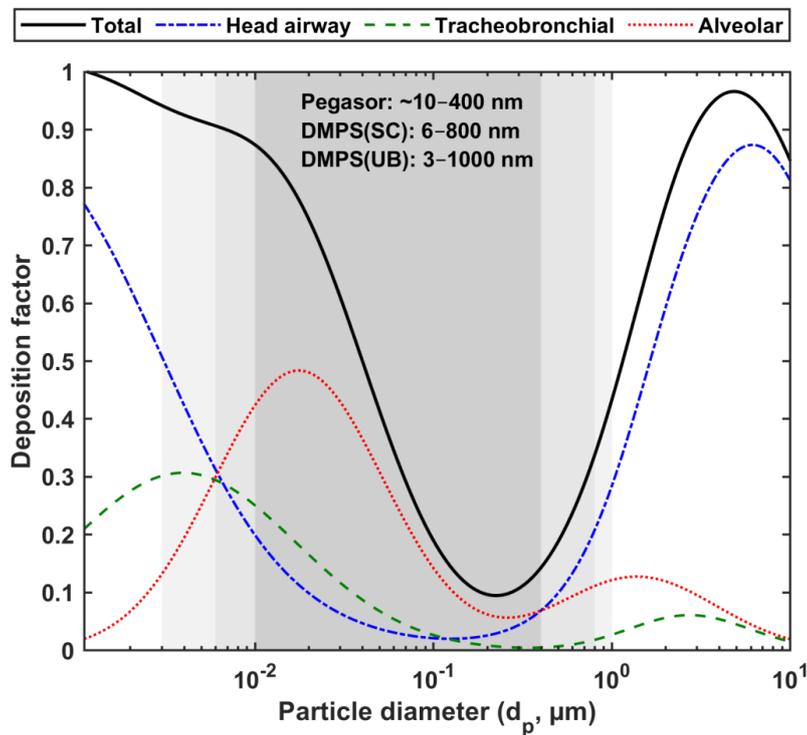
Figure 1. Lung deposition factor of a spectrum of particle size distribution based on the equation (ICRP, 1994). Black solid line represents the total deposition factor while blue, green and red dotted line refer to deposition factor in head airway, tracheobraonchial and alveolar region, respectively. Pegasor AQ Urban measured the alveolar LDSA concentration of particles in the ~10–400 nm size range (dark grey). DMPS at SC and UB were used to calculate alveolar LDSA in selected size fractions in the 6–800 nm and 3–1000 nm size range, respectively.

