Spatial analysis of PM₂.₅ using a Concentration Similarity Index applied to air quality sensor networks

Rósín Byrne[1, 2], John. C. Wenger[1, 2], Stig Hellebust[1, 2].

[1]Centre for Research into Atmospheric Chemistry, School of Chemistry, University College Cork, Ireland.
5 [2]Environmental Research Institute, University College Cork, Ireland.

Correspondence to: Stig Hellebust (s.hellebust@ucc.ie)

Abstract. Air quality sensor (AQS) networks are useful for mapping PM$_{2.5}$ in urban environments, but quantitative assessment of the observed spatial and temporal variation is currently under-developed. This study introduces a new metric - the Concentration Similarity Index (CSI) - to facilitate a quantitative and time-averaged comparison of the
10 concentration-time profiles of PM$_{2.5}$ measured by each sensor within an air quality sensor network. Following development on a dataset with minimal unexplained variation and robust tests, the CSI function is ensured to represent an unbiased and fair depiction of the air quality variation within an area covered by a monitoring network. The measurement data is used to derive a CSI value for every combination of sensor pairs in the network, which can then be compared with others in the network, yielding valuable information on spatial variation in PM$_{2.5}$. This new method is applied to two separate AQS
15 networks in Dungarvan and Cork City, Ireland. Dungarvan yielded a lower mean CSI, indicating lower overall similarity between locations in the network, possibly due to the town's coastal location giving rise to higher variation within the network. In both networks, the average diurnal plots for each sensor exhibit an evening peak in PM$_{2.5}$ concentration due to emissions from residential solid fuel burning, however, there is considerable variation in the size of this peak. Clustering techniques applied to the CSI matrices identify two different location types in each network; locations in central or
20 residential areas which experience more pollution from sold fuel burning and locations on the edge of the urban areas which experience cleaner air. Furthermore, the examination of isolated data periods (January and May) indicates higher PM$_{2.5}$ levels during periods of increased residential solid fuel burning act as a major driver for greater differences (lower similarity indices) between locations in both networks. Additionally, the CSI method facilitates the assessment of the representativeness of the PM$_{2.5}$ measured at regulatory air quality monitoring locations with respect to population exposure,
25 showing here that location type is more important than physical proximity in terms of similarity assessment. Applying the CSI in this manner can allow for the placement of monitoring infrastructure to be optimised. The findings of this work underscore the influence of solid fuel combustion as a local contributor to PM$_{2.5}$ and the variation it can cause between the measurements at different monitoring locations in a network while also highlighting the importance of including wintertime PM data for accurate comparisons. The CSI method developed here could be a valuable tool for quantitative comparisons of
30 air quality within a monitoring network, offering insights for further regulatory monitoring and exposure assessments.

1 Introduction

Air pollution affects the environment, quality of life and is a major cause of premature death and disease (Lelieveld et al., 2015; Cesaroni et al., 2013; Raaschou-Nielsen et al., 2013; Pedersen et al., 2013). The category of air pollutant with the

35 largest impact on human mortality and health is fine particulate matter, i.e. atmospheric particles with an aerodynamic diameter of 2.5 micrometres or less ($PM_{2.5}$) (Pope et al., 2020; Samoli et al., 2013; Pope and Dockery, 2012). In many regions around the world, air quality monitoring and management have become critical endeavours to mitigate the detrimental effects of air pollution, and especially $PM_{2.5}$, on citizens and the environment.

Over the years, technological advances have provided valuable tools to enhance our understanding of air pollution, and low-

40 cost air quality sensors (AQS) are emerging as promising instruments for collecting real-time air quality data at an improved spatial and temporal resolutions (Kumar et al., 2015; Munir et al., 2019). When used in networks, air quality sensors offer immense potential for enhancing and supplementing regulatory monitoring and assessment (Malings et al., 2020). However, further work needs to be carried out to assess the effectiveness of sensor networks and how to make best use of the data for gaining further insights into air pollution within a locality. Careful consideration must be given to the quality of the data

45 provided by sensors and the requirement for calibration must be assessed (Diez et al., 2022). Recent studies have shown that the performance and calibration of a $PM_{2.5}$ sensor is dependent on the type of sensor and often on the measurement location, suggesting the need for site-specific and individual calibrations (Wang et al., 2015; Zamora et al., 2020; Kaur and Kelly, 2023; Sayahi et al., 2019). When these factors are considered and accounted for, AQS networks offer an unprecedented opportunity to gain further insights into the complex dynamics of air pollution in localised areas, such as urban

50 environments, industrial zones, and residential neighbourhoods (Hodoli et al., 2023; Heimann et al., 2015; Crawford et al., 2021; O'Regan et al., 2022; Frederickson et al., 2022).

Assessing the spatial variation of air quality is of paramount importance because air pollution is not homogenous and can exhibit significant variations across different areas even on a local scale (Kassomenos et al., 2014; Wang et al., 2018; Frederickson et al., 2023, 2022). The variability of air pollution can be influenced by a multitude of factors such as traffic

55 patterns, industrial activities, meteorological conditions, and local topography. Consequently, relying on single monitoring locations or limited data resolution can provide an incomplete picture and inadequate understanding of local air quality in a certain area (Li et al., 2019). Understanding these variations is crucial for targeted interventions and policy decisions aimed at improving air quality and safeguarding public health. Spatial analysis, facilitated by sensor networks allows for a more accurate and nuanced understanding of how air quality, and therefore exposure to pollution, varies across a population

60 centre.

In a recent study, we used data collected by a $PM_{2.5}$ sensor network in the city of Cork, Ireland, to estimate the contribution of local pollution sources as separate and distinct from regional or transported air pollution (Byrne et al., 2023). The results highlighted the very localised nature of $PM_{2.5}$ caused by residential solid fuel burning during winter, which is a significant

65 problem in many towns and cities in Ireland and elsewhere (Dall'osto et al., 2013; Ovadnevaite et al., 2021; Wenger et al., 2020; Kourtchev et al., 2011; Lin et al., 2018; Zhang et al., 2021; Lin et al., 2019).

In this work, we propose a new approach for assessing the spatial profile of air quality using an AQS network. The method yields a time-averaged concentration similarity index (CSI) for quantitative assessment of the similarity between the complete data series produced by different sensors within the network. The CSI is built on the premise that sensors exposed to similar ambient conditions and pollutant sources will produce comparable $PM_{2.5}$ temporal trends. Conversely, sensors

70 subject to different conditions might display divergent $PM_{2.5}$ concentration trends. The motivation for the development of an assessment method based on the temporal variation over an extended period is the realisation that the annual average is a poor representation of true population exposure, which is experienced from hour to hour and day to day. It is therefore not adequate to merely compare annual averages of $PM_{2.5}$ levels in different locations in order to compare the exposures to $PM_{2.5}$ experienced by the populations in the respective locations. This method aims to translate this idea into a quantifiable metric

75 by calculating the time-averaged degree of similarity between two sensor datasets. Method development and testing is performed on an AQS network in the town of Dungarvan in Ireland to identify areas that may be experiencing persistently elevated or very localised $PM_{2.5}$ pollution compared to others. Clustering techniques are used to group sensors based on the similarity of their $PM_{2.5}$ measurements. The CSI method is also retrospectively applied to the data collected in the Cork City network to investigate the transferability of the method between sensor networks and to explore any differences between the

80 locations.

2 Methodology

2.1 Data collection, preprocessing, and calibration

The collection, preprocessing, and calibration of the data collected by the $PM_{2.5}$ sensor networks in Dungarvan and Cork City was carried out using the Julia programming language (Bezanson et al., 2017). Since low-cost AQS are not of

85 regulatory standard, great care needs to be taken with quality assessment and quality control of the data. In particular, the degree to which changes or differences in $PM_{2.5}$ measurements between devices can be trusted needs to be considered. The methodology proposed here addresses these inherent issues to deliver an approach for assessing the spatial representativeness of any monitoring location, i.e. what is the extent of the geographical area that the location meaningfully represents, in air quality monitoring terms, and what environment type is it comparable to regardless of geographical

90 distance to the location?

