Network Calibration Methodology

1. Introduction

Conventionally, routine monitoring of air pollution relies on the use of expensive, high maintenance instruments and infrastructure. The Breathe London project deployed 100 lower-cost air quality sensor nodes each measuring nitric oxide (NO), nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}) and carbon dioxide (CO_2) at a 1-minute time resolution. The project provided an excellent opportunity to study the performance of a large network of lower-cost sensors and to determine their reliability and accuracy compared to London's well-established reference network. A key challenge of operating any air quality monitoring network is the level of effort required to calibrate the devices, particularly for large networks. The cloud-based network scaling method, one approach used to calibrate the Breathe London measurements, remotely calculates individual calibration factors for each sensor within the network. The method builds on previously published work reported in Heimann et al. (2015) and Popoola et al. (2018), and dramatically reduces the level of effort needed to operate a lower-cost network by allowing calibration of large numbers of sensors without extensive co-location studies. The network calibration approach was trialled in the Breathe London project, which provided an excellent testbed for this approach, as numerous physical co-locations of sensors with reference instruments were undertaken during the study for comparison purposes. This methodology has the potential to be scaled in other networks (with sufficient temporal resolution) to reduce the measurement uncertainty of networked sensors.

2. Methodology

2.1 Separating pollution scales for use in calibration

Atmospheric gas and particle signals observed at a given receptor point can be conceptualised as being broadly made up of three signals: local, near-field and far-field/background. The local contributions are due to emissions near the receptor (1 – 10 m), the near-field contributions vary on scale of 10 – 1000 m and represent the contribution from dispersed sources local to the environment, while the far-field/background contribution is uniform over 1-100 km. The background signal represents contributions from long range transport upwind of the monitoring domain and this signal is frequently relatively uniform across the field of monitoring. Intuitively, if the network is properly calibrated, the backgrounds signals detected at each measurement location will be very similar. Conversely, for a poorly calibrated network, the site-specific backgrounds, while exhibiting good inter-site correlations, will show site to site differences which can be characterised by a range of correction factors (e.g. gains and offsets).

The basis of the cloud calibration approach is to extract the far-field signals from the high temporal resolution data for each uncalibrated node within the network and to use them to provide series of

relative calibration factors for the entire network. Once this analysis is completed, the *absolute* calibration of the network is obtained by calibrating a single selected pod, by collocation, against a suitable reference site. The statistical estimators obtained from this relationship represent the calibration parameters for each node in the entire network. Previous studies have shown ways of separating the different pollution scales for an already calibrated network (Heiman et al., 2015, Popoola et al., 2018), and the methods for the Breathe London network derive from this.

2.2 Introducing correction methodology-conceptualised case

The following simulated data set illustrates the concept of local, non-local (near- and far-field pollutant signals), that is, the different fractions that make up the total pollutant signal recorded at any ambient monitoring location. The data shown in **Figure 1** is a simulation to illustrate two receptor locations (clean and polluted). We have assumed in this illustrative case that the two sensors are identical, i.e. signals have same calibration (gain=1 and offset=0) so that the differences in the overall signal are solely driven by the characteristics of the emissions at the two receptor locations. This assumption simplifies the presentation below, but of course the argument translates without change to cases where sensor (gains, offsets) differ. The clean site is assumed to be characterised by sparse low magnitude local pollutant signals, while in contrast, the polluted site has both relatively high magnitude and more frequent local signals which is typical of urban kerbside stations, both representing behaviour similar to real-world locations. Both locations have the same far-field contribution but different near field and, obviously, different local contributions. At each site, all three signals can be separated into two signals: local and non-local, the latter being the combination of the near- and far-field signals. These simulated data are characteristic of fast temporal observation data (≤ 1 minute), the typical time resolution required for distinguishing the different pollution scales as described in previous work (Heiman e al., 2015, Popoola et al., 2018). The hourly averages for the two receptor locations are also presented (**Fig 1 (e)**) to simulate the typical time resolution for the data reported by conventional monitoring station.

