Building and Environment 127 (2018) 138–147

Contents lists available at ScienceDirect

Building and Environment

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

# Spatiotemporal distribution of indoor particulate matter concentration with a low-cost sensor network

Jiayu Li[a], Haoran Li[b], Yehan Ma[b], Yang Wang[a], Ahmed A. Abokifa[a], Chenyang Lu[b], Pratim Biswas[a,*]

[a] Aerosol and Air Quality Research Laboratory, Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA

[b] Cyber-Physical Systems Laboratory, Department of Computer Science & Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA

ARTICLE INFO

*Keywords:*
Low-cost sensor
Wireless
Spatial temporal distribution
Kriging
Machine learning

ABSTRACT

Real-time measurement of particulate matter (PM) is important for the maintenance of acceptable air quality. The high cost of conventional instruments typically limits the number of monitoring sites, which in turn undermines the accuracy of real-time mapping of sources and hotspots of air pollutants with sufficient spatial resolution. In this study, a wireless network of low-cost particle sensors that can be deployed indoors was developed. To overcome the well-known limitations of low sensitivity and poor signal quality associated with low-cost sensors, a sliding window and a low pass filter were developed to enhance the signal quality. Utility of the networked system with improved sensitivity was demonstrated by deploying it in a woodworking shop. Data collected by the networked system was utilized to construct spatiotemporal PM concentration distributions using an ordinary Kriging method and an Artificial Neural Network model to elucidate particle generation and ventilation processes.

## 1. Introduction

Particulate matter (PM) is a routinely monitored air pollutant in outdoor and indoor environments [1–3]. High $PM_{2.5}$ exposure levels tend to trigger cardiovascular disease and mortality via various mechanisms including pulmonary and systemic inflammation, accelerated atherosclerosis, and altered cardiac autonomic function [4,5]. Worldwide, outdoor $PM_{2.5}$ pollution accounts for 6.4 million deaths annually [6]. Indoor PM, carrying allergens and endotoxins, may exacerbate asthmatic symptoms [7]. Due to these adverse health effects, many countries have enacted regulations in an effort to lower PM concentrations, and regulatory agencies commonly require long-term measurements to monitor air quality [8,9]. The designated US Environmental Protection Agency (US EPA) federal reference method (FRM), gravimetric sampling, measures PM mass concentration by collecting the particles on a filter for a set time in a high-volume air sampler [10–12]. There are around hundreds of monitoring sites across the country that provide the daily concentrations of total suspended particles (TSP), $PM_{10}$, and $PM_{2.5}$. Using these data to generate a spatiotemporal distribution map showing how the pollutants vary with location and time aids exposure assessment and health effect studies. To generate the spatiotemporal distribution on the basis of limited data from scattered monitoring sites, researchers need to predict the pollutant concentration at unsampled locations.

Geostatistical interpolation and land use regression (LUR) are common methods to predict the spatiotemporal distribution in outdoor atmospheric studies. Geostatistical interpolation (also called spatial interpolation) characterizes the relationship between pollutant concentrations and their locations, and utilizes the relationship to predict the pointwise pollution concentration. There are four general *weighted average* algorithms for geostatistical interpolation: spatial averaging, nearest neighbor, inverse distance weighting, and Kriging [13]. Among the four algorithms, since Kriging produces the best linear unbiased estimate of the pollution surface [14], it has become the most widely used algorithm for predicting air pollution distribution [15]. Using Kriging, Jerrett et al. [16] interpolated the $PM_{2.5}$ concentrations from 42 monitoring sites and demonstrated that these concentrations are relevant to ischemic heart disease. Kriged ozone concentrations have been used for monthly exposure assessment in the southeastern United States [17], and have been applied to correlate exposure with pediatric asthma presentation rates [18]. LUR, the other predictive method, associates pollution data with multiple variables, including the wind field, traffic count, land use, population, and emissions [19]. LUR has been used to predict the PM concentration distributions across New York City

* Corresponding author.
*E-mail address:* pbiswas@wustl.edu (P. Biswas).

https://doi.org/10.1016/j.buildenv.2017.11.001
Received 4 September 2017; Received in revised form 1 November 2017; Accepted 2 November 2017
Available online 04 November 2017

and Los Angeles [20,21]. Neither of these predictive methods consistently outperformed the other. The distributions they predicted may vary according to their principles [13,22]. Furthermore, the scattered monitoring sites limit the resolution and the accuracy of the spatio-temporal distribution map, which will further undermine the confidence of the spatiotemporal distribution.

Recently, advances in the low-cost particle sensor techniques have altered the conventional data collecting and data mining processes. Conventional gravimetric sampling is off-line and laborious, whereas low-cost particle sensors offer adequate accuracy, are compact, and require only modest maintenance. The networking capability of particle sensors enhances the possibility of wider application. Laboratory evaluations have demonstrated that low-cost particle sensors operate with high linear correlation to standard commercial instruments for fixed PM sources [23–25]. In combination with other sensors and wireless communication chips, low-cost particle sensors can be networked to collect air quality data efficiently and conveniently. The data mining process of the networked low-cost particle sensor is a trending topic. By distributing 8000 low-cost iSPEX (a smart phone add-on) sensors across Netherlands, Snik et al. [26] obtained a map of aerosol optical thickness with higher spatial resolution than conventional maps generated by the satellite. In a similar study, Shinyei PPD sensors were deployed in Xi'an, China, to determine the spatiotemporal variations of $PM_{2.5}$ [27]. However, one shortcoming of low-cost particle sensors is their low signal to noise ratio, which allows accurate measurements only under higher concentration scenarios or after long periods of averaging to increase data quality. To eliminate high-frequency noise and to accurately represent the measurement data, digital filters were added to the sensor system. Common digital filters include sliding window filter, low-pass filter (e.g. finite impulse response (FIR) filter and fast Fourier transform (FFT) filter), and model-based filter (e.g. Kalman filter) [28–30]. The advantages of sliding window filter and low-pass filter are model-independent, light computational weight, and specifically tailored for filtering high frequency noises [28,31].