2.1.1 Dungarvan $PM_{2.5}$ sensor network

The Dungarvan sensor network consisted of 18 solar powered Clarity Node-S devices (Clarity Movement Co., USA) which utilise the Plantower PM6003 sensor to measure $PM_{2.5}$ within the range 1-1000 $\mu g\ m^{-3}$ and at a resolution of 1 $\mu g\ m^{-3}$ (Clarity Movement Co., 2023; Node-S technical sheet, 2023). By default, the Node-S devices take measurements every

Atmospheric
Measurement
Techniques
Discussions

95   15 minutes, allowing sufficient data upload and battery sleep time in between sampling periods. However, this can be adjusted to higher or lower frequencies. The highest sampling frequency achievable during winter without significantly affecting the battery performance was 8 minutes.

The Clarity Node-S devices were typically attached to street light poles between 2 and 4 metres above the ground. The sensors were positioned in a range of different environments including urban background, residential, coastal, and roadside
100   locations (Figure S1). Many of these locations were a mix of the different environments. The majority of devices were operational from 1 November 2022 to 31 May 2023, however three devices (AP7, AY9N, AY93) were only deployed from 12 January 2023. Measurements were taken over a continuous period covering different meteorological seasons (mainly Winter and Spring/early Summer), thus ensuring temporal variations in $PM_{2.5}$ concentrations were captured comprehensively.

105   Prior to and after deployment in Dungarvan, the Clarity Node-S devices were co-located on the roof of the Ellen Hutchins Building, University College Cork (51.895136, -8.516146) to compare their performance. Details of the three co-location periods are outlined in Table S1. Although some devices were not available for all three co-location periods, the three periods combined provide a comparison between the sensors across different seasons. This co-location dataset enabled the CSI method to be developed on measurements that in theory should be equal and the function could then be modified, if
110   necessary, to allow for sensor behaviour, uncertainties, errors, and potential limitations.

The raw sensor data from the co-location periods and field deployment, underwent a series of preprocessing steps to mitigate potential sources of error in the measurement and ensure data quality and consistency. Data points outside of the operational range of the sensors (> 1000 µg m^{-3}) were identified and removed, although instances of these were minimal. The 8-minute data were averaged to produce hourly measurements. Missing data points could potentially affect the temporal continuity of
115   the data; however, the data coverage was overall very good for the co-location and measurement campaign periods. On average, the devices had an hourly measurement coverage of 87 % for the field measurement campaign. This corresponds to an average of 4443 hourly measurements per device for the campaign period.

Assessing the consistency of measurements across the sensor network was paramount. Although the $PM_{2.5}$ readings were very well correlated when the devices were co-located (Table S2), a data harmonisation procedure was performed to ensure
120   the uniformity of sensor measurements, which is a prerequisite for the subsequent development of the Concentration Similarity Index. Since there was no reference-grade $PM_{2.5}$ data available during the co-location periods, the $PM_{2.5}$ concentrations from each sensor were scaled to a common reference point, represented by the mean of all data points across the whole co-location dataset (Figure S2). The data series for each sensor was then individually compared with the calculated mean dataset and subsequently harmonised to the common reference point using a simple linear regression
125   approach. The equations resulting from this harmonisation procedure were applied to the measurements collected from all devices during the subsequent field measurement campaign. While this procedure did not convert the measured $PM_{2.5}$ to reference-equivalent concentrations, it minimised sensor output variability and facilitated a more equitable comparison between sensor measurements (Table S2).

2.1.2 Cork City PM$_{2.5}$ sensor network

130 The Cork City sensor network consisted of 16 PurpleAir PA-II-SD units which each contain two Plantower PMS5003 sensors to measure PM$_{2.5}$ within the effective range 0-500 μg m^{-3}, with a maximum range of 1000 μg m^{-3}, and at a resolution of 1 μg m^{-3} (PMS5003 series data manual, 2022). In this study, data recorded by the devices in the network for the periods 01 January 2021 to 31 May 2021 and 1 September 2021 to 31 December 2021 were collated and analysed. However, four devices were found to have limited data capture for the specified periods (< 50 %) and were therefore omitted from the

135 analysis. The 12 sensors used in this analysis had an average data capture of 85 % for the specified periods; their locations are shown in Fig. S3.

Due to logistical constraints, it was not possible to co-locate all of the PurpleAir devices together to assess variability in PM$_{2.5}$ concentrations. However, low inter-sensor and inter-unit variability was exhibited by four co-located PurpleAir devices in our previous study on the Cork City network (Byrne et al., 2023). Moreover, the PM$_{2.5}$ concentrations measured

140 using the four PurpleAir devices were highly correlated with hourly values of PM$_{2.5}$ obtained using a Met-One (USA) Beta-Attenuation Monitor (BAM-1020) and a co-location dataset was used to derive calibration factors incorporating the effects of temperature and relative humidity. The data processing procedures for obtaining the PM$_{2.5}$ concentrations reported here are identical to those reported by Byrne et al. (2023).

The Cork City dataset spans a similar measurement period to the Dungarvan dataset to allow for comparable results due to

145 the known seasonality of PM$_{2.5}$ pollution in Ireland (Ovadnevaite et al., 2021). Although the year 2021 included some periods of COVID-19 pandemic restrictions, such measures mainly affected NO$_2$ concentrations and were not shown to have a significant impact on PM levels in Ireland (Environmental Protection Agency (EPA), 2020).

2.2 Development of the Concentration Similarity Index

The Concentration Similarity Index (CSI) derived here quantifies the degree of likeness between PM$_{2.5}$ concentration profiles

150 from two sensors for a defined period of time and forms the basis for assessing the spatial disparities in PM$_{2.5}$ measurements within sensor networks. The methodology proposed was developed through multiple iterations in order to adjust and improve the procedure. An overview of the development is described, showing the evolution towards the final method.

2.2.1 Original function application

The first phase of development was based directly on the work carried out by Piersanti et al. (2015), who used a

155 concentration similarity function to assess the spatial representativeness of PM$_{2.5}$ and O$_3$ monitoring stations in the Italian air quality monitoring network. By comparing point measurements to a dataset of modelled hourly air pollutant data covering Italy with a 4 × 4 km^2 grid cell resolution, Piersanti et al. (2015) produced maps showing how representative certain sites in the Italian monitoring infrastructure were. The application proposed here compares point measurement to point measurement as opposed to point measurement to modelled grid cell data, however the underlying principle of comparing two

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

160   concentration-time profiles to produce a single indication of similarity between them still applies. The function value

$f_{site}(x, y)$ used by Piersanti et al. (2015) to assess the spatial coverage of point measurements is given in Eq. (1):

$$f_{site}(x,y) = \frac{\sum_{i=1}^{N_t} flag}{N_t}, where\ flag = \begin{cases} 1, & \dfrac{|C(X_{site}, Y_{site}, t_i) - C(x, y, t_i)|}{C(X_{site}, Y_{site}, t_i)} < 0.2 \\[3mm] 0, & \dfrac{|C(X_{site}, Y_{site}, t_i) - C(x, y, t_i)|}{C(X_{site}, Y_{site}, t_i)} > 0.2 \end{cases}$$

(1)

165   Where, $C(x, y, t_i)$ represents the surface concentration from the modelled data at time t_i, $C(X_{site}, Y_{site}, t_i)$ represents the point

measurement of a specific monitoring site at time t_i, and N_t is the total number of time steps. The study defined a point

measurement as representative of a grid cell area if the condition $f_{site}(x, y) > 0.9$ is true.

In the first step of our approach, this function was applied to the hourly average PM$_{2.5}$ data obtained from the co-located

Clarity Node-S units by comparing two sensor data series at a time, with the reference and modelled concentration inputs

170   substituted for sensor PM concentration values, $C(A, t_i)$ and $C(B, t_i)$. Over a total of 1565 co-located hours, the mean number

of comparable data points per $C(A, t_i)$, $C(B, t_i)$ pair was 654, due to devices being present at different stages during the co-

location periods (Table S1).

It might be expected that the function value comparing two sensor data series would be 1, given that the measurements were

collected in the same location and were known to all represent the same air parcel at each point in time. However, it was

175   found that the function was not comprehensive enough to allow for an acceptable comparison of the sensor data. While in

theory, the results for all sensor pairs should be 1, the results showed discrepancies between some device pairs, because the

function value deviated significantly from 1 in many cases (Table 1) and was as low as 0.51 in some cases, with an overall

mean of 0.82.