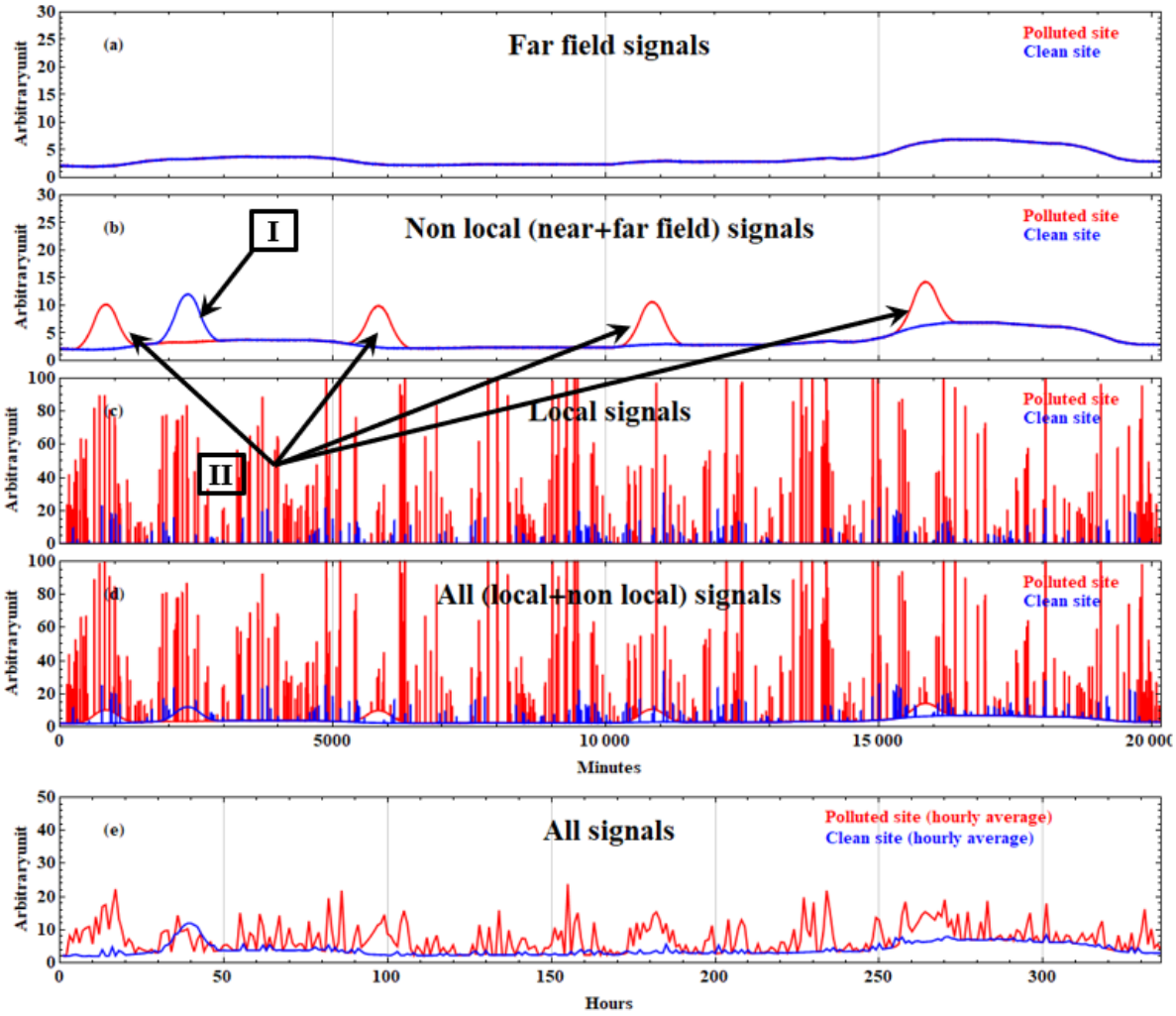

Figure 1. *Time series showing simulated pollution data typical of two sites (polluted and clean). (a) far-field signals for 1-minute data, (b) non-local signals (combination of far and near field) for 1-minute data, (c) local signals for 1-minute data , (d) combination of local and non-local (all) signals for 1-minute data and (e) hourly equivalent of all signals. Note I and II represent the near-field signals contribution to the non-local signals for the clean and polluted sites respectively for 1-minute data.*

2.3 Estimating calibration parameters from non-local signals

Figure 2 shows scatter plots of the 1-minute data and hourly averaged data at the two receptors locations. For clarity, the number of data points at each aggregated point, a useful measure of where most of the data points lie within the scatter plot, are also shown. A simple linear fit to the paired sensor data is also shown in each plot (as expected, these are similar, characterised by low gradients << 1 and similar offsets ~ 4) along with the 1:1 line for comparison.

