Process Safety and Environmental Protection 99 (2016) 149–158

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

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

An assessment of adding value of traffic information and other attributes as part of its classifiers in a data mining tool set for predicting surface ozone levels

Amin Nawahda *

Environmental Research Center (ERC), Sohar University, PO Box 44, 311 Sohar, Oman

ARTICLE INFO

Article history:
Received 13 May 2015
Received in revised form 12 October 2015
Accepted 4 November 2015
Available online 14 November 2015

Keywords:
Ozone
Bagging
DOAS
Prediction
Traffic
Sohar

ABSTRACT

This study seeks to examine to what extent traffic information can improve the prediction of surface ozone levels from mobile sources when coupled with a state of the art air quality monitoring system and the application of data mining tools. For the purpose of the experiment an open-path Deferential Optical Absorption Spectroscopy (DOAS) instrument is used and 10 min video samples obtained from Sohar's main highway (SHW) (Sultanate of Oman). This traffic information is collated to recognize, classify, and count three types of vehicles passenger car; light duty vehicle; and heavy duty vehicle. The DOAS is deployed to measure the following gases; ambient nitrogen dioxide (NO_2); ozone (O_3); sulfur dioxide (SO_2); and BTX (benzene, toluene, xylene) across SHW. The ambient concentrations of these gases are measured in situ at time resolutions that vary from 30 s to 1 min along with simultaneous measurements of meteorological parameters. The Waikato Environment for Knowledge Analysis (WEKA) (Witten and Frank, 2005) software was used for the data mining part of the study. To identify which classifiers in WEKA would be the most suitable in predicting surface O_3 levels the following five indexes were used: correlation coefficient (CC); mean absolute error (MAE); root mean square error (RMSE); relative absolute error (RAE); and root relative squared error (RRSE). It was found that the Bagging and M5P classifiers were the most robust when compared to others within the software when measured against the fives indexes. It was identified that with the additions of time and day of the week as well as changing of the parameters as part of the classifiers in WEKA the robustness of the predictions was not enhanced significantly. However, the findings did illustrate that the analysis of traffic information does improve the robustness of the prediction of surface O_3 levels.

1. Introduction

In recent years, the environmental risks caused by exposure to air pollution from mobile sources globally are increasing on an annual basis. Kagawa (1994) and Shima et al. (2003) find that traffic emissions affect the health of the people living along high-traffic roads in Japan. Sun et al. (2012) analyzed the traffic emissions close to a highway in New York City using a mobile monitoring station equipped with an aerosol mass spectrometer (AMS). Sun et al. (2012) found that there was high exposure of air pollutants for people living in close proximity to the highway.

Based on these studies it has been identified that a better understanding of the emissions of air pollutants from highways is required. This is necessary to ascertain a more realistic assessment of any environmental risks. However, traffic

* Tel.: +968 26720101; fax: +968 26720102.
E-mail address: anawahda@soharuni.edu.om
http://dx.doi.org/10.1016/j.psep.2015.11.004
0957-5820/

emissions in an inventory of air pollutants is the most difficult one to estimate as it is affected by to irregular traffic behavior and the air pollutants interaction with meteorology.

The focus of this study is to look at and predict surface ozone (O_3), which is a secondary air pollutant. It is either transported from other regions or formed by a photochemical reaction with the following three gases: nitrogen oxides (NO_x), volatile organic compounds (VOCs), and the oxygen (O_2) in the air. This will be the first study of its kind that will attempt to predict surface O_3 concentrations. This will be done by taking account of: traffic information; monitoring meteorological parameters; and measuring the concentration of gases by a Deferential Optical Absorption Spectroscopy (DOAS) instrument.

The history and development of data mining tools have been reviewed by number of researchers such as Lovell (1983), Heckerman (1996), Han and Kamber (2000), and Mikut and Reischl (2011). Data mining has been defined as "a step in the knowledge discovery from databases (KDD) process that consists of applying data analysis and discovery algorithms to produce a particular enumeration of patterns (models) across the data" Mikut and Reischl (2011). Data mining and machine learning tools have been extensively used for predicting ground level O_3 and for estimating the missing values in a time series data U.S. EPA (1999), Comrie (1997), Gardner and Dorling (1998), Hubbard and Cobourn (1998), Ruiz-Suarez et al. (1995), and Ryan (1994). However, Cobourn et al. (2000) listed a number of challenges and obstacles that limit the real application of many of the published O_3 prediction models. These are mainly due to the nonlinear behavior and the spatial variability of the chemical interactions of surface O_3 in the atmosphere.