180

Table 1: Function values, $f_{site}(x, y)$, for hourly averaged PM$_{2.5}$ measured by a range of co-located Clarity Node-S devices. Device labels in the columns were set as $C(X_{site}, Y_{site}, t)$ and device labels in the rows were set at $C(x, y, t_i)$.

	A3	A4	A8H	A8Z	A9	AQ	AZ	A7	A6P	AJ3	AP7	AQV	ARF	AW6	AWF	AY9N	AY93	AYG
A3	1	0.83	0.88	0.8	0.8	0.86	0.88	0.87	0.87	0.99	0.67	0.9	0.8	0.99	0.8	0.61	0.8	0.98
A4	0.84	1	0.81	0.89	0.83	0.87	0.89	0.87	0.87	0.98	0.8	0.9	0.9	0.97	0.85	0.72	0.79	0.99
A8H	0.86	0.79	1	0.79	0.78	0.87	0.89	0.83	0.78	0.76	0.66	0.85	0.72	0.92	0.71	0.62	0.72	0.73
A8Z	0.82	0.91	0.81	1	0.78	0.88	0.89	0.85	0.84	0.97	0.87	0.88	0.9	0.97	0.86	0.81	0.73	0.98
A9	0.76	0.82	0.8	0.76	1	0.75	0.8	0.78	0.78	0.78	0.69	0.77	0.74	0.67	0.78	0.62	0.76	0.69
AQ	0.88	0.87	0.9	0.87	0.77	1	0.94	0.88	0.86	0.97	0.87	0.97	0.84	1	0.81	0.81	0.75	0.94
AZ	0.89	0.88	0.89	0.88	0.8	0.94	1	0.91	0.9	0.97	0.8	0.97	0.82	0.99	0.82	0.75	0.84	0.93
A7	0.87	0.87	0.86	0.83	0.82	0.88	0.91	1	0.89	0.98	0.68	0.91	0.81	0.99	0.8	0.63	0.74	0.96
A6P	0.87	0.86	0.79	0.82	0.77	0.87	0.92	0.89	1	0.89	0.77	0.9	0.83	0.75	0.82	0.8	0.78	0.89
AJ3	0.99	0.97	0.76	0.97	0.75	0.97	0.98	0.98	0.89	1	0.72	0.88	0.91	0.76	0.82	0.77	0.81	0.89
AP7	0.71	0.85	0.63	0.88	0.65	0.86	0.86	0.74	0.77	0.71	1	0.75	0.81	0.54	0.8	0.83	0.75	0.8
AQV	0.9	0.89	0.86	0.86	0.75	0.96	0.96	0.91	0.88	0.88	0.77	1	0.85	0.77	0.8	0.76	0.82	0.85
ARF	0.82	0.9	0.72	0.91	0.74	0.86	0.85	0.83	0.85	0.9	0.8	0.84	1	0.74	0.82	0.81	0.81	0.92
AW6	0.96	0.94	0.92	0.94	0.65	1	0.99	0.98	0.72	0.72	0.5	0.73	0.7	1	0.69	0.51	0.49	0.7
AWF	0.8	0.88	0.7	0.86	0.76	0.82	0.82	0.81	0.82	0.82	0.81	0.8	0.83	0.73	1	0.81	0.77	0.86
AY9N	0.62	0.76	0.59	0.81	0.6	0.79	0.75	0.65	0.78	0.77	0.81	0.74	0.8	0.55	0.8	1	0.72	0.8
AY93	0.76	0.76	0.67	0.65	0.72	0.69	0.74	0.65	0.78	0.84	0.75	0.8	0.81	0.55	0.77	0.75	1	0.82
AYG	0.99	0.99	0.72	0.99	0.67	0.95	0.95	0.98	0.87	0.87	0.78	0.83	0.9	0.75	0.84	0.79	0.76	1

2.2.2 Function parameter optimisation and introduction of PM limit

Analysis of the results obtained from direct application of the original function showed that the conditions set out by it were
185 too strict to apply to the sensor data given the variations that can occur in AQS measurements, especially bearing in mind that the entire sensor network could be within the original single grid cell size analysed by Piersanti et al. (2015). Moreover, the high hourly PM$_{2.5}$ variation and very localised effects exhibited in a typical Irish winter PM$_{2.5}$ profile is not suited to the original function (Byrne et al., 2023). Thus, a PM mass concentration limit, PM_{lim}, was introduced to the function, with different relative concentration limits for the upper and lower PM values, $C_{lim, upper}$ and $C_{lim, lower}$, respectively. Treating larger
190 and smaller PM$_{2.5}$ values when assessing the similarity between two data series is useful in capturing the nuanced relationships and patterns in the data. It allows for the real-world significance of the data to be reflected, acknowledging the varying implications of PM$_{2.5}$ measurements based on the magnitude. Higher PM$_{2.5}$ values can indicate a pollution episode or specific local pollution sources, while lower values can represent background levels. Therefore, treating lower PM$_{2.5}$ values

Atmospheric
Measurement
Techniques
Discussions

with more leniency in the similarity assessment recognises that minor fluctuations in low hourly concentrations might not be

195   as concerning as similar deviations in higher concentrations and the health-related considerations associated with these high

concentrations.

Another potential advantage of the PM limit concerns the varying degrees of accuracy of the AQS measurements. Allowing the leeway introduced here in assessing the similarity of lesser measurement values considers potential measurement uncertainties with these devices. However, it is important to note that this approach is not accommodating sensor limitations

200   at the expense of accuracy but rather it is a strategy to ensure that the assessment remains faithful to the underlying air quality dynamics while accounting for the potential deficiencies in measurement equipment.

The differentiation between higher and lower PM values in the concentration similarity assessment is a strategic choice which acknowledges the complexity of $PM_{2.5}$ data, the varying significance of concentration levels, and the limitations of sensors. It allows for a more accurate representation of similarities while considering real world implications and

205   measurement uncertainties and minimises the potential biases that could arise from an indiscriminate approach, thus ensuring an impartial and unbiased evaluation.

When the function is applied to a pair of sensors, the resulting CSI can differ slightly depending on which sensor was classified as sensor A or sensor B in Equation (1) when computing the difference at each timestep. Due to the nature of the function, the denominator value of the relative difference calculation, the concentration of sensor A at a given timestep, is

210   what makes the difference. To counteract this and to avoid the possibility of large discrepancies between the CSI values for a sensor pair depending on which sensor is taken as A or B, the function was modified to have the geometric mean, or the square root of the product, of $C(A, t_i)$ and $C(B, t_i)$ used as the denominator. This ensured symmetry in the function so that the CSI values were identical regardless of which sensor was classified as A or B in a sensor pair.

Equation (2) shows the next form of the concentration similarity function (function notation has been modified to be more

215   suitable for this application).

$$CSI_{A,B} = \frac{\sum_{i=1}^{N_t} f}{N_t}, where\ f = \begin{cases} 1\ if\ \dfrac{|C(B,t_i) - C(A,\ t_i)|}{\sqrt{C(A,\ t_i) \times C(B,t_i)}}\ < 0.2\ and\ C(A,t_i)\ or\ C(B,t_i) > 15\ \mu g\ m^{-3} \\[3mm] 0\ if\ \dfrac{|C(B,t_i) - C(A,\ t_i)|}{\sqrt{C(A,\ t_i) \times C(B,t_i)}}\ > 0.2\ and\ C(A,t_i)\ or\ C(B,t_i) < 15\ \mu g\ m^{-3} \\[3mm] 1\ if\ \dfrac{|C(B,t_i) - C(A,\ t_i)|}{\sqrt{C(A,\ t_i) \times C(B,t_i)}}\ < 0.7\ and\ C(A,t_i)\ or\ C(B,t_i) > 15\ \mu g\ m^{-3} \\[3mm] 0\ if\ \dfrac{|C(B,t_i) - C(A,\ t_i)|}{\sqrt{C(A,\ t_i) \times C(B,t_i)}}\ > 0.7\ and\ C(A,t_i)\ or\ C(B,t_i) < 15\ \mu g\ m^{-3} \end{cases}$$

(2)

Where $C(A, t_i)$ and $C(B, t_i)$ are the $PM_{2.5}$ measurements from devices *A* and *B* at time, t_i.