800
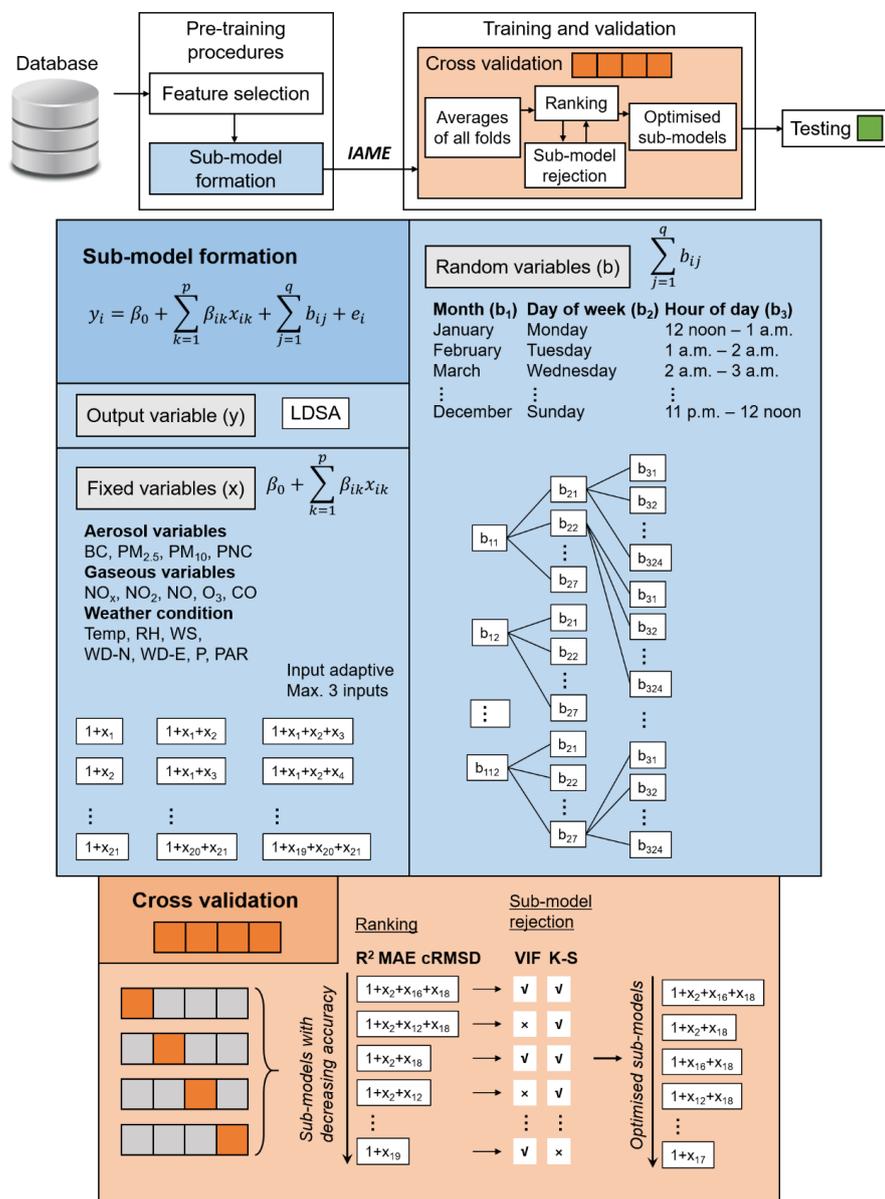
Figure 2. The block diagram of the proxy procedures (top). The blue and orange blocks are explanatory notes to the sections of sub-model formation and cross validation, respectively.

801
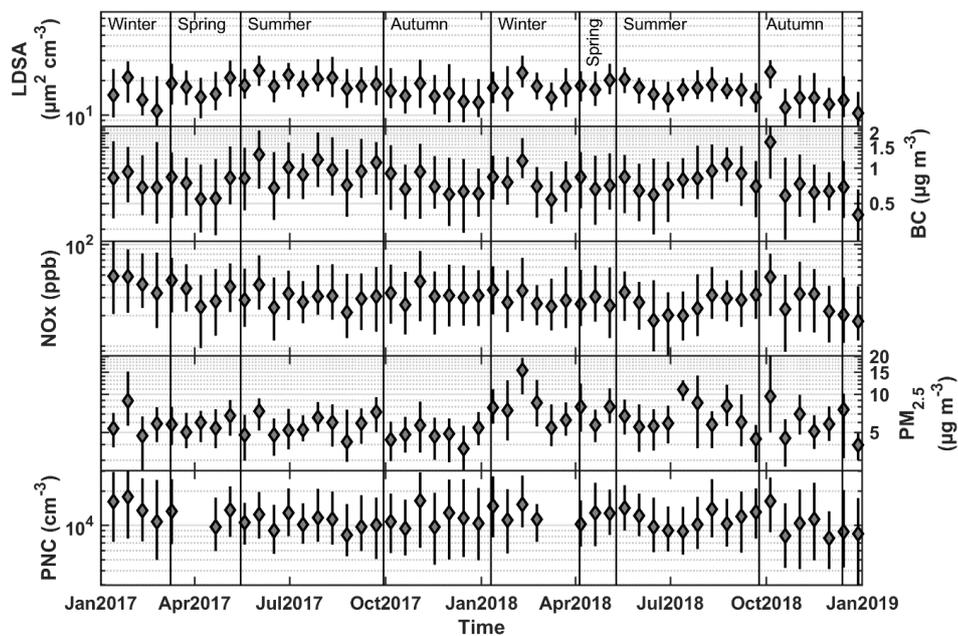
Figure 3. Time series of the selected air pollutant parameters (First to end row: LDSA (μm² cm⁻³), BC (μg m⁻³), NOₓ (ppb), PM₂.₅ (μg m⁻³) and PNC (cm⁻³)) at Mäkelänkatu SC site during the measurement period from 1 January 2017 and 31 December 2018. Each bar represents a period of two weeks where the shaded diamond marker is the median and the vertical error bars are the 25th and 75th percentiles. Seasons are thermally separated.