In addition to the conventional geostatistical models, such as ordinary kriging, machine-learning techniques were also previously used for the spatial interpolation of environmental variables as a cost-effective method where monitoring resources are limited [32,33]. For example, Chowdhury et al. [34] implemented Artificial Neural Networks (ANNs) for the spatial mapping of complex patterns of groundwater arsenic levels based on sampling data at finite locations. They demonstrated that the use of non-linear pattern learning techniques, such as ANNs, could yield more accurate results than the ordinary Kriging method. Antonic et al. [35] used neural networks to build empirical spatio-temporal models for various climatic variables such as temperature, relative humidity, precipitation, solar irradiation, and evapotranspiration. In addition, ANN models were used to forecast outdoor particulate matter concentrations such as $PM_{10}$ and $PM_{2.5}$ [36,37].

Apart from atmospheric measurements, low-cost particle sensors can perform multi-point indoor measurements. Indoor air quality, referring to PM concentrations and trace gas concentrations, is critical to human health, since a human being spends on average approximately 88% of their time inside buildings [38–43]. The application of low-cost sensors and their networks enables sampling PM and trace gas under various scenarios [44–47]. Generally, the exposure level estimated from indoor or personal low-cost sensors is more accurate than the Kriging or LUR predicted values from scattered fixed monitoring sites. Compared to outdoor field measurements, indoor measurements are usually limited by confined space and room arrangements. It is common to neglect the indoor spatiotemporal distribution and use a single-point measurement to represent the whole room, which introduces errors to exposure intake estimation [48]. The sensors' low price and the compact size allow deploying multiple sampling points in households, which is very helpful for understanding ventilation process and monitoring occupancy [49–51]. However, very few studies using networked sensor systems reported the dynamic evolution of the particle concentrations as a function of location and time. Rajasegarar et al. [52] conducted one of such studies that reported the PM concentration distribution mapped by networked low-cost particle sensors in a garage. Patel et al. [49] deployed low-cost particle sensors in a household to monitor the transport of particles produced from biomass burning. Leavey et al. [53] implemented several wireless PM sensors, gas sensors, and temperature sensors in an auditorium room and analyzed the energy consumption of different operation modes. However, neither of these studies reported the spatial evolution of the PM concentration.

The focus of indoor air quality mapping should be different from that of outdoor atmospheric studies. The scale of an atmospheric study is obviously large, possibly also ranging from county to country in scale, while indoor measurements are confined to several hundreds or thousands of square feet. Due to these space limitations, indoor measurements usually involve fewer than ten sampling locations, but the density (sampling locations/unit area) is high. In addition, there are no boundary conditions for atmospheric measurements, but the PM concentrations at the boundaries of a room should be zero, since it is a confined space and particles are scavenged at the wall. Additional variables (e.g., traffic and land use) that can be incorporated in atmospheric measurements are inapplicable for indoor sampling. Furthermore, in Kriging and the LUR method, the PM concentration distributions are considered steady and stable, hence yearly-average concentrations are commonly used as inputs. For indoor measurements, capturing instant emission events is of major interest. In general, spatially depicting the dynamic evolution of the PM concentration with a limited number of sensors in a confined space is the goal of deploying low-cost sensors for indoor measurements.

In this study, a networked wireless particle sensor system coupled with a sliding window filter and a low pass filter to enhance the sample quality by increasing the signal to noise ratio, while preserving the time resolution is presented. After calibrating the networked wireless sensor system, its use is demonstrated by conducting spatiotemporal measurements in a student woodworking shop to identify PM concentration hotspots. Kriging interpolation and artificial neural network (ANN) methods are used, and the pros and cons of each are compared. The total exposure to PM of woodworkers is estimated from calculations based on the predicted spatiotemporal PM concentration distribution.

## 2. Materials and methods

The networked wireless sensor system consists of multiple end devices to monitor the PM concentration, and a base station that receives the data from the sensors for further processing. For each end device, a Sharp GP2Y1010AU0F (GP2Y, Sharp Corp., Osaka, Japan) PM sensor, an Arduino Nano ATmega328 (Arduino, Arduino Inc., S.R.L, Italy), and an XBee radio (Digi International Inc., Minnetonka, MN) were mounted on a printed circuit board. The base station that collects and translates the data package sent from the end devices integrates a Raspberry Pi 2 embedded computer (Adafruit Industries., New York City, NY) and an XBee radio. The system and the major components are shown in Fig. 1.

### 2.1. Major components

The Sharp GP2Y measures a scattered light signal that is correlated with aerosol concentration. When the infrared emitting diode inside the Sharp GP2Y is powered with a square wave voltage with a 32 ms pulse width, the particles passing through the testing location are illuminated and the light is reflected towards a phototransistor. The light is reflected or scattered more at higher aerosol concentrations since more particles alter the path of light. The infrared-sensitive phototransistor converts the scattered light intensity into a voltage signal. An earlier study [23] showered that the Sharp GP2Y demonstrates the highest linearity against commercialized instruments among the low-cost particle sensors tested, and is stable under humidity and temperature