Sensors and Actuators B 267 (2018) 245–254

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

Using statistical methods to carry out in field calibrations of low cost air quality sensors

José María Cordero*, Rafael Borge, Adolfo Narros

ETSII-UPM, José Gutiérrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

Article history:
Received 28 November 2017
Received in revised form 14 March 2018
Accepted 4 April 2018
Available online 11 April 2018

Keywords:
Electrochemical sensors
Neuronal networks
Uncertainty
Nitrogen dioxide
Air quality

ABSTRACT

The poor air quality found in big cities is harmful to human health. The aim of the LIFE PHOTOSCALING Project (LPP) is to assess the effectiveness of different photocatalytic pavements in reducing NO_2 pollution. The objective of this preliminary study is to determine how well low cost AQmesh sensors can accurately enough measure NO_2 concentrations to be able to determine the effects of using photocatalytic pavements. Data was collected from AQmesh sensors that were installed in two Air Quality Stations in order to monitor NO_2 under traffic and urban background conditions. The NO_2 measurements were unreliable, resulting in an unsatisfactory level of accuracy. A two-step calibration method was devised in order to overcome this limitation. This method consisted of supervised statistical machine learning regression algorithms. A first Multivariate Linear Regression provided a new explanatory variable that contained valuable information about the error. This variable was fed into more sophisticated equations, such as Random Forests, Support Vector Machines and Artificial Neural Networks. The various models were evaluated by calculating statistics of errors and relative expanded uncertainties, and through Taylor Diagrams. After a careful calibration, AQmesh sensors met the Air Quality Directive's standards of accuracy at high concentrations of NO_2. However, we found that each individual sensor behaves differently and thus, each unit requires the development and application of a specific calibration model.

1. Introduction

Worldwide, poor air quality leads to 3 million deaths annually [1]. Ambient air pollution is a cause of premature deaths and various respiratory conditions, such as throat irritation, congestion and asthma [2]. Motor vehicles are the principle source of NO_2. High concentrations of this pollutant are routinely detected in areas of traffic. Therefore, the source of NO_2 emissions is found to be proximal to pedestrians, causing a harmful situation.

Within the framework of its current sustainable development initiatives, the European Union (EU) approved The Air Quality Directive (AQD) [3]. The AQD established thresholds for different types of pollutants. EU member states are bound by this directive to apply the necessary measures to bring pollutant concentrations down below these thresholds.

However, NO_2 is routinely detected at higher than permitted levels at Air Quality Stations (AQS) in large cities that have high motor vehicle traffic, such as Madrid. AQS networks continuously monitor pollutants. From these monitoring activities we know that, under adverse weather conditions, pollutants cannot dissipate, leading to levels of NO_2 well above the AQD thresholds. Measures are taken to minimize the impact of these weather events, such as restricting motor vehicle traffic. As a consequence, the normal circulation of motor vehicle traffic is altered, disrupting the daily routines of the citizenry. Alternatives to the traffic restrictions are technologies for reducing the NO_2 in the atmosphere such as installing photocatalytic pavements or paints in roads, sidewalks and facades [4–6] or photocatalytic urban textiles in sunshades [7]. These technologies rely on TiO_2-based materials that are activated by UVA radiation. Upon activation, these materials are able to oxidize NO_2 into nitrates.

The high cost of AQS networks means that pollutants can only be measured in a limited number of locations at a time. This makes it impossible to model a complete time space distribution. Passive samplers have been used to study the distribution of pollutants at the microscale level [8]. However, passive sampling is a cumulative method and cannot be used to carry out absolute measurements at discrete time points. As a consequence, complex distributions in the horizontal and vertical profiles cannot be performed via passive sampling. In addition, the effect of other meteorological or ambient variables on NO_2 measurements cannot be determined using passive methods.

---

* Corresponding author.
E-mail address: jmcordero@etsii.upm.es (J.M. Cordero).

https://doi.org/10.1016/j.snb.2018.04.021

Low-cost sensors (LCS) are being developed through initiatives such as AirSensEUR [9,10] and are based in both electrochemical or solid state microsensors designed to monitor urban ambient air gases like NO_x, CO, SO_2 or O_3. Networks of mobile or fixed LCSs gas sensors permit the mapping of wider areas of the city [11] with a high time resolution. The sensors are normally calibrated against the measurements recorded in nearby AQSs in a co-location campaign. LCSs are versatile, autonomous, and can make measurements at several points over wide areas [12], as well as remote spaces [13], reducing the cost of detection in comparison to AQSs and allowing the generation of large observation datasets. However, this technology has its challenges as well: LCSs show cross-sensitivity between the target gas and other gases in the urban atmosphere. In particular, NO_2 and O_3 electrochemical sensors are prone to show mutual cross-sensitivities [14]. LCSs are also dependent on weather conditions, and suffer from short term and long term drift. In addition, units of the same type are highly variable [15], poorly selective, and not very stable [16–18].