802

803

804
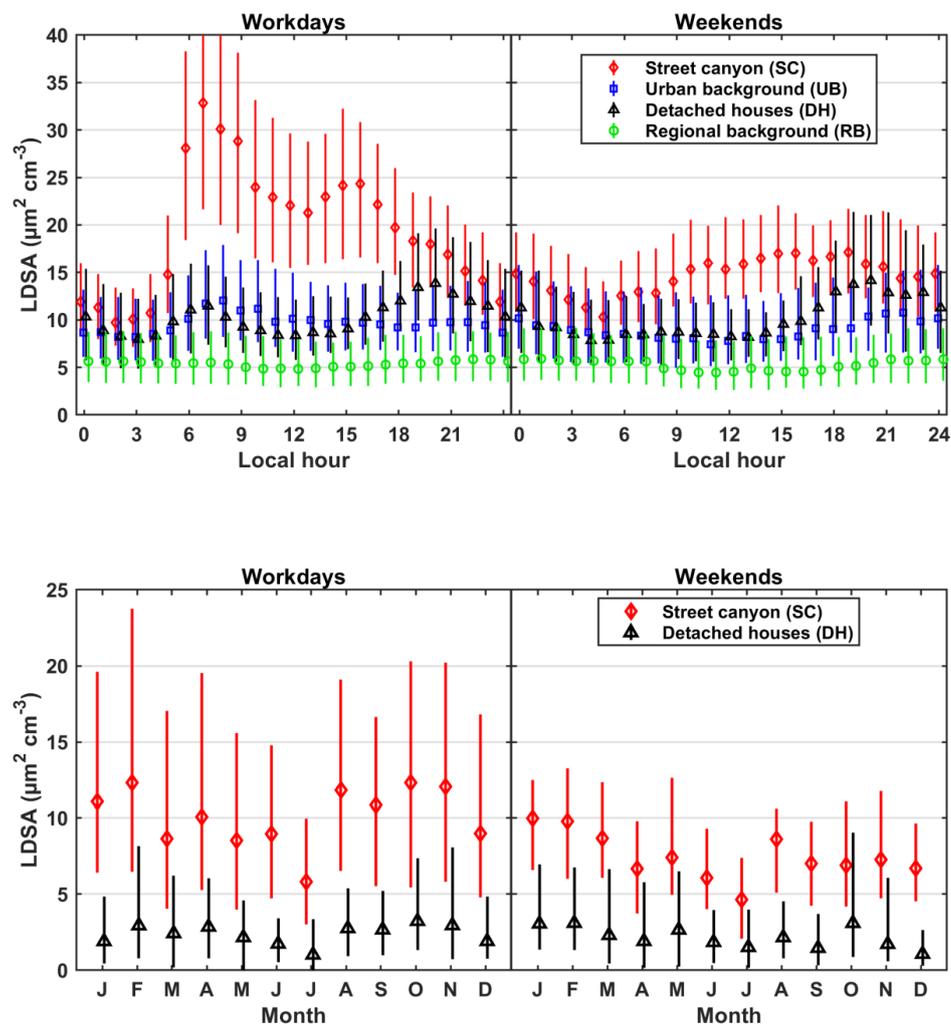
Figure 4. Upper panel: Diurnal cycles of LDSA concentrations (μm^2 cm^{-3}) at SC (red diamond, 2017–2018), UB (blue square, 2017–May 2018), DH1–3 (black triangle, 2018) and RB site (green circle, 2018) on workdays and weekends with error bars of 25th and 75th percentiles. Lower panel: Monthly averages in year 2018 of local LDSA increments at the SC (red diamond) and DH1–3 (black triangle) site (LDSA concentration at the hotspot site – LDSA at RB site) on workdays and weekends with error bars of 25th and 75th percentiles.

