Environmental Pollution 199 (2015) 209–218

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

A comparison of strategies for estimation of ultrafine particle number concentrations in urban air pollution monitoring networks

Matteo Reggente [a, *], Jan Peters [a], Jan Theunis [a], Martine Van Poppel [a], Michael Rademaker [b], Bernard De Baets [b], Prashant Kumar [c, d]

[a] VITO, Flemish Institute for Technological Research, Boeretang 200, B-2400 Mol, Belgium
[b] Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, 9000 Gent, Belgium
[c] Department of Civil and Environmental Engineering, Faculty of Engineering and Physical Science (FEPS), University of Surrey, GU2 7XH, United Kingdom
[d] Environmental Flow (EnFlo) Research Centre, FEPS, University of Surrey, GU2 7XH, United Kingdom

ARTICLE INFO

Article history:
Received 11 November 2014
Received in revised form
24 January 2015
Accepted 27 January 2015
Available online

Keywords:
Ultrafine particles estimation
Urban air pollution
Pollution monitoring network
Gaussian process regression
Statistical modelling

ABSTRACT

We propose three estimation strategies (local, remote and mixed) for ultrafine particles (UFP) at three sites in an urban air pollution monitoring network. Estimates are obtained through Gaussian process regression based on concentrations of gaseous pollutants (NOx, O_3, CO) and UFP. As local strategy, we use local measurements of gaseous pollutants (local covariates) to estimate UFP at the same site. As remote strategy, we use measurements of gaseous pollutants and UFP from two independent sites (remote covariates) to estimate UFP at a third site. As mixed strategy, we use local and remote covariates to estimate UFP. The results suggest: UFP can be estimated with good accuracy based on NOx measurements at the same location; it is possible to estimate UFP at one location based on measurements of NOx or UFP at two remote locations; the addition of remote UFP to local NOx, O_3 or CO measurements improves models' performance.

1. Introduction

Exposure to traffic-related pollution, especially UFP and nitrogen oxides (NOx), is of great concern in urban environments because of their adverse impact on human health (Hong et al., 2002; de Hartog et al., 2003; Atkinson et al., 2010; Jacobs et al., 2010; Bos et al., 2011; Kumar et al., 2011a, 2014).

UFP are commonly defined as particles having a diameter of less than 100 nm (Morawska et al., 1998), and the consensus is that these particles contribute most (around 80%) of the total particle number concentration (PNC) (Heal et al., 2012; Kumar et al., 2011b; Morawska et al., 2008; Charron and Harrison, 2003), whereas their corresponding mass accounts for less than 20% of the total particle mass concentration (Kittelson, 1998). UFP can be classified into the “nucleation”, “Aitken” and “accumulation” modes. In terms of size ranges, the nucleation, Aitken and accumulation modes typically encompass 1–30, 20–100 and 30–300 nm, respectively. Particles with a diameter below 30 nm contain nearly 30% of total PNC (Morawska et al., 2008; Kumar et al., 2010).

Road vehicle emissions in polluted urban environments can contribute up to 90% of the total PNC (Kumar et al., 2010; Pey et al., 2009). The UFP along the roadside show an association with the vehicle flow characteristics. For instance, increasing vehicle speed increases the emissions of UFP (Kittelson et al., 2004). Among the road vehicles, diesel engines dominate road traffic emission of UFP, and heavy duty vehicles have an average factor of magnitude of two with respect to the light duty engine (Beddows and Harrison, 2008).

UFP vary spatially between the sources and the receptors living or travelling close to the roads (Kumar et al., 2014). This variation depends on many factors such as source type and strength, meteorological and dilution conditions, location geometry and transformation processes, among others (Heal et al., 2012; Goel and Kumar, 2014).

Currently there is no limit value to control ambient UFP. Consequently, there are not many UFP monitors deployed as part of the governmental monitoring stations. On the other hand, NOx, ozone (O_3) and carbon monoxide (CO) are regulated pollutants (Directive 2008/50/EC) and their monitors are spread all over

* Corresponding author. Present address: Atmospheric Particle Research Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.
E-mail address: reggente@gmail.com (M. Reggente).

http://dx.doi.org/10.1016/j.envpol.2015.01.034
0269-7491/© 2015 Published by Elsevier Ltd.

Europe. Nitric oxide (NO) and nitrogen dioxide (NO_2) together make NOx. Emissions of NOx are associated with all types of high-temperature combustion, but similar to UFP, their most important sources in urban areas remain road vehicles (Westmoreland et al., 2007; Alvarez et al., 2008; Kumar and Imam, 2013).

The dispersion modelling of pollutants mostly fits into two categories: deterministic and statistical. Deterministic dispersion models provide a link between theory and measurements and account for source dynamics and physico-chemical processes explicitly (Holmes and Morawska, 2006). A drawback of these models is that they need detailed information (e.g. boundary conditions), which is not always available. Statistical models do not describe the actual physical processes, but they treat the input data as random variables to derive a statistical description of the target distribution using a set of measurements. A few studies have used a statistical approach in the past (Hussein et al., 2006; Clifford et al., 2011; Mølgaard et al., 2012; Sabaliauskas et al., 2012; Reggente et al., 2014).

