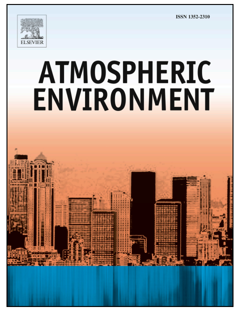
Accepted Manuscript



Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment

Nicholas E. Johnson, Bartosz Bonczak, Constantine E. Kontokosta

PII: S1352-2310(18)30250-4

DOI: 10.1016/j.atmosenv.2018.04.019

Reference: AEA 15952

To appear in: *Atmospheric Environment*

Received Date: 21 August 2017

Revised Date: 9 April 2018

Accepted Date: 10 April 2018

Please cite this article as: Johnson, N.E., Bonczak, B., Kontokosta, C.E., Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment, *Atmospheric Environment* (2018), doi: 10.1016/j.atmosenv.2018.04.019.

Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment

Nicholas E. Johnson[a,b], Bartosz Bonczak[b], Constantine E. Kontokosta*[b]

[a]*University of Warwick, United Kingdom*
[b]*New York University, United States*
[c]*corresponding author: ckontokosta@nyu.edu*

Abstract

The increased availability and improved quality of new sensing technologies have catalyzed a growing body of research to evaluate and leverage these tools in order to quantify and describe urban environments. Air quality, in particular, has received greater attention because of the well-established links to serious respiratory illnesses and the unprecedented levels of air pollution in developed and developing countries and cities around the world. Though numerous laboratory and field evaluation studies have begun to explore the use and potential of low-cost air quality monitoring devices, the performance and stability of these tools has not been adequately evaluated in complex urban environments, and further research is needed. In this study, we present the design of a low-cost air quality monitoring platform based on the Shinyei PPD42 aerosol monitor and examine the suitability of the sensor for deployment in a dense heterogeneous urban environment. We assess the sensor's performance during a field calibration campaign from February 7th to March 25th 2017 with a reference instrument in New York City, and present a novel calibration approach using a machine learning method that incorporates publicly available meteorological data in order to improve overall sensor performance. We find that while the PPD42 performs well in relation to the reference instrument using linear regression (R^2=0.36-0.51), a gradient boosting regression tree model can significantly improve device calibration (R^2=0.68-0.76). We discuss the sensor's performance and reliability when deployed in a dense, heterogeneous urban environment during a period of significant variation in weather conditions, and important considerations when using machine learning techniques to improve the performance of low-cost air quality monitors.

Keywords: Machine learning, Low-cost sensing, Air quality, Urban, Calibration

1. Introduction

Air quality is an important quality of life concern with well-established links to serious respiratory illnesses, cardiovascular disease, and increased mortality rates (Pope III and Dockery, 2006). Cities in particular often experience high levels of fine particulate matter (PM2.5), especially in developing countries where industrial expansion has created unprecedented levels of poor air quality (Cheng et al., 2016). In order to monitor and evaluate levels of PM2.5, government agencies often operate air quality monitoring stations that provide ambient PM2.5 concentration measurements. These networks, however, often fail to capture the granular spatiotemporal variations in PM2.5 levels that can occur over short distances (<1km) (Castell et al., 2017). Urban environments, in particular, contain widely varying mixing ratios with diverse and complex emission sources that require high resolution spatial and temporal monitoring networks to adequately quantify and describe air quality (Mead et al., 2013).

The proliferation of low-cost sensor technologies offers new opportunities to monitor and study air quality in urban environments. A growing body of research has begun to use low-cost aerosol monitors to provide high resolution spatiotemporal measurements by creating dense spatial networks that can inform local and regional emission sources' contribution to total pollution levels, as well as increase the ability to identify pollution hot-spots (Heimann et al., 2015; Jerrett et al., 2005; Shusterman et al., 2016; Manikonda et al., 2016; Moltchanov et al., 2015). Furthermore, these low-cost technologies are often compact, low-powered, and easy to operate, thus offering the ability to establish and facilitate participatory networks(Jovašević-Stojanović et al., 2015; Snyder et al., 2013). High density air qual-