LCSs have been extensively checked and calibrated in laboratory. However, the results change dramatically when the devices are deployed in urban sites [15,17–19]. This is because the conditions of the urban atmosphere cannot be exactly reproduced in the laboratory. Consequently, in-field calibrations of these devices, under the conditions in which the LCSs will be used, are essential for obtaining accurate readings in the urban atmosphere. Therefore, a calibration method has to be designed for each individual device under the particular conditions it will be used.

These limitations make the calibration of these devices difficult, and complex equations are required to assure the reliability of these instruments. As part of the MacPoll Project, the Joint Research Centre has published several reports on this subject and has designed a calibration protocol for LCSs [20]. In most of these studies, statistical methods, such as linear regression (LR), multivariate linear regression (MLR) and artificial neural networks (ANN) [13,21,23,25] are used to fit the signals. A recent work provides a comprehensive comparison of several machine learning algorithms, including training, optimization and validation [22]. Some ANN studies have relied on predictive variables from different sensors as inputs [19,20], an approach that solved the primary problem of interference between sensor measurements and external variables. This strategy has led to more accurate results. Another approach has been to use a reference station to update the parameters of the calibrated sensors, thereby correcting for the drift [24]. In other study [26] other sophisticated machine learning algorithms such as boosted regression trees and Gaussian process emulators were successfully applied to process the raw signal of their LCSs, solving the main complex interferences in the measurements. Here we report the results of using ANNs and less commonly utilized methods, such as Support Vector Machines (SVM) and Random Forests (RF), to fit the data using AQS reference data as objective variables.

The goal of the LIFE PHOTOSCALING Project (LPP) is to achieve the sustainable use photocatalytic technologies on urban pavements. As part of the LPP, different photocatalytic materials will be tested for using as pavements [27]. To evaluate the effectiveness of these materials, accurate and precise methods for measuring NO_2 are crucial. The objective of this study is to validate the LCS gas sensor technology for use in the context of the LPP. We report here a more advanced method for calibrating LCS devices, making them more accurate and precise.

AQmesh units (PODs) were used as model LCSs. Manufacturer calibration of these devices was not found to be satisfactory in terms of accuracy. Four PODs were installed in two AQSs in the Community of Madrid. These AQSs covered two different ambient conditions: traffic and urban background. Data was gathered from both the PODs and AQSs. Various statistical methods were applied to fit the data. Interactions between variables that obscured the correct measurements were evaluated. This process had to be repeated for every POD under each ambient condition. This study should be repeated in order to account time drift [20] and seasonally variable meteorological conditions.

2. Materials and methods

2.1. Air quality stations

Reference measurements were collected from two AQSs. One of these was located in a traffic area in a neighbourhood of Arturo Soria street in the city of Madrid. The other was located in an urban background area in the town of Arganda del Rey, located about 25 km southeast of Madrid city. As a consequence, the behaviour of the AQmesh PODs could be studied under two different environmental conditions. The measurement techniques utilized were:

- Chemiluminescence, to measure levels NO and NO_2.
- UV absorption, to measure levels of O_3.

Meteorological parameters were recorded only in Arganda del Rey because the Arturo Soria AQS did not have a meteorological station.

The Arturo Soria AQS is managed by the Madrid City Council and the Arganda del Rey AQS is managed by the Comunity of Madrid autonomous government.

2.2. AQmesh PODs

AQmesh PODs were selected as model LCSs. Our group had extensive experience with LCSs that were either electrochemical or MeO based. AQmesh was chosen because they provide accurate readings over a wide range of values: from 0 to 4000 ppb for NO and NO_2, and from 0 to 1800 ppb for O_3. All the gas sensors in the AQmesh PODs tested were Alphasense electrochemical sensors [28]. Detailed technical specifications can be found in reference [29]. The POD's supplier applies a proprietary algorithm which converts the electrical signals into concentration units. The user could download the resulting information from a website. From here on in, we will refer to this downloaded data as "raw" despite the prior treatment. However, treating the data with the proprietary algorithm alone did not result in the required level of accuracy. Therefore, the data required further treatment.

2.3. Calibration by statistical models

Four PODs were placed on an iron stand near the sampling points of selected AQ stations. This operation was carried out between 11–28 March 2017 in Arturo Soria, and between 28 March to 10 April 2017 in Arganda del Rey. Once the operation was finished, both the data from the reference stations and the data from the PODs were utilized for the calibration. The calibration of the PODs was carried out using statistical machine learning methods.

The data treatment was performed using the open source language R and using the IDE Rstudio software [30,31].

At first sight, an acceptable correlation was observed between the station and the POD raw data. Only a small variability between the data sets was observed. For being low cost sensors, AQmesh performed relatively well, but did not provide the accuracy required for use by the LIFE PHOTOSCALING Project. Therefore, the data was treated further.

To carry out the data treatment, a novel regression algorithm was applied in two steps (see Fig. 1). First, a MLR model was fitted using explanatory variables measured by the sensors (C_{NO2}, C_{NO}, C_{O3}, T). RH was discarded applying a backward elimination