IOP Conf. Series: Journal of Physics: Conf. Series **1065** (2018) 212023 doi:10.1088/1742-6596/1065/21/212023

Uncertainty estimation of road-dust emissions via interval statistics

M de-Angelis[1], V Ricciardi[2] and E Dalmau[2]

[1] Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, UK
[2] Automotive Engineering Group, Technische Universität Ilmenau, Ilmenau, ,Germany

E-mail: marco.de-angelis@liverpool.ac.uk

Abstract. Particulate matter, a.k.a. particle pollution, is a complex mixture of small particles and liquid droplets that are present in the air. Once inhaled, these particles can affect the heart and lungs and cause serious health problems. A recent study, based on geographically referenced datasets of pollutant emissions has shown that non-exhaust related pollution is at present dominant and increasing. Emissions from paved roads are poorly estimated due to the lack of knowledge about the resuspension process. Recent literature works have attempted to provide a reliable framework for the estimation of emission factors. Estimations are obtained by linear regression with a single-valued discriminant for the acceptance/rejection of the experimental dataset based on the evaluation of the r-squared coefficient. In this paper, we explore alternative methods to evaluate the "quality" of the data and consequently discriminate whether a given sample can be accepted to provide estimation of the emission factors. Uncertainties are characterised both in the data and in the statistical model. Measurements are expressed with interval-valued datapoints to include the experiment precision directly within the estimation process. Alternative fitting techniques that avoid the use a single-valued discriminant are also explored for an inclusive estimation of the emission factors.

1. Motivations for the study

Road transport emissions are a major contributor to ambient particulate matter (PM) concentrations and have been recognised as a cause of life threatening health effects [1]. Motor vehicle emissions, which represent a main source of urban pollutant emissions, can be divided into two main categories depending on their source: exhaust (engine) emissions and non-exhaust emissions. Exhaust emissions result from incomplete fuel combustion and lubricant volatilisation that occur during the combustion process. Non-exhaust emissions arise from brakes, tires, general vehicle wear and resuspension due to traffic-induced turbulence [2]. The London Atmospheric Emissions Inventory [3] database of geographically referenced datasets of pollutants emissions and sources in Greater London envisages that non-exhaust related PM emissions are at present dominant and the situation is going to worsen if countermeasures are not taken.

It is expected that the progressive transition towards the electric mobility, and the consequent increasing vehicle weight, may produce higher emissions of non-exhaust PM, and therefore higher deposition of dust on the roads [4].

2. Uncertainty estimation

The proper assessment of emissions, especially non-exhaust related, for the quantification of emission factors is a difficult task and, at present, there is still no standardised testing protocol and/or measurement method. Uncertainty comes in different forms and affects the experiment setting, the measurements, and the model used to obtain the emission factors. Particularly uncertain is the on-road testing and measurement of emitted dust particles in real driving conditions, in contrast to laboratory testing, which measurements are performed under controlled conditions. This is due to the continuously changing factors, like flow conditions, driving dynamics, fine dust particles from other sources, etc.

The lack of a unified procedure for the assessment of road-dust emissions makes the experiment setting vary depending on the country, road and air conditions, e.g. state of pavement, temperature, humidity, rainfall, climate, etc. Several approaches have been proposed in the literature, however lacking a unified framework. Strong assumptions are usually made to quantify the mass of pollutant obtained during the experiments. The assumptions mainly entail:

- the physical model used to quantify the emission flux of particles;
- the statistical model used to estimate the emission factors.

In this study the focus is on the second assumption, the statistical model.

The PM10 emission factors are obtained by means of a specific sampler utilised by Orza et al. [7] that measures the micro-scale vertical profiles of particle concentration. PM10 values are obtained by fitting the physical model of dust mass balance to the experimental data.

3. Current practice for the estimation of road emissions

Emission factors are estimated using the micro-scale vertical profiles of particle matter concentrations [7]. The technique estimated the horizontal aeolian sediment fluxes. The sampler consists of a number of cylindrical canisters of fixed height and diameter (usually 6cm height and 14 cm diameter). The canisters are vertically arranged and spaced by 1 cm.

The samplers are exposed to ambient air for passive collection of particulate matter [6].

3.1. Physical model for dust deposition flux

The terminal concentration c divided by the area of deposition and the sampling time provides the deposition flux J which can then be converted to emission rate [5].

The approach makes use of a simple model based on mass balance of dust deposited on the road surface. Let M be the mass of road dust per unit area at a time t, and J the dust deposition flux. Accepting that road dust removal (i.e. emission flux) is a continuous process described by an unknown function of M, denoted by f(M), the time evolution of M must obey:

$$\frac{dM}{dt} = J - f(M) \tag{1}$$

Two additional assumptions are made: (i) J is constant in time, and (ii) f(M) increases monotonically with M. Indeed, these hypotheses are often tacitly assumed.