Atmospheric
Measurement
Techniques
Discussions

2.2.3 Development and testing of the modified equation

220   The PM limit and associated concentration similarity limits introduced were chosen by iteratively testing the similarity function on the co-location data using different limits. Each co-located sensor pair was tested with different PM_{lim} values (5, 10, 15, 20 µg m^{-3}) and with C_{lim} values ranging from 0.1 to 2.0 in steps of 0.1 for both the upper and lower limits. This produced a C_{lim} vs CSI comparison for each A-B pair for data above and below the corresponding PM_{lim} value. It was clear that larger PM value comparisons (> 15 µg m^{-3}) tended to produce higher CSI values than lower PM values as expected. The

225   C_{lim} value for each sensor comparison which gave a minimum CSI value of 0.95 was recorded with the overall mean of these C_{lim} values above and below each PM_{lim} value taken forward. The mean C_{lim} pair values were then applied to the co-location measurements with the respective PM_{lim} values to give final CSI values for each sensor pair, highlighting how the PM$_{2.5}$ concentration profile of each sensor compares to that of all the other sensors. The highest mean CSI value for all A-B pairs was found for $PM_{lim} = 15$ µg m^{-3}, $C_{lim, upper} = 0.2$, and $C_{lim, lower} = 0.7$. When applying these new limits, all sensor pairs gave

230   CSI > 0.85, with 99 % of pairs above 0.90 with an overall CSI mean of 0.98. These final limits enabled a good comparison for the hourly co-located AQS measurements (Table 2).

   The CSI function was also applied to data obtained from the four co-located PurpleAir devices in order to make sure that the function was applicable across the two AQS types. The data was harmonised by following the same procedure as the Clarity data, through scaling each data from each sensor to the mean data series of all four sensors. Although this co-location period was shorter than that of the Clarity dataset used for the function development, it still allowed for the CSI to be calculated

235   from around 250 common data points per sensor pair. All device pairs reported a CSI close to 1.0, with a mean CSI of 0.99 (Table S3).

240 **Table 2: Concentration Similarity Indices for hourly averaged PM$_{2.5}$ measured by a range of co-located Clarity Node-S devices.** PM_{lim} = 15 µg m^{-3}, $C_{lim, upper}$ = 0.2, $C_{lim, lower}$ = 0.7.

	A3	A4	A8H	A8Z	A9	AQ	AZ	A7	A6P	AJ3	AP7	AQV	ARF	AW6	AWF	AY9N	AY93	AYG
A3	1	0.97	0.99	0.96	0.92	0.99	1	0.99	0.99	1	0.96	1	0.97	1	0.96	0.94	0.99	1
A4		1	0.97	0.99	0.95	0.98	0.99	0.99	0.99	1	0.99	0.99	0.99	1	0.98	0.97	0.97	1
A8H			1	0.96	0.92	0.99	1	0.99	0.99	0.97	0.94	0.99	0.96	0.98	0.96	0.94	0.97	0.98
A8Z				1	0.94	0.97	0.99	0.98	0.98	1	0.98	0.98	1	1	0.98	0.96	0.97	1
A9					1	0.92	0.92	0.94	0.95	0.97	0.96	0.95	0.97	0.92	0.97	0.95	0.96	0.96
AQ						1	1	1	0.99	1	0.97	1	0.97	1	0.96	0.97	1	1
AZ							1	0.99	0.99	1	0.97	1	0.98	1	0.96	0.95	1	1
A7								1	0.99	1	0.95	0.99	0.98	1	0.96	0.94	0.97	1
A6P									1	0.99	0.97	0.99	0.99	0.96	0.97	0.97	0.99	1
AJ3										1	0.98	0.98	0.99	0.95	0.99	0.98	1	0.99
AP7											1	0.96	0.99	0.85	0.99	0.99	0.97	0.98
AQV												1	0.98	0.98	0.97	0.96	0.99	0.99
ARF													1	0.96	0.98	0.99	0.99	1
AW6														1	0.92	0.87	0.9	0.97
AWF															1	0.98	0.97	0.98
AY9N																1	0.97	0.99
AY93																	1	0.99
AYG																		1

The function described in Eq. 2 was further tested by comparing one sensor, A6P, to numerous sets of synthetic data created from that sensor's measurements to assess the impact of a range of scenarios. Comparing the A6P data to itself establishes a

245 baseline for the comparison where the CSI is 1 and any subsequent adjustments to the data to create the synthetic data can be explored, resulting in a new CSI. The first scenario investigated changes in CSI when outliers are present in the data. To explore this, the A6P data was changed so a certain number of data points could be considered outliers (n = 1, 10, 500, 1000). To classify a data point as an outlier, the selected data point was increased by 100 µg m^{-3} in order to ensure discrepancy between it and the original value. The function was then tested in a scenario where the mean remained constant

250 but the variance of the data was increased, and it was also tested in a scenario where the entire data was merely offset by 5, 10, 15, and 20 µg m^{-3}. The effects of these tests on the CSI results for A6P are shown in Table 3. It is clear that in the case of data with higher variability but the same overall mean, the CSI is impacted, because even when the variance is increased by just a factor of 1.5 the CSI is significantly reduced, indicating that such a dataset is dissimilar to the original. Offsetting the

Atmospheric
Measurement
Techniques
Discussions

data by different degrees also shows a major effect which means that such a dataset is deemed dissimilar by the method.
255   However, the method is quite robust with respect to outliers. As the method yields a time-averaged result, low numbers of
outliers do not hugely affect the index for a given sensor pair. So, two datasets that are generally similar, but where one
experiences some outliers, will be deemed similar by the method. The development and analysis of the similarity index
function in this way provided a basis for what to consider when applying the function to the field data.

Table 3: Influence of data outliers and other factors on CSI determined in test scenarios.

Number of outliers	CSI	Standard Deviation factor increase ($\mu g\ m^{-3}$)	CSI	$PM_{2.5}$ positive offset ($\mu g\ m^{-3}$)	CSI	$PM_{2.5}$ negative offset ($\mu g\ m^{-3}$)	CSI
0	1	0	1	0	1	0	1
1	1	1.5	0.52	5	0.54	5	0.34
10	1	2	0.29	10	0.05	10	0.02
500	0.89	4	0.10	15	0.01	15	0.01
1000	0.77			20	0.01	20	0.004

260   **2.3 Application to sensor networks and analysis of spatial trends**

The CSI methodology developed above was subsequently applied to the Dungarvan and Cork City sensor networks to
evaluate the similarity and spatial variations in $PM_{2.5}$. A systematic pairwise comparison approach was employed, wherein
each sensor was individually compared to every other sensor within the network.

Hierarchical clustering and fuzzy c-means (FCM) clustering were both performed on the CSI results, to identify groupings
265   based on each sensors' relationship to other sensors in the network which can then be reflected spatially. Cluster analysis is a
valuable unsupervised analysis technique used to identify natural groupings in a dataset by classifying the data into distinct
groups, or clusters, without needing pre classified or labelled data to train the algorithm. It systematically works to separate
the data by minimising within group variation and maximizing between group variation. Cluster analysis is often used in air
quality analysis, including describing pollution diurnal variation, identifying distinct diurnal patterns, pollution source
270   identification, and identifying spatial patterns in particle compositions (Austin et al., 2012; Flemming et al., 2005; Austin et
al., 2013).

Hierarchical clustering does not separate the data into a defined number of clusters in a single step, but rather consists of a
series of separations which typically goes from a single cluster containing all of the data, to n clusters each containing an
individual sample (when the data is an $n \times m$ matrix for n samples and m data points in each sample) (Everitt et al., 2011).
275   The procedure typically includes a dendrogram showing the tree-like structure of the nested clusters. This type of clustering
gives an advantage over partition-based algorithms, whereby the user is not required to specify the number of clusters.