Sohar (Sultanate of Oman) is under rapid development and is expected to face major issues regarding air pollution due to its ongoing and future industrialization. It is located on the Gulf of Oman outside the Strait of Hormuz and highly affected by land-sea breeze circulation and long periods of stagnation (Charabi et al., 2013). Sohar Industrial Port (SIP) is the only seaport in Sohar and has been in partnership with the port of Rotterdam (Netherlands) for over 10 years. The port accommodates a number of heavy industries including steel, refinery and petrochemical industries, etc. In addition there is an ongoing expansion plan of over 4500 ha identified for a free zone and the recent construction of a local airport. These industrial changes have resulted in a major increase in the volume of traffic on SHW. Furthermore, an indirect result of the industrialization, which has also contributed to the increase in traffic, has been job creation that has seen and led to more people having to commute to work and the down steaming of businesses to feed the larger industries. Both of these factors are positive from a social-economic perspective but do have serious environmental implications. Information from the Ministry of Transport and Communications in Oman states that for Sohar's main highway (SHW) the annual mean daily traffic in 2006 was 31,206 vehicles and increased to 45,390 vehicles in 2012.

Therefore, predicting the level of traffic emissions of surface O_3 from SHW is recommended to: first, reduce the level of uncertainty in the assessment of environmental risks (WHO, 2000 and Nawahda, 2013, Nawahda et al., 2013), and second to manage the volume of traffic and the different types of vehicles. It will provide an invaluable tool to obtain a better understanding about the current and future traffic emissions of surface O_3 from SHW.

2. Methodology

2.1. *Instrumentation*

The DOAS system and its comparison with other air quality monitoring systems such as: chromatography; chemiluminescence; electromechanical; gravimetric; and matrix isolation can be found in Nawahda (2015), Mejía et al. (2013), Platt et al. (1979), Kalabokas et al. (2012), Kourtidis et al. (2000), Edner et al. (1993), Hao et al. (2006), and Psiloglou et al. (2013).

The deployment of the DOAS instrument in the measurement of traffic emissions has a number of advantages such as: multi-analysis; open-path measurements; and rapid measurements of gaseous species. These characteristics eliminate the effect of wind direction, which is a major concern in the conventional point-measurement systems. In the DOAS system light is projected or reflected across an open-path. The gaseous molecules absorb light from certain ranges of the spectrum. The light is captured and then transmitted using an optical fiber cable to the analyzer. The spectrometer splits the light into narrow wavelength bands using an optical grating. The bands are processed and evaluated to get the best estimation of the concentration of gases in the light path using the Beer–Lambert's absorption law. This law illustrates the relationship between the quantity of light absorbed and the concentration of gaseous species in the light path as illustrated by the following equation;

$$I_1 = I_0 e^{-aLC} \tag{1}$$

where I_1 is the intensity of the reflected or received light, I_0 is the intensity of the emitted light in the light path, L is the length of the measuring light path, C is the concentration of the gaseous specie to be calculated, and a is the absorptivity coefficient. A detailed mathematical description of the derivation of different forms of Beer–Lambert's law is found in Rozanov and Rozanov (2010) and Petraskis et al. (2003).

The light source in the emitter unit of the DOAS system is a high-pressure xenon lamp. This light source radiates an almost smooth spectrum within the range from 200 nm to 500 nm. Within these ranges NO_2, O_3, SO_2, and BTX molecules provide specific absorption spectra. The collected light is directed to a grating which refracts it towards a rapid scanning slit where a photo-multiplier sensor detects certain parts of the spectrum. This generates an electrical current in the sensor, which is converted into a digital signal. This signal is stored in a logger for an evaluation based on Eq. (1). In order to capture the rapid variations of the concentrations of gaseous species in the monitoring path, the evaluations of the captured light in the DOAS instrument is analyzed every thirty seconds for measuring the concentrations of O_3, NO_2, and SO_2, and 1 min for measuring the concentrations of BTX with simultaneous measurements of meteorological parameters (OPSIS AB, 2012).

2.2. *Air quality data*

The open-path measurements taken by the DOAS system of air quality and meteorological parameters were conducted in front of Sohar University (SU). This instrument was installed and calibrated in March 2103. The light beam travels a round trip of 477 m from the light emitter to the reflector. The emitter was installed on the roof of SU and the reflector was installed