As already stated, the sensors at the two sites are assumed to have the same calibrations, although they are sites with different levels of assumed pollution. As we have shown from the simulated 1-minute data in section 2.2., the data that lie near the 1:1 line (**Fig. 2**) will be dominated by the far field signals, while the outliers will mainly be due to the local signals and to a lesser extent, the near field signals. As can be seen, in both plots in **Figure 2**, the distributions are dominated by the (different) local signals at each site and no direct comparison of sensor performance can be made.

However, at higher time resolution (**Figure 2 (b)**), while the distributions are visually dominated by small numbers of highly displaced points representing individual pollution events at the two sites, there is a significant core of points lying on or close to the 1:1 line which represent (in this case) the backgrounds measurements of the two identical sensors. The essence of the calibration methodology is to identify and remove the extremes shown in **Figure 2 (b)**, leaving the core background measurements which defined the relative calibration of the sensors at the two sites.

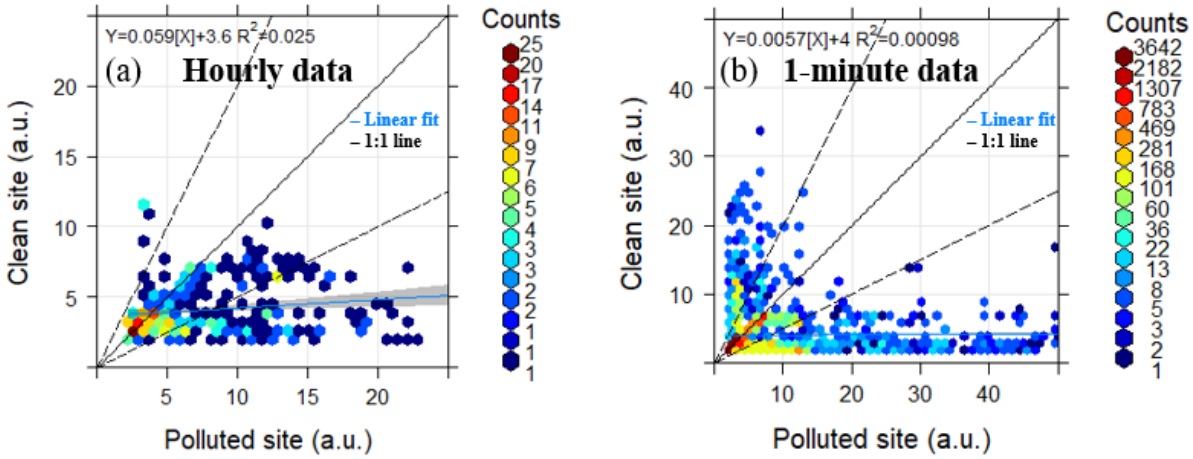

Figure 2. *Scatter plots for the all signals at the two receptor points. (a) hourly averaged and (b) 1-minute data equivalent. The 1:1 line is shown as the solid line, while the 2:1 and 0.5:1 line are presented as dashed lines either side of the 1:1 line. Note the total number of data points are 20161 and 336 for (a) and (b) respectively. Even though the same colour palette is used in the two scatter plots, the maximum data count in (a) is 3642 compared to 25 for (b). The linear model line (solid blue) and 95% confidence interval (grey shading). Also note the axes range are different in (a) and (b).*

The analysis method is in three stages: 1) Identifying and extracting the local signals to leave the non-local signals (near-and far-field), 2) separating the non-local signals to leave the far-field signals, and 3) Establishing the relative sensitivities of the two sensors from the far-field signals.

We address the first stage by exploiting the information from high temporal resolution data, which makes it possible to identifying the local signals (which is not generally possible using hourly resolution data).

Once stage 1) is complete, the simplest model that might be thought relevant is a linear regression model, where the slope and intercept represent the relative gains and offsets of the two sensors. However, such a model is still sensitive to the different near-field signals at the two sites (the signals "I" and "II" in **Fig. 1 (b)**), as such, the fit parameters obtained (slope=0.25 and offset=2.3 (see green line in **Fig. 3 (a)**) do not reflect the relative calibrations of the two sensors (i.e. gain=1 and offset=0).