Atmospheric
Measurement
Techniques
Discussions
EGU
Open Access

Fuzzy c-means clustering is an example of a partitional clustering technique, where the number of clusters must be predefined. However, another distinctive feature separating it from hierarchical clustering is that it is a soft clustering method. In hard (i.e. non-fuzzy) clustering, each point belongs exclusively to a single cluster, whereas in soft clustering, the

280 output is a membership score or probability likelihood of a data point belonging to each of the pre-defined clusters (Gentle et al., 1991). The assignment of a member to a group is a distribution over all available clusters. The partition that gives the closest hard clustering to the fuzzy output can be obtained by assigning each object to the cluster in which it has the largest membership score. However, the information achieved with soft clustering can be particularly useful when dealing with datasets exhibiting overlapping patterns or uncertainties in classification as opposed to directly partitioning into hard clusters

285 (Gentle et al., 1991).

With both clustering techniques, the quality of cluster assignments can be assessed with various evaluation metrics to choose the optimal number of clusters. As the "true" cluster classifications are not known here, validation must be performed using the clustering algorithm itself. To assess the quality of the hierarchical clustering assignments, the Silhouette metric was used along with the Calinski-Harabasz index to assess the FCM assignments (Caliñski and Harabasz, 1974; Rousseeuw,

290 1987). The Silhouette score, ranging from -1 to +1, can be calculated for each member of a cluster and then the mean Silhouette score from all members indicates an overall assignment quality for members of that cluster, with a high score closer to 1 indicating higher quality clusters, and a low or negative score indicating poorer cluster assignments. The Calinski-Harabasz index also quantifies the quality of cluster assignments with higher scores indicating better quality. The metrics were used to test for the optimal number of clusters for each algorithm.

295 Both clustering approaches were selected to provide further understanding of the inherent spatial structures concealed within the CSI results. Hierarchical clustering offers the hierarchical representation of clusters, aiding in the identification of nested relationships, while FCM allows for a more flexible approach to the cluster assignments when the number of clusters is not known a priori.

3. Results and Discussion

300 **3.1 Dungarvan PM$_{2.5}$ sensor network**

Analysis of the harmonised data obtained from the sensors in the Dungarvan PM$_{2.5}$ network was conducted to determine CSI values and assess the spatial variation of air pollution across the town. Although the PM$_{2.5}$ concentrations are not as accurate as those collected by reference instrumentation, any relative differences between the sensors and individual sensor data trends can be regarded as genuine due to the low inter-sensor variation observed after data harmonisation procedures.

305 The temporal and spatial trends of PM$_{2.5}$ across the Dungarvan sensor network are reflected in the average diurnal plots obtained for each sensor, Fig. 1. These diurnal profiles all show large evening peaks in PM$_{2.5}$, which are typical for towns and cities in Ireland affected by residential solid fuel burning during winter evenings (Healy et al., 2010; Dall'Osto et al., 2014; Wenger et al., 2020). However, there are clear disparities in some of the average evening peak values between the

sensors. One group of sensors has maximum values above 35 µg m^{-3} (A3, A4, A8H, A9, AQ, A7, AW6, AQV), while the

310 sensors with maxima below 35 µg m^{-3} can be further divided into three smaller groups. Sensors labelled AJ3, AWF, and AZ all have a maximum PM$_{2.5}$ concentration around 30 µg m^{-3}; sensors AY9N, AY93, ARF, A8Z, AYG, and A6P all have maxima in the 20-26 µg m^{-3} range, while AP7 has a significantly lower evening peak than all other devices.

Most sensors exhibited the diurnal maximum around the same time of day, between 18:00 and 20:00, however AP7 and ARF, showed a slightly delayed peak from 20:00 to 22:00. AP7 had the lowest peak concentration and did not exhibit the

315 sharp rise and subsequent decrease associated with evening solid fuel burning that the other sensors showed. AP7 was located on the south-western edge of the town and since the predominant wind direction is south westerly, did not experience as much local pollution as the other locations in the network.

Summary statistics obtained for the 18 sensors in the Dungarvan network are listed in Table 4. Unsurprisingly, most of the devices with diurnal maxima > 35 µg m^{-3} have the highest mean, median, and maximum values. Out of this subset of

320 devices, AQV has the lowest overall mean (15 µg m^{-3}), but still has a relatively high standard deviation (22 µg m^{-3}), indicating the PM$_{2.5}$ values tend to vary widely but are lower on average. This could be indicative of fluctuating particle concentrations, consistent with intermittent pollution sources such as residential solid fuel burning.
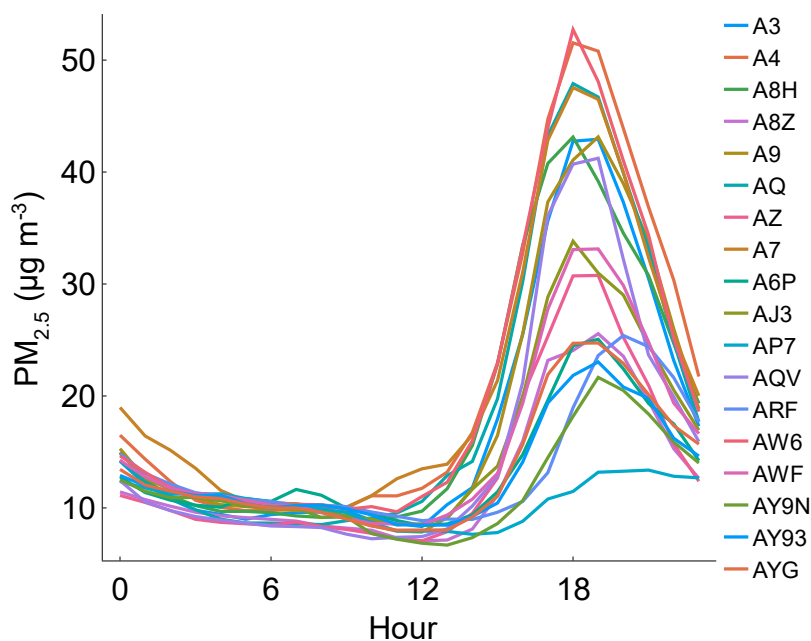
Figure 1: Diurnal profiles for hourly averaged measurements of PM$_{2.5}$ in the Dungarvan sensor network (September 2022 to May 2023).

Table 4: Summary statistics of hourly average PM$_{2.5}$ concentrations obtained for all sensors in the Dungarvan sensor network (September 2022 to May 2023).

	Mean	Median	Standard Deviation	Maximum hourly value	Maximum diurnal value	Hour of maximum diurnal value
ID	µg m^{-3}	µg m^{-3}	µg m^{-3}	µg m^{-3}	µg m^{-3}	
AP7	11	7	12	153	12	21
AY9N	12	7	14	136	23	19
A8Z	13	7	16	274	25	19
A6P	13	8	18	311	26	19
AY93	13	8	15	176	22	19
AZ	14	7	18	286	30	19
ARF	14	8	18	243	26	20
AYG	14	8	16	259	24	19
AQV	15	8	22	281	44	19
AJ3	16	9	19	270	33	18
AWF	16	9	17	189	31	19
A3	18	9	27	412	45	19
A9	18	9	27	409	45	19
A8H	19	9	28	482	40	18
AQ	19	10	28	480	48	18
A4	21	12	27	370	52	18
A7	21	12	27	361	45	18
AW6	21	11	26	319	51	18

3.1.1 Concentration Similarity Index

330   The matrix of CSI values obtained for the Dungarvan sensor network is shown in Table 5. The results can be analysed in a number of ways. Firstly, the indices for one sensor can be used to assess how similar or dissimilar the measurements are to all other sensors in the network, thus providing information on the spatial representativeness of that particular location. Secondly, the indices of all sensors can be looked at together to elucidate any potential relationship between sensor measurement locations.

335   The minimum CSI value (0.85) determined during the co-location deployment can act as the lower limit for when two sensor locations can be considered very similar. The reported CSI values for Dungarvan sensors ranged from 0.48 (ARF vs A7) to 0.79 (AYG vs AWF) with a mean of 0.61, indicating a significant difference in air quality representation between locations

Atmospheric
Measurement
Techniques
Discussions
Open Access
EGU

across the town. The device with the lowest mean of its CSI values with respect to the other locations was A4 (0.55), and although device ARF was only slightly above this (0.57), it reported a larger range of CSI values, including the lowest of the

340 entire dataset. AJ3, AQV, and AYG all shared the highest mean CSI values (0.66).

