Transportation Research Part D 34 (2015) 104–115

Contents lists available at ScienceDirect

Transportation Research Part D

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

Coherent approach for modeling and nowcasting hourly near-road Black Carbon concentrations in Seattle, Washington

Runze Yu [a], Xiaoyue Cathy Liu [b], Timothy Larson [c], Yinhai Wang [c,*]

[a] East Asia and Pacific Region, World Bank, 17th Floor, China World Office 2, No. 1 Jianguomenwai Avenue, Beijing 100004, PR China
[b] Department of Civil & Environmental Engineering, University of Utah, 110 Central Campus Drive, Suite 2000, Salt Lake City, UT 84112, United States
[c] Department of Civil & Environmental Engineering, University of Washington, Box 352700, Seattle, WA 98195, United States

ARTICLE INFO

Keywords:
Black Carbon
Time series analysis
Nowcast
Near-road pollution

ABSTRACT

With a growing awareness of the importance of near-road air pollution and an increasing population of near-road pedestrians, it is imperative to "nowcast" near-road air quality conditions to the general public. This necessitates the building hourly predictive models that are both accurate and easy to use. This study demonstrates an approach to model the hourly near-road Black Carbon (BC) concentrations given on-road traffic information and current meteorological conditions using datasets from two urban sites in Seattle, Washington. The optimal set of prediction variables is determined with a Bayesian Model Averaging (BMA) method and three different model structures are further developed and compared by goodness-of-fit. An innovative approach is proposed to translate wind direction from numerical values to categorical variables with statistical significance. By modeling the autocorrelation within the BC time series using an AR(1) component, the model achieves a satisfactory prediction accuracy. The conditional heteroscedasticity and heavy-tailed distribution of the model residuals are successfully identified and modeled by the General Auto Regressive Conditional Heteroscedasticity (GARCH) model, which provides valuable insights to the interpretation of prediction results. The methodological procedure demonstrated in selecting and fine-tuning the model is computationally efficient and valuable for further implementation onto online platforms for near-road BC nowcasting. A comparison between the two sites also reveals the effectiveness of local freight regulation for mitigating the environmental impacts from a heavy truck fleet.

Introduction

Numerous studies have identified high air pollutant concentrations in the ambient environment near major roadways, to which the exhaust emitted by the on-road traffic accounts for a major contributor (EPA, 2008; Dockery et al., 1993; Peretz et al., 2008; Van Hee et al., 2009). With a population of over 40 million living within 300 feet (91 m) of major transportation infrastructures (American Housing Survey, 2011) and approximately 250 million vehicles on the roads, air pollution from road traffic has posed great challenges to the sustainability of the modern society.

The U.S. Environmental Protection Agency (EPA) therefore has strengthened the monitoring practice and national standard for near-road air pollutants. Additionally, the increasing number of pedestrians and bicyclists, encouraged by social

* Corresponding author. Tel.: +1 206 616 2696; fax: +1 206 543 1543.
E-mail addresses: runze@uw.edu (R. Yu), cathy.liu@utah.edu (X.C. Liu), tlarson@uw.edu (T. Larson), yinhai@uw.edu (Y. Wang).

http://dx.doi.org/10.1016/j.trd.2014.10.009
1361-9209/

campaigns of sustainable transportation, intensifies the public's concern regarding human exposures to near-road air pollution, and additional interest from the research community (Morgenstern et al., 2007; Bigazzi et al., 2012). Investigating near-road human exposure requires a finer temporal resolution for air quality monitoring and forecasting than the traditional daily average reporting. The latter fails to capture the large temporal and spatial span of human activities and meteorology throughout the day. The concentrations of many air pollutants vary to a great extent, e.g. peak hours' concentrations far exceed the daily averages (Grivas and Chaloulakou, 2006). Short-term exposure has been found to have profound public health impact (Gold et al., 2000), which highlights the practical importance of hourly forecasts. Nowcasting, a concept used to describe short-term forecasting, has been widely executed in weather and economics prediction. In the context of this study, nowcasting would allow the general public to assess immediate air quality conditions and to plan their activities along the roadways (e.g. jogging or biking) accordingly. Air quality nowcasts can be delivered to the public via Geographic Information System (GIS) based online platforms. Previous research (Briggs et al., 1997; Jensen, 1998; Brayerm et al., 2003; Mavroulidou et al., 2004; Vienneau et al., 2009) has demonstrated that those systems have the capability of not only housing a diverse amount of datasets, but also broadcasting forecasts from a backstage computational engine in real-time. This requires the computational engine to optimally select impact factors, adopt computationally inexpensive model structures, and quickly yield satisfying predictions.