The final step is therefore to use a multivariate normal distribution to characterise the non-local signals. Other statistical methods can be applied to establish the relationship between the two signals provided it excludes the outliers (near field signals). Using the bivariate normal method, we

obtain two scaling parameters (G_{corr} and O_{corr}) which are subsequently adopted as the relative slopes and offsets which are then applied in a simple linear fit equation to scale the raw sensor data as shown in **Eqn. 1**. Details of all the derivations and statistics (estimators) are presented in the manuscript in preparation.

$$corrected\ data = (raw\ data) * G_{corr} + O_{corr} \quad (1)$$

The correlation plot for the non-local signals and the corresponding plot for the optimal bivariate normal joint pdf are presented in **Figure 3**.

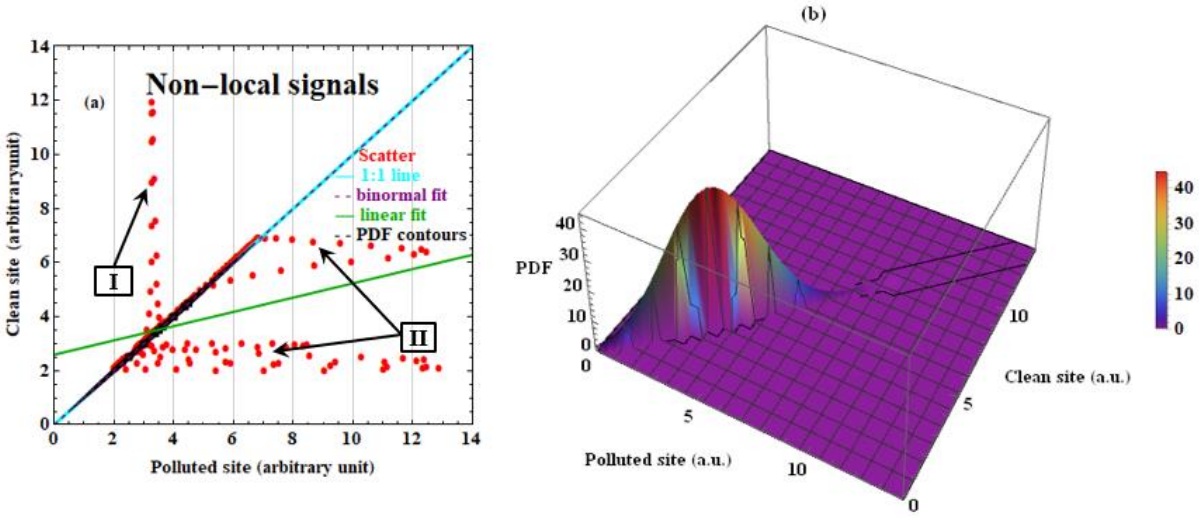

Figure 3. *Scatter and pdf plots for non-local signals at the two receptor points. (a) scatter showing the bivariate normal, linear model fit lines and the contour plot (black) of the pdf for the retrieved non-local signals and (b) 3-D probability density function, pdf, plot for the optimal bivariate normal relationship describing the relationship between the two extracted non-local signals. Note the outliers identified as "I" and "II" in Fig. 2 (a) are related respectively to the distinct near-field contributions to the non-local signals for the clean site and the four distinct near-field contributions to the non-local signals for the dirty site in Figure 1.*

The results presented in this section show that the method reproduces the expected calibration (values of 1 and 0 for the relative gains and offsets) for the two pods within acceptable statistical uncertainty.

2.4 Cloud-based Network Calibration Protocol-Breathe London project data

This analysis method presented for the simulated data was replicated for the Breathe London 1-minute data set with (n=100) pods (high spatio-temporal resolution data). We identify one pod (gold pod) within the network for co-location, while the remaining (the n-1) pods in the network are

classified as the independent variable resulting in n-1 unique scatter relationships between the gold pod and the individual pods. This involves 3 steps with as shown below.

1) Derive the *relative* calibrations for all the individual pods relative to the gold pod using the extracted non-local signals.

2) Estimate the *absolute* calibration by co-locating the gold pod with a reference station for a short period, typically 1 week. A suitable calibration period is required where the atmospheric variability of the pollutant is sufficient to allow a meaningful calibration of the two devices.