If it is assumed that, at a certain time, the value of M is such that f(M) less than J, at the next (infinitely close) time, M will increase as will f(M), so that the difference J e f(M) will become smaller. In an infinite succession of such events, f(M) will progressively augment as its magnitude balances with J, and dM/dt approaches zero. As a result, in a sufficiently long time, an equilibrium value of M will be reached.

3.2. Regression analysis for parameter estimation

The regression is performed by fitting the exponential model of Eq. 2 to the data as shown in Figure 1. The obtained parameters a, b and c are then used to estimate the emission factors.

$$k(x) = a\ e^{-b\ x} + c \tag{2}$$

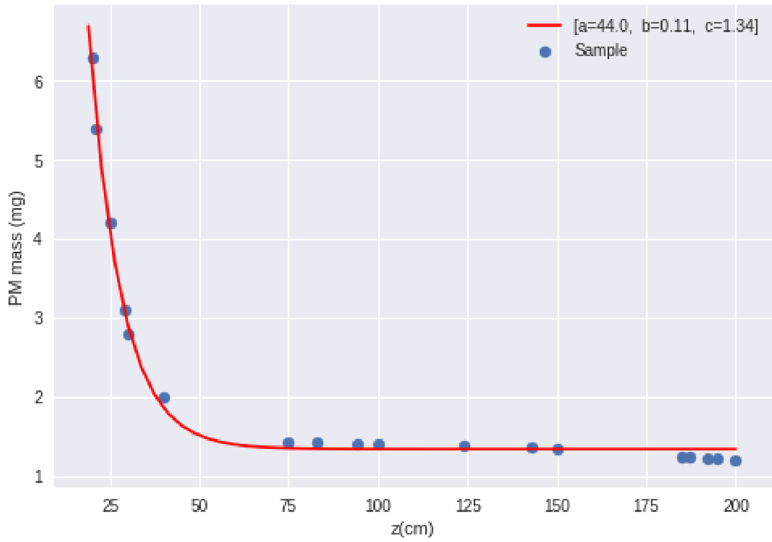

Figure 1. Augmented interval dataset.

Experimental data are often scattered and associated with limited precision. The coefficient of determination, r^2, a.k.a. r-squared coefficient (see Eq. 3) from the fitting procedure is used to judge the acceptability of a given sample.

$$r^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \overline{y})^2}, \quad \overline{y} = \frac{1}{N} \sum_i^N y_i \tag{3}$$

In Amato et al. [6] samples with values of r^2 strictly smaller than 0.7 are discarded. Contrarily, any sample with $r^2 \geq 0.7$ is accepted.

The procedure described in Amato et al. 2012 [6] does not consider the experimental uncertainty inherently present in the data. Taking the experimental uncertainty into account will remove the fictitious assumption of single-valued data points. Interval datasets can be treated equivalently to single-valued ones and fitted to a mathematical formula. The resulting fitted model will be expressed as an envelope of curves rather than a single one. This procedure adds rigour and robustness to the statistical fitting procedure and provides a better discriminant to assess the validity of an experimental sample set. The r-squared coefficient will be obtained as a whole range which will then be compared against a given discriminant.

3.3. Limitations of the existing practice

The existing estimation procedure makes several assumptions. The most prominent ones are the following:

- The dust removal function f(M) is unknown and assumed dependent on M only
- f(M) is strictly increasing
- The exponential model of Eq. 2 is a consequence of the above model's assumptions
- The discriminant to accept samples is a single-values coefficient of determination

Including the measurement precision inherently in the data and performing robust regression will weaken the fourth assumption. In addition, using a more reliable prediction numerical tool for the regression analysis, such as IPM or ANN, will remove the third assumptions and will provide more insights into the physical model.

4. Statistics for imprecise data

Regardless of the instrument accuracy the precision of a single datum is always finite. In fact, substituting single measurements with real numbers is a sometime strong assumption. This is even more applicable when measurements come from devices that were not originally designed for that specific purpose, like in the case of experimental settings to quantify the amount of non-exhaust particles on paved roads.

When the data set has imprecision, computing statistics can be challenging. For example, for data in the form of intervals, using naïve interval analysis yields results with inflated uncertainty because of the repeating variables problem. Moreover, finding optimal bounds on many basic statistics is a NP-hard problem that grows in difficulty with the size of the data set. It is practically impossible to solve these problems for large data sets with simple space-filling strategies, such as random sampling, in which, for instance, the formula for the variance is treated like a black box evaluated for many possible configurations of the data points within their respective intervals.