805

806
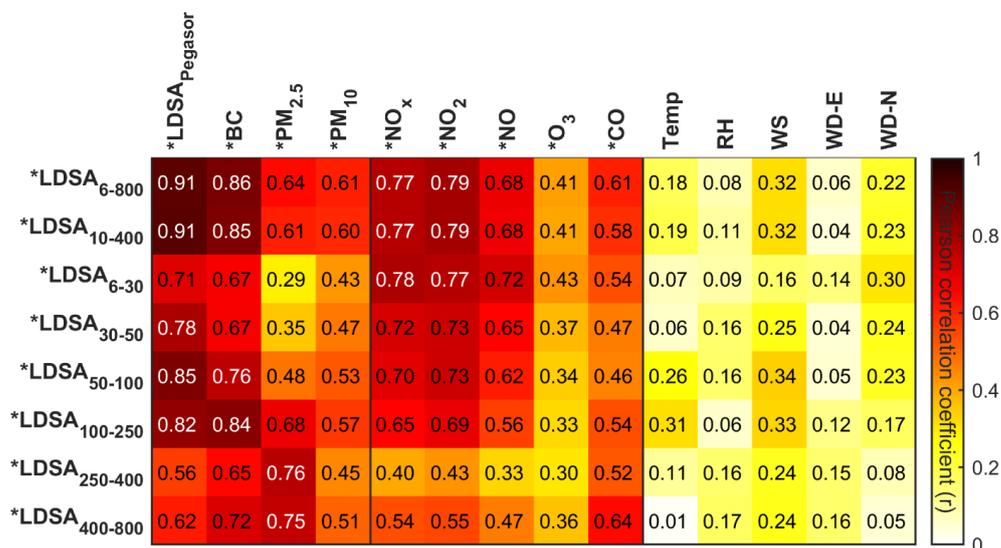
Figure 5. Heatmap showing Pearson correlation coefficient (r, corrected to 2 significant figures) of LDSA of different particle size sections (in nm) by ICRP lung deposition model and the other air pollutant parameters at Mäkelänkatu SC site. Dark red indicates a high correlation while pale yellow indicates a low correlation. Parameters with an asterisk represent natural logarithm. LDSA_Pegasor represents the measured LDSA concentrations.