A variety of statistical models have been used to build such a computational engine (Gardner and Dorling, 1999; Grivas and Chaloulakou, 2006; Chen et al., 1998; Hrust et al., 2009). Sfetsos and Vlachogiannis (2010) applied Local Models with Clustering Algorithms (LMCA) and Hybrid Clustering Algorithm (HCA) to predicting hourly PM_{10} concentrations. The authors identified that the linear approach marginally outperformed Artificial Neural Network (ANN) approaches. Ballester et al. (2002) built pure predictive models for one-day-ahead forecasts of the hourly ozone concentrations across three sites in Spain. Methods applied involved autoregressive-moving average with exogenous inputs (ARMAX), Multilayer Perceptron (MLP) and FIR neural networks. For ARMAX models, *R*-squared values ranged from 0.75 to 0.8 and the Index of Agreement (IA) ranged from 0.9 to 0.93, on par with both neural networks methods. While exploring different modeling methods, most of previous studies did not sufficiently emphasize the modeling process, including data preparation, variable selection, and model structure. However, the modeling process has considerable influence on accurately capturing the underlying nature of the governing relationship between air pollutant concentrations and impacting factors, and, if conducted carelessly, it may weaken the model performance.

Besides the modeling process, another issue emerges on the data side. Modeling the relationship between hourly concentrations and on-road factors requires traffic data with high resolution and good quality. Traditional traffic counters are able to capture general traffic fluctuations (Lau et al., 2008; Wang et al., 2006), yet fail to provide vehicle classification information, which is desirable for discovering the relationship between air pollutants and diesel fuel vehicles. Some researchers (Johnson and Ferreira, 2001; Reynolds and Broderick, 2000; Zhang et al., 2007) have utilized video data to overcome this issue. Other researchers (Gao and Niemeier, 2007 and Gao and Niemeier, 2008) have relied on Weigh-In-Motion (WIM) systems for high-resolution truck data. However, video data requires large storage space and computing power for processing, and thus can only be collected for a short period of time for validation purposes. Also, WIM stations are generally not located in urban areas and therefore cannot meet the needs of urban air pollution modeling. Additionally, meteorological conditions have a direct impact on observed near-road concentrations (Annand and Hudson, 1981; Grivas and Chaloulakou, 2006; Baldauf et al., 2008; Baldauf et al., 2009; Liang et al., 2013). Among the variety of meteorological parameters, wind direction significantly affects hourly concentration modeling and forecasting (Arain et al., 2007). Some studies used air dynamic model to estimate vehicle emission and emission factors. For example, Kim Oanh et al. (2008) developed a monitoring program that produces data to determine on-road hourly fleet emission and emission factors of vehicles in a street canyon in Bangkok. In this study, the emission factors are back calculated using the Operational Street Pollution Model on the basis of monitored pollutants and traffic conditions. In another study, Giang and Kim Oanh (2014) measured PM2.5 and BTEX to generate input data for CALINE (California LINE Source Disperson Model) inverse modeling. The air dispersion model often requires intensive data generated from a specifically designed monitoring program. And when applied for prediction, the individual vehicle information is required as well. Statistical model, on the other hand, estimates the air pollution concentrations at a relatively macro scale, which loses the constraint on the data resolution.

This paper presents a coherent methodological approach to perform nowcasting of the hourly Black Carbon (BC) concentrations using datasets at two urban sites in Seattle, Washington. BC refers to the soot particles produced during incomplete combustion, with diameter less than 0.88 micrometers (880 nm). Puget Sound Clean Air Agency (PSCAA) monitors BC using Aethalometer™. The Aethalometer™ is a tape sampler that takes in air at a controlled flow rate, and then passes the air through a special glass fiber filter tape. It measures the light attenuation of the BC channel and further converts it into micrograms per cubic meter (BC concentration unit). As a primary component of fine particular matter ($PM_{2.5}$), it is associated with public health risks and global warming effects (Keill and Maykut, 2003; Bond et al., 2013). The objectives of this study are: (1) to address the challenging issues identified from previous research regarding the near-road air pollutant concentration modeling process, including variable selection, model structure, and residual analysis; (2) to investigate the contributions of on-road traffic and meteorological conditions to hourly near-road BC concentrations, particularly by proposing an innovative method to incorporate wind fields into the model; (3) to develop a computationally inexpensive model with desirable prediction accuracy for hourly BC concentration nowcasting and to provide scientific tool for agencies to make decisions about road projects addressing health concerns; and (4) to understand the changes in heavy duty vehicle emissions over time.