Ferson et al. [8] provided a suggestive solution to the problem by looking at the different classifications of interval data. In this work, it is shown how both bounds on the variance can be computed in linear time in several cases, without the need to solve a quadratic programming problem. The work detailed in [8] is being developed via accessible software libraries [9] that are intended to provide convenient access to basic statistics for interval and censored data. The library of algorithms develop on-line and stand-alone software for analyzing data sets containing imprecision as well as sampling uncertainty. The algorithms in the library require users to make fewer dubious assumptions about the data set than currently popular methods for handling data censoring, missingness, and lack of independence. The library currently supports methods to compute over two dozen measures of location, dispersion and distribution shape, including confidence intervals, as well as several inferential methods for linear and logistic regressions, t-tests, F-tests, and outlier detection.

5. Regression analysis with imprecision

Robust estimations are obtained performing regression analysis on the interval dataset. The regression analysis consists in fitting a three-parameter exponential to the data as shown in Figure 1. When the data are in the form of intervals regression analysis needs to be performed virtually for every possible combination of datapoints within the intervals. The results of interval regression are in the form of an envelope of optimal fitting curves which include all the possible solutions of the least square problem. The numerical technique is compared with a double-loop optimisation, where the inner loop computes the optimal fitting via least square minimisation and the outer loop finds the limiting curves of the regression envelope.

6. Conclusions

Statistical fitting techniques may hinder the actual physics of the phenomenon under study if used to determine the parameters of the model.

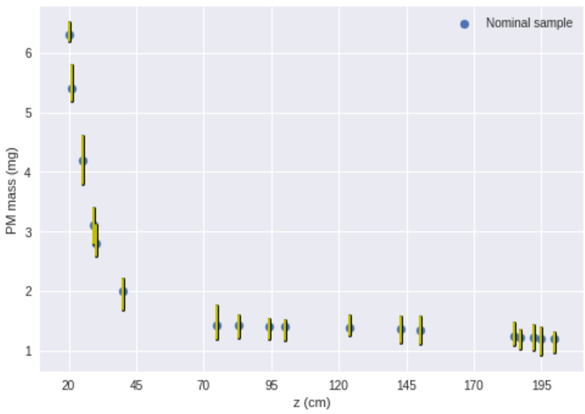

Figure 2. Augmented interval dataset.

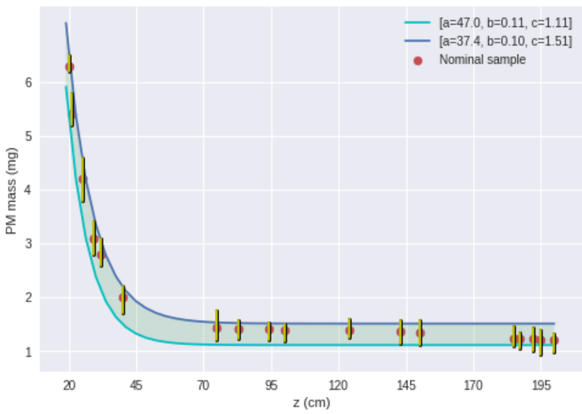

Figure 3. Robust interval regression.

Assumptions on the model are strong and should be supported by appropriate uncertainty models.

7. Future work

The work will benefit from replacing the least-square regression with a kernel based IPM that can also be used for outlier detection to produce more reliable statistics. This links with improving the model of the dust removal function, and provide better connection to the emission flux of particles.

References

[1] Kim KH, Kabir E and Kabir S 2015 *Environ. Internat.* **8** 74
[2] Thorpe A and Harrison RM 2008 *Sci. of the Total Environ.* **13** 400
[3] Authority Greater London 2008 *London Atmospheric Emissions Inventory*
[4] Timmers V RJH and Achten P AJ
[5] Escrig A, Amato F, Pandolfi M, Monfort E, Querol X, Celades I, Sanfélix V, Alastuey A, Orza JAG 2011 *J. of environ. manag.* **5** 92
[6] Amato F, Karanasiou A, Moreno T, Alastuey A, Orza JAG, Lumbreras J, Borge R, Boldo E, Linares C, Querol X 2012 *Atmospheric environ.*
[7] Orza JAG, Cabello M, Lidón V, Martínez J 2009 *Advances in Studies on Desertification*
[8] Ferson S, Kreinovich V, Hajagos J, Oberkampf W, Ginzburg L 2009 *Sandia Nat. Lab., SAND2007-0939*
[9] de-Angelis M, Ferson S, Green L 2017 *Proceedings of ISIPTA2017*