807
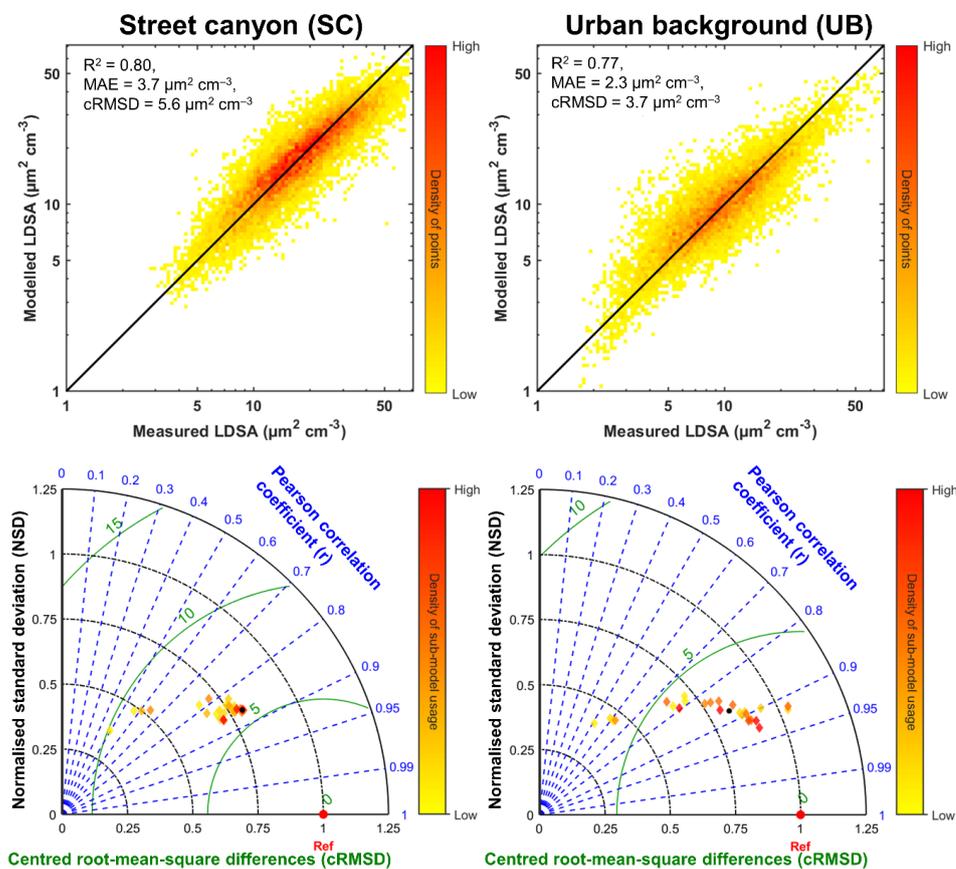
Figure 6. The upper panel shows the Taylor's diagrams (Taylor, 2001) at Mäkelänkatu SC site (first column) and at Kumpula UB site (second column). Each diamond marker in the Taylor's diagrams represents each sub-model used in the final estimation by IAME (solid black dot), compared with the reference data (solid red dot). Hues of colours represent how frequent the sub-model was used. The lower panel shows the scatter plots of modelled LDSA against the measured LDSA at Mäkelänkatu SC site (first column) and at Kumpula UB site (second column). Hues of colours represent the density of points on the figure.
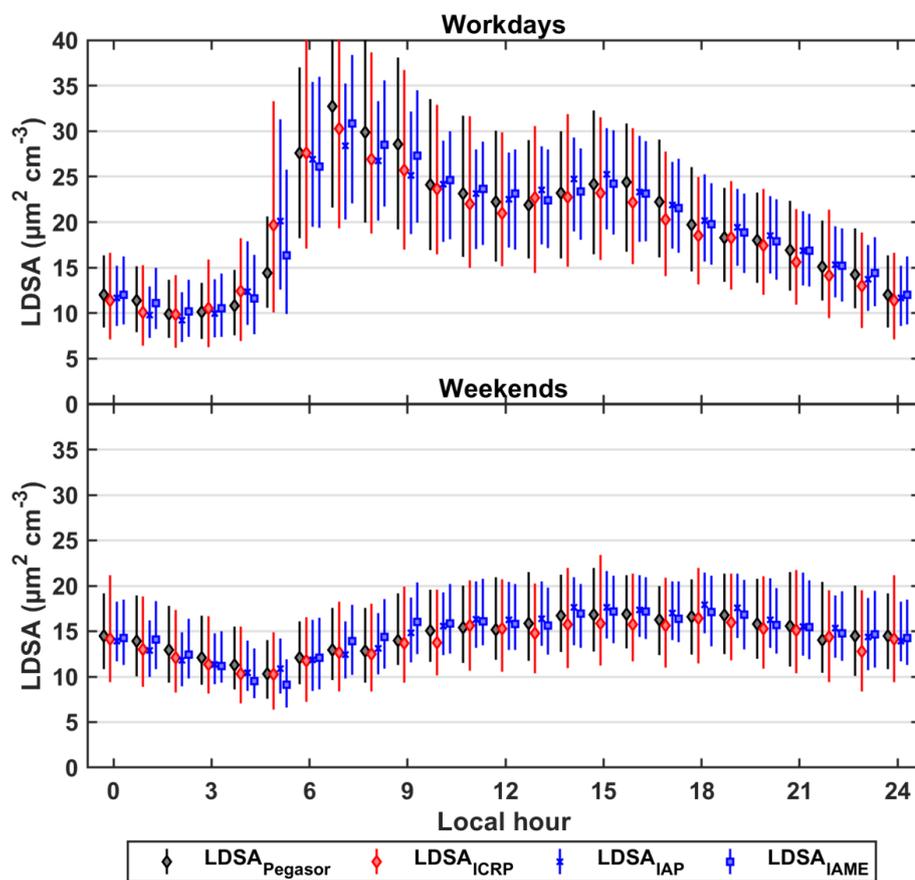
808