We employ a statistical modelling approach – Gaussian process (GP) regression – to estimate UFP in an urban air pollution monitoring network based on local and remote concentrations of NOx, O_3, CO and UFP.

2. Materials and methods

2.1. Instrumentation

We recorded UFP and gaseous pollutants for one month at a sampling frequency of 5 min and then averaged on a half-hourly basis.

Measurements of UFP were obtained using the GRIMM Nano-Check model 1.320. The Nano-Check can count total PNCs between 25 and 300 nm, and provides the mean diameter of the measured size range.

Chemiluminescence (EN 14211), ultraviolet photometry (EN 14625) and non-dispersive infrared (EN14626) analysers (Airpointer) were used to measure NOx, CO and O_3, respectively. The lowest detectable concentration was 1 $\mu g\ m^{-3}$ for NOx and O_3, and 50 $\mu g\ m^{-3}$ for CO.

Vehicle counts were recorded in four categories (cars, vans, small and big trucks/buses) using double inductive loop detectors at sites 1 and 3; video counting was performed to obtain traffic data at site 2 (Table 1).

2.2. Description of the sampling locations

Measurements were carried out in the Borgerhout district (51° 13′ N and 4° 26′ E) of Antwerp, Belgium. Borgerhout is a typical urban commercial and residential area with busy traffic. Measurements were carried out simultaneously for one month (12/02/2010–12/03/2010) at three different sites (Fig. 1). Sites 1 and 2 were located in two street canyons with two traffic lanes and moderate levels of traffic. The monitoring devices were deployed in parking lots (few metres far from the traffic). Site 3 was located in a parking area ~30 m far from a major access road with busy traffic intersections and four lanes (two in each direction) and ~200 m far from a highway.

Table 1
Description of the measurement sites.

	Distance from traffic (m)	Weekday traffic volume (veh/day)	Weekend traffic volume (veh/day)	Heavy duty vehicle on weekday (weekend) (%total)
Site 1	~3	5000	4000	5% (2%)
Site 2	~2	4000	3000	4% (2%)
Site 3	~20–30	37,000	25,000	7% (3%)

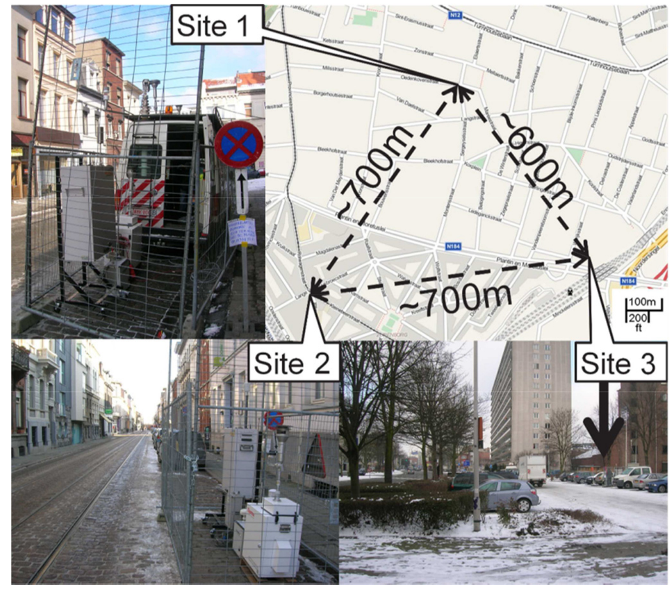

Fig. 1. Map of the measurement sites (Antwerp, Belgium) and their distances from each other. The images show the deployed instrumentation at each site. The black arrow in the image of site 3 shows the location of the deployed monitors.

2.3. Description of the model

2.3.1. Gaussian process regression

We treat the estimation problem as a non-parametric regression problem, and solve it using Gaussian process (GP) regression.

Definition: *A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution* (Rasmussen and Williams, 2006). We want to learn, from a set of measurements (D) a function $f(\cdot)$ of the relationship existing between the set of covariates $\mathbf{x}$ (NOx, CO, O_3, UFP) and the target variable, UFP ($\boldsymbol{y}$), assuming that the observed data $\boldsymbol{y}$ is generated with Gaussian noise around the underlying function f

$$\boldsymbol{y} = f(\mathbf{x}) + \varepsilon \tag{1}$$

Because of the nature of the dataset used, we do not assume an independent noise, and the dependencies are modelled adding a noise term to the covariance function (k_{Noise}). This method has been suggested by Rasmussen and Williams (2006) and by Murray-Smith and Girard (2001).

Prior beliefs about the properties of the latent function are included in the mean $m(\boldsymbol{x})$ and covariance $k(\mathbf{x},\mathbf{x}')$ functions. In order to estimate UFP based on data, we consider the joint Gaussian prior of the training observations $\boldsymbol{y}$ and the test outputs $\boldsymbol{f}_*$. The posterior distribution is obtained by conditioning the prior on the observed training outputs, such that the conditional distribution of $\boldsymbol{f}_*$ only contains those functions from the prior that are consistent with the training data

$$p(\boldsymbol{f}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \tag{2}$$

where