3) Transferring the *absolute* calibration from step 2 onto the relative calibrations from step 1 to obtain the *absolute* network calibration parameters.

Steps 1-3 can be repeated for varying time windows to obtain time-dependent calibration parameters if the sensors are not stable over a long time period (1-2 years).

The general protocol for the cloud-based network calibration methodology adopted for the Breathe London project is summarised in the schematic shown in **Figure 4**

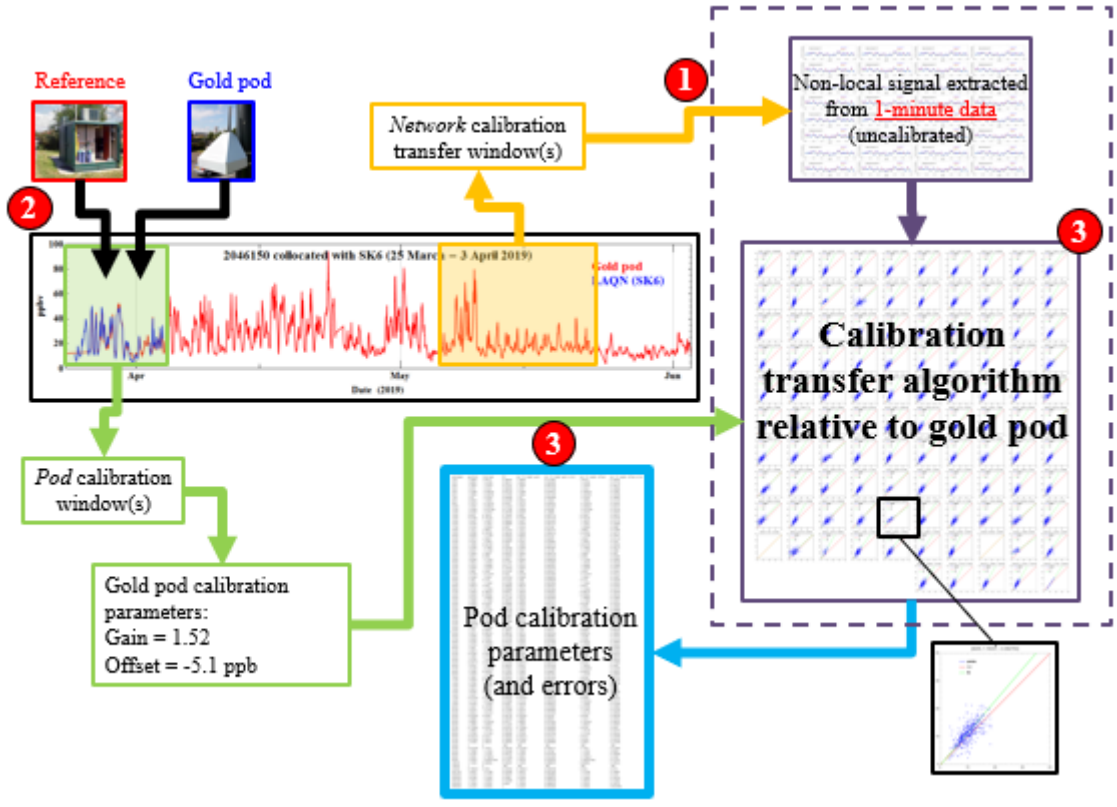

Figure 4. *Framework developed for Breathe London cloud-based calibration transfer methodology.*

3. Evaluation of the Network Calibration Method

The Breathe London project carried out numerous physical co-locations of sensor pods at regulatory monitoring sites, allowing us to evaluate the performance of the network calibration method by comparing sensor measurements (with and without calibration) to the co-located reference instrument measurements. This work is described in detail in **Appendix 4**. In brief, we found that the network method reduced the magnitude of overall network bias and reduced measurement error compared to uncalibrated data. The results also show comparable performance to the physical calibration method for similar data comparisons.

4. Replicability

The cloud-based network calibration method discussed above is designed to provide a low-cost but high-quality alternative to the traditional collocation or gas calibration methods which are traditionally used for maintaining networks of air quality instruments. The context to this is that the maintenance and quality control of any air quality network is key to ensuring sound and reliable scientific outcomes, and with this, is essential for well characterised and meaningful political interventions and policies. For traditional networks the cost of maintaining the QA/QC protocols is often, annually, a significant fraction of the initial capital costs, so that any economic advantage of using a low capital cost network is lost unless alternative calibration methods can be developed and deployed.