809
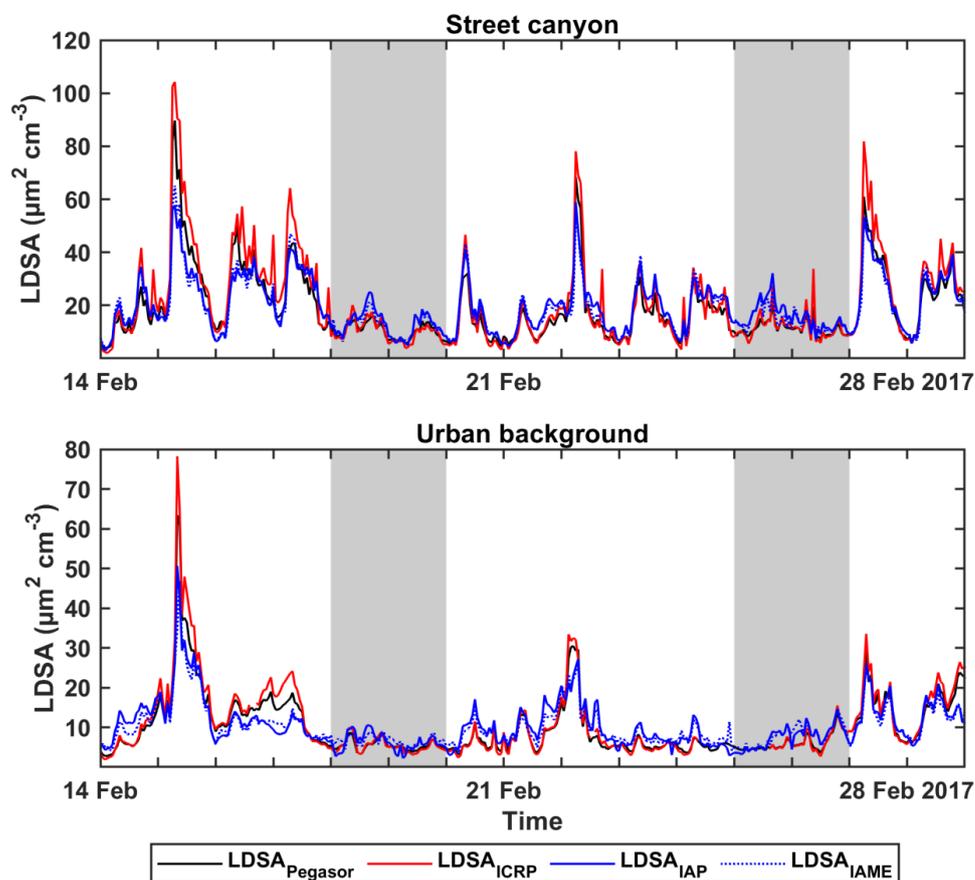
Figure 7. Timeseries of measured LDSA (LDSA$_{Pegasor}$, black), deposition model derived LDSA by ICRP (LDSA$_{ICRP}$, red), modelled LDSA by IAP (LDSA$_{IAP}$, blue solid line) and modelled LDSA by IAME (LDSA$_{IAME}$, blue dotted line) during a selected measurement window between 14 and 28 February 2017. Shaded regions represent weekends, otherwise workdays.

810

811

Figure 8. Diurnal cycles of measured (LDSA$_{Pegasor}$, black), deposition model derived (LDSA$_{ICRP}$, red) and modelled (LDSA$_{IAP}$ and LDSA$_{IAME}$, blue) LDSA concentrations with error bars of 25th and 75th percentiles on workdays (left) and weekends (right). LDSA$_{IAP}$ and LDSA$_{IAME}$ can be differentiated by their markers, cross for the former and square for the latter.

812

813