To further investigate the effect of solid fuel burning on local air quality, the CSI function was applied to data from two isolated months – January and May 2023. The purpose of this assessment was to evaluate the extent to which residential solid fuel burning dictates the CSI between two sensors, given that one month (January) will have higher $PM_{2.5}$ levels with measurements heavily influenced by solid fuel burning, and the other will not (May). For both months, all sensors had data

345 capture above 65 % and the mean capture was 94 % for January and 92 % for May. The January mean CSI from all comparisons was 0.51, and the May mean CSI was 0.84 (Table S4, Table S5). The large discrepancy between the mean CSI for January and May is most likely due to the higher variation typically seen in wintertime $PM_{2.5}$ due to residential solid fuel burning. This highlights the importance of seasonality when assessing the spatial representativeness of monitoring network locations.

350

Table 5: Concentration Similarity Indices for the hourly averaged PM$_{2.5}$ concentrations measured by Clarity Node-S devices in the Dungarvan sensor network.

	A3	A4	A8H	A8Z	A9	AQ	AZ	A7	A6P	AJ3	AP7	AQV	ARF	AW6	AWF	AY9N	AY93	AYG
A3	1	0.56	0.64	0.56	0.58	0.67	0.58	0.62	0.64	0.59	0.53	0.61	0.5	0.63	0.55	0.59	0.6	0.57
A4		1	0.53	0.52	0.55	0.53	0.56	0.58	0.53	0.57	0.55	0.58	0.49	0.54	0.55	0.56	0.55	0.55
A8H			1	0.58	0.6	0.61	0.59	0.57	0.61	0.61	0.57	0.64	0.54	0.61	0.56	0.61	0.62	0.61
A8Z				1	0.62	0.55	0.69	0.54	0.62	0.65	0.71	0.66	0.6	0.56	0.62	0.66	0.62	0.67
A9					1	0.53	0.6	0.57	0.58	0.63	0.6	0.63	0.52	0.58	0.6	0.61	0.58	0.6
AQ						1	0.58	0.57	0.6	0.59	0.5	0.62	0.49	0.63	0.53	0.56	0.58	0.56
AZ							1	0.58	0.63	0.68	0.68	0.71	0.6	0.62	0.66	0.62	0.61	0.74
A7								1	0.57	0.58	0.53	0.6	0.48	0.62	0.55	0.56	0.57	0.55
A6P									1	0.72	0.62	0.66	0.58	0.63	0.64	0.65	0.68	0.71
AJ3										1	0.64	0.76	0.6	0.64	0.76	0.67	0.7	0.78
AP7											1	0.64	0.67	0.52	0.64	0.65	0.63	0.69
AQV												1	0.59	0.65	0.72	0.65	0.68	0.77
ARF													1	0.56	0.59	0.62	0.61	0.63
AW6														1	0.61	0.6	0.62	0.62
AWF															1	0.66	0.68	0.79
AY9N																1	0.59	0.67
AY93																	1	0.72
AYG																		1

3.1.2 Clustering

Clustering techniques were employed on the CSI matrix to uncover any inherent spatial relationships between different locations in the network. Hierarchical clustering produced a dendrogram showing the hierarchical relationship between the sensor locations and was used to identify clusters (Figure 2). The highest mean Silhouette score was found with 2 clusters

(Figure S4). However, it was not a high Silhouette score (0.19), indicating that the quality of the cluster assignments was low. The highest Calinski-Harabasz index corresponded to the assignment of members to 2 clusters when applying the FCM clustering (Figure S5).

360   From both the dendrogram (Figure 2) and the FCM membership weights (Figure 3), it is clear that devices A4 through to AQ are grouped together in one cluster (Cluster 1), and devices AQV to AP7 are grouped in another cluster (Cluster 2). This split is very similar to the easily visualised groupings shown in the diurnal profile maxima (Figure 1), with the only difference being device AQV. The devices in Cluster 1 are also those with the highest mean $PM_{2.5}$ for the measurement period. The mean CSI for each sensor mostly corresponds to the cluster assignments, with Cluster 1 devices having a mean CSI equal to

365   or below 0.6, and all devices in Cluster 2 have a mean CSI above 0.6, except for device ARF. Interestingly, this grouping also appears to have spatial importance too, as shown in Fig. 4. Cluster 2 devices are mainly located around the edge of the town and generally experience cleaner air, while Cluster 1 devices are located in central and residential areas, which are more polluted during winter months.
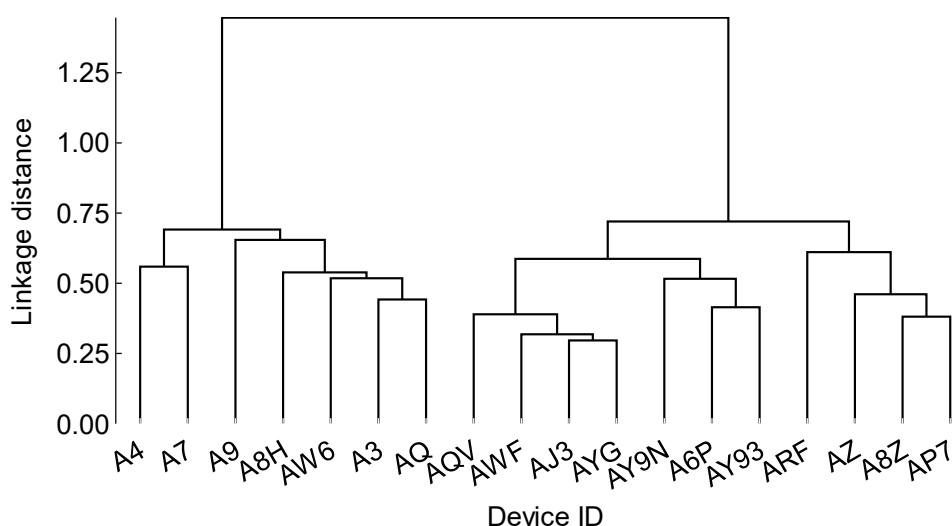
370   **Figure 2: Dendrogram output from hierarchical clustering of the CSI data from the Dungarvan sensor network.**

Atmospheric
Measurement
Techniques
Discussions
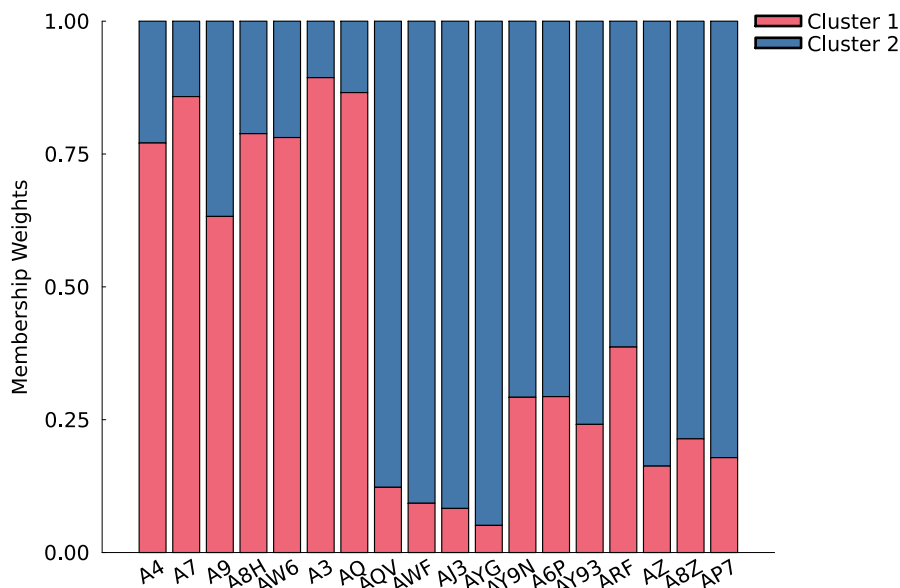
Figure 3: Membership weights from FCM clustering of the CSI data from the Dungarvan sensor network.
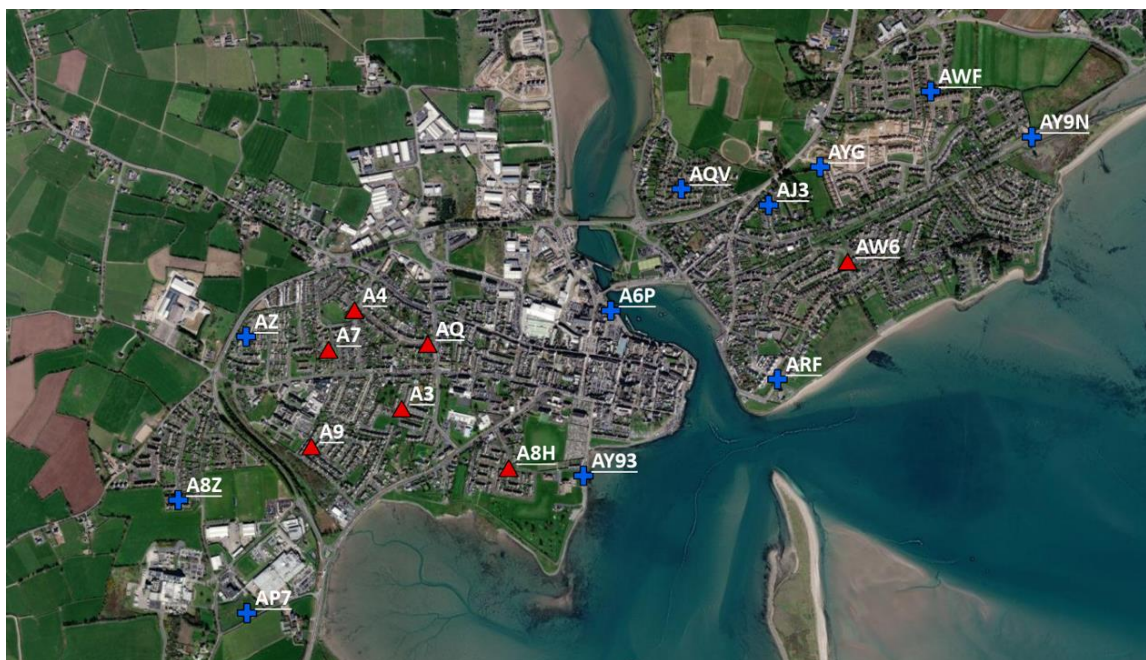
Figure 4: Dungarvan AQS locations with two cluster groups indicated. Cluster 1 devices (red triangle markers) are mainly located in central and residential areas, while cluster 2 devices (blue cross markers) are mainly located on the edge of the town. (Map obtained from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community)

3.2 Cork City PM$_{2.5}$ sensor network

380   The same approach was used to analyse the data collected by the Cork City AQS network. In this case, the corrected measurements are indicative of the actual PM$_{2.5}$ experienced in each location. The diurnal plots for each sensor in the Cork City network are similar to those observed in Dungarvan, with a sizeable evening peak in PM$_{2.5}$ concentrations (19:00-21:00) due to emissions from residential solid fuel burning. Again, there is considerable variation in the peak concentration of PM$_{2.5}$ (Figure 5). Device MTU showed the lowest diurnal average maximum of 9 µg m^{-3}. This device is located on the western side

385   of the city and has few upwind pollution sources contributing to air pollution at the location as the prevailing wind direction is from the South-West. Devices CCC12 and CCC9 both showed the highest diurnal average maximum, 17 µg m^{-3}. CCC12 is located northeast of the city, and so likely experiences urban PM$_{2.5}$ sources up-wind from it or has strong localised sources. Similarly, CCC9 is located to the east of the city, in a residential area. Table 6 contains summary statistics for each of the sensors in the Cork City network. Some devices had very high PM$_{2.5}$ maxima, e.g. 201 µg m^{-3} for CCC11, which were

390   more than double the maxima of other devices, e.g., CCC8 which had the lowest overall maximum of 47 µg m^{-3}. Device MTU had the lowest diurnal maximum value, indicating that this location is the least affected by local emissions from solid fuel burning. However, it measured a significant overall PM$_{2.5}$ maximum of 99 µg m^{-3} and significant spikes in pollution were occasionally observed, likely due to meteorological conditions or specific localised effects. When looking at all of the parameters listed in Table 6, CCC11 stands out. This sensor has the highest maximum hourly average PM$_{2.5}$ concentration in

395   the network, but the standard deviation (8 µg m^{-3}) is in the middle of the range, indicating that the location had relatively stable PM$_{2.5}$ levels throughout the measurement period with less variation than other devices but was still susceptible to occasional spikes in PM$_{2.5}$.
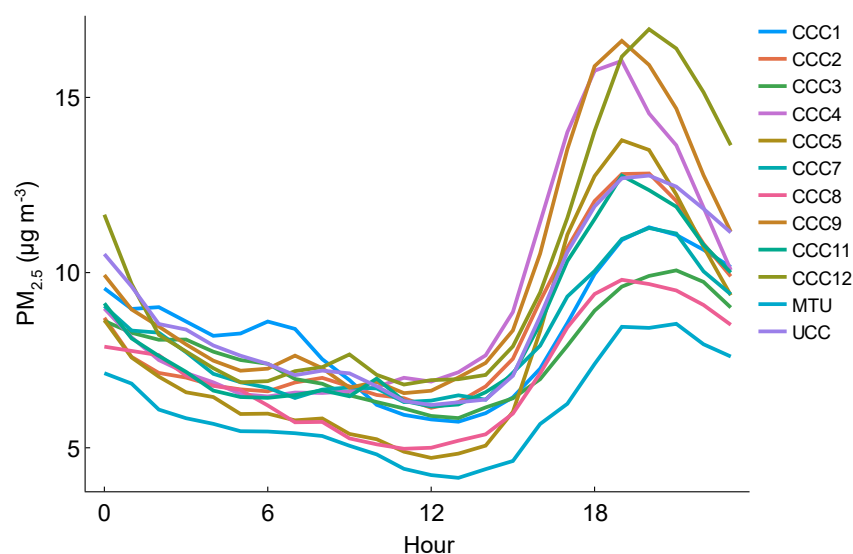

Figure 5: Diurnal PM$_{2.5}$ profiles for all AQS in the Cork City network (January to May and September to December 2021).

Atmospheric
Measurement
Techniques
Discussions

400   **Table 6: Summary statistics of hourly averaged PM2.5 obtained for all sensors in the Cork City network (January to May and September to December 2021).**

Device Label	Mean	Median	Standard Deviation	Maximum	Maximum diurnal value	Hour of maximum diurnal value
	µg m⁻³	µg m⁻³	µg m⁻³	µg m⁻³	µg m⁻³	
MTU	6	4	6	99	9	21
CCC8	7	5	6	47	10	19
CCC3	8	5	7	61	10	21
CCC5	8	5	10	181	14	19
CCC	8	6	7	71	11	20
CCC11	8	6	8	201	13	19
CCC1	8	6	8	92	11	20
CCC2	8	6	8	122	13	20
UCC	9	6	8	108	13	20
CCC4	9	7	8	97	16	19
CCC9	10	7	10	158	17	19
CCC12	10	7	10	117	17	20

3.2.1 Concentration Similarity Index

The matrix of CSI values obtained for the Cork City sensor network is shown in Table 7. The values range from 0.52 (CCC12 vs MTU and CCC9 vs MTU) to 0.85 (CCC2 vs CCC11) with a mean of 0.71. The high maximum CSI indicates a

405   high degree of similarity between those locations in the network, and overall, the Cork City locations show a higher degree of similarity compared to those in Dungarvan.

The isolated CSI results for the months of January and May 2021 were also assessed for Cork City. The average data coverage during both periods was 92 %. The mean CSI value in January (0.55) was considerably lower than that observed in May (0.82), Table S6, Table S7. This result is similar to that found for the Dungarvan network, again indicating that the

410   large difference in mean scores between the two months can be attributed to higher wintertime PM2.5 variation by residential solid fuel burning.

Table 7: Concentration Similarity Indices for the hourly averaged PM$_{2.5}$ concentrations measured by PurpleAir devices in the Cork City AQS network.

	CCC1	CCC2	CCC3	CCC4	CCC5	CCC7	CCC8	CCC9	CCC11	CCC12	MTU	UCC
CCC1	1	0.73	0.76	0.68	0.65	0.71	0.66	0.64	0.76	0.67	0.66	0.76
CCC2		1	0.79	0.82	0.65	0.77	0.68	0.73	0.85	0.78	0.61	0.79
CCC3			1	0.73	0.76	0.8	0.8	0.65	0.82	0.7	0.76	0.8
CCC4				1	0.63	0.73	0.64	0.76	0.82	0.78	0.56	0.77
CCC5					1	0.65	0.74	0.7	0.66	0.6	0.69	0.71
CCC7						1	0.68	0.65	0.78	0.7	0.66	0.73
CCC8							1	0.7	0.67	0.6	0.67	0.74
CCC9								1	0.72	0.74	0.52	0.8
CCC11									1	0.79	0.61	0.84
CCC12										1	0.52	0.77
MTU											1	0.62
UCC												1

3.2.2 Clustering

The two clustering algorithms were applied to investigate the CSI results of the Cork City network. The Silhouette Scores for each number of assigned clusters (2 to 5) were low, with 2 clusters showing the highest mean score (Figure S6). Similarly, with the FCM analysis, 2 clusters showed the highest score with the Calinski-Harabasz indices (Figure S7).

The dendrogram produced from the hierarchical clustering and the membership weights for 2 clusters from FCM clustering are shown in Fig. 6 and Fig. 7, respectively. It is clear that devices MTU, CCC5, and CCC8 are all grouped together in one branch, Cluster 2, with the remainder of the devices in the other branch. The one assignment difference between the two clustering methods is CCC3, which has a higher membership weight towards Cluster 2 with the FCM method but does not branch with that cluster in the dendrogram. However, its membership weight is close to 0.5. CCC1 also shows a membership weight close to 0.5, however it is showing a higher weight towards Cluster 1, as per the hierarchical clustering results. Devices in Cluster 2, except for CCC3, all have the lowest mean CSI value.

Similar to the Dungarvan results, there appears to be a spatial component to the cluster groupings, with devices in Cluster 2 being mainly on the western side of the city, Fig. 8. Interestingly, device CCC7, located in a commuter town on the western side of the city boundary, is grouped in Cluster 1, along with devices mainly in urban residential type sites, instead of being grouped with other devices on the western edge of the city. This indicates it has a more comparable CSI profile to the urban residential sites than the locations closer to it, further emphasising the importance of location type over physical proximity.

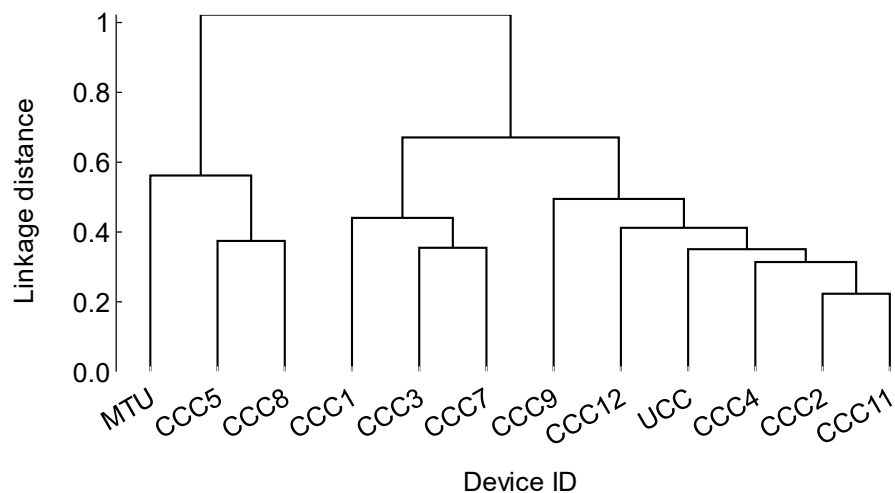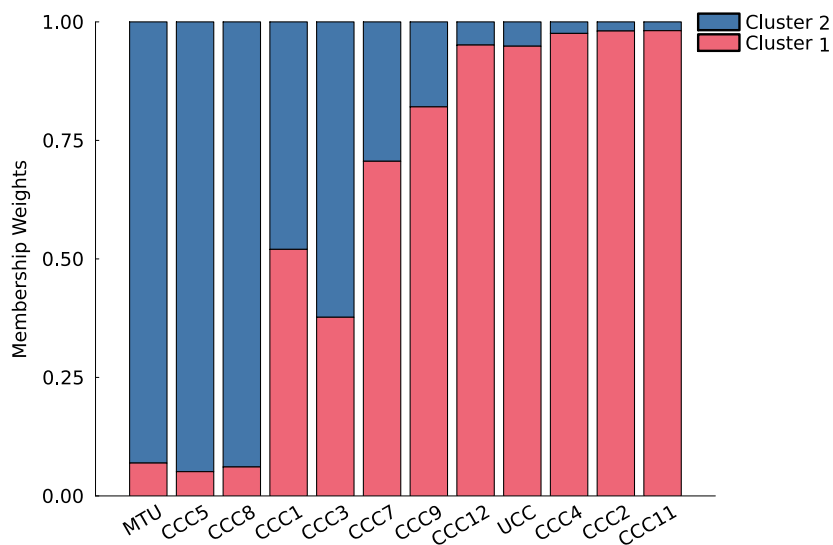**Figure 6: Dendrogram output from hierarchical clustering of the CSI data from the Cork sensor network.**

Figure 7: Membership weights from FCM clustering of the CSI data from the Cork sensor network.

435

Figure 8: Cock City AQS locations with 2 cluster groups indicated. Cluster 1 devices (red triangle markers) are located in the city centre and east/northeast, while Cluster 2 devices (yellow circle markers) are mainly located on the western side of the city. (Map obtained from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community)

440 **3.3 Application of the CSI to assess representativeness of air quality monitoring locations**

One key benefit of the CSI metric for AQS networks is that one sensor can be singled out and its overall degree of similarity to measurements from other locations can be determined. This analysis can be used to assess the spatial representativeness of a given location in the AQS network by quantitively exploring how similar its $PM_{2.5}$ profile is to other locations. If a network sensor is co-located with a reference instrument, then the CSI values for that sensor can be used to provide a measure of the

445 representativeness of the designated monitoring location and how well it informs the assessment of population exposure to air pollution.

In Dungarvan, the device A6P was co-located with a $PM_{2.5}$ instrument (Osiris, Turnkey) deployed as part of the national air quality monitoring network. The instrument is not a reference instrument but is certified to provide indicative measurements of $PM_{2.5}$ (National Ambient Air Quality Monitoring Network , 2023; Osiris, 2024). The similarity indices for A6P are

450 included in Table 5 and represented spatially in Fig. 9. All CSI values are below the minimum threshold of 0.85 for two Clarity S-node devices in the Dungarvan network to be considered very similar. The most similar devices are found to the north and south of this location, AQV and AY93. Interestingly, the similarity of PM profiles does not decrease with increasing distance from A6P. Devices on the furthest western (AZ, A8Z, AP7) and eastern (AWF, AY9N) edges of the

town are within 0.6 to 0.7, yet devices A4, A7, AQ, and A9 are all below 0.6 despite being physically closer to A6P. This
455   suggests that the location type is more important when it comes to assessing the similarity of locations within Irish towns as
opposed to physical proximity, as A4, A7, AQ, and A9 are all fully surrounded by residential areas, whereas the other
mentioned devices are in more open areas.

Figure 9: Dungarvan AQS locations with CSI results indicated in coloured circles (blue = lowest CSI, yellow = highest CSI), and
460   **A6P location indicated with red pin marker. (Map obtained from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX,**
Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community)

One of the devices in the Cork City senor network, UCC, was co-located alongside a reference instrument (BAM-1020) at
the national air quality monitoring location on UCC campus. The CSI values for device labelled UCC are shown in Fig. 10,
showing how similar the measurements at this site are compared to the rest of the locations in the sensor network. The CSI
465   scale on the map has been adjusted for these values. Similar to the Dungarvan case, there are devices which show among the
highest similarity (CCC4, CCC2, CCC12, CCC1) with UCC which are not located nearby.

24

Atmospheric
Measurement
Techniques
Discussions