A cloud-based calibration method has several ancillary benefits, including real time readily repeatable network QA/QC, in contrast to physical calibration and co-location methods. Additional benefits include scale separation for source attribution and emission ratio evaluation which are discussed elsewhere in this document.

The BL project has allowed the performance of the cloud calibration method to be evaluated (see elsewhere in this document). This confirms that the calibration concept works, at least in the London context, on a par with physical co-location methods for NO_2 and $PM_{2.5}$. A simpler variant of the calibration method has also been applied to NO, again in the London context, with success. There are several caveats to this which are discussed below.

There are key steps towards replicability of the cloud calibration methodology to other locations:

1) <u>High time resolution pollutant measurements</u>

Fundamental to the application of the calibration method is the scale separation of local emissions from those which are non-local and regional in origin. As is discussed above and elsewhere in this document, the key to achieving this is that the instruments used must be capable of making

pollutant measurements at a suitably fast rate (in this case one-minute resolution for the static measurements) at each measurement site.

This is a key prerequisite for replicability of the calibration method to other locations.

2) Chemical and Emissions Environment

While frequently above WHO health thresholds for some species, Western Europe, within which London falls, can be considered an intermediate pollutant level environment. In this context this means that spatially average concentrations of NO_X and O_3 are broadly similar. The upshot is that spatially homogeneous background fields are generated for NO_2 by the combined chemical and meteorological processes in play, which means that the cloud calibration methodology for NO_2 works effectively, as detailed elsewhere in this document. However, the same processes mean that NO does not develop a homogeneous background field, so that while a zero offset for NO can be obtained, the gains of the NO sensors cannot and must be derived by alternative means.

For other environments where NOx levels are higher, the expectation is that a homogeneous background field can be obtained for NO, allowing "full" calibration (gain and offset) for NO sensors.

Similarly, the work within this project has demonstrated that $PM_{2.5}$ (and PM_{10}) sensors can be calibrated very effectively using the cloud method because of the relative dominance of long-range transport to observed PM values, this has again been tested in relatively unpolluted western European environments, and there may be regions where local emissions represent a significant larger fraction of the total, so that adaptations of the algorithm would be necessary for it to be applied effectively.

The upshot is that chemical and emissions environment specific to different regions must be taken into account when tailoring the calibration algorithm to optimise its performance in different regions.

3) Sensor Characteristics and Local Environment

Subject to the comments made above, and subject to the requirement that sensors used must have sufficient sensitivity to detect the pollutants levels expected to be observed in a specified region, the cloud calibration method is sensor agnostic.

As is discussed elsewhere in this document, a significant cross interference for O_3 has been observed in the NO_2 sensor measurements in the BL network. Discussion of the methodology which has then been applied to correct for this cross interference is covered elsewhere in this document.

However, while the O_3 cross-interference is clearly a limitation of the NO_2 sensor used in this project, and it has added complexity to the analysis (see elsewhere), it should not be viewed as a fundamental limitation of the cloud calibration algorithm methodology.

It does raise the importance of the need to understand sensor limitations for different environments (e.g. high SO_2 environments not encountered during this project) and how they impact sensor performance.

Similarly, regional changes in PM composition which may impact particle hygroscopicity could impact elements of the PM cloud-based calibration algorithm and again consideration of this should be factored in when considering replicability.

The bottom line is that sensor characteristics and the local environment must be considered when selecting sensors for projects, while accounting for cross interferences, and adjustments to the core calibration algorithm must be considered when considering replicability.

References

1. Popoola, O. A., Carruthers, D., Lad, C., Bright, V. B., Mead, I.M., Stettler, M., Saffell, J. and Jones, R.L. (2018). The use of networks of low cost air quality sensors to quantify air quality in urban settings. Atmospheric Environment 194, 58-70. https://doi.org/10.1016/j.atmosenv.2018.09.030

2. Heimann, I., Bright, V.B., McLeod, M.W., Mead, M.I., Popoola, O.A.M., Stewart, G.B. and Jones, R.L. (2015). Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. Atmospheric Environment. 113, 10-19. https://doi.org/10.1016/j.atmosenv.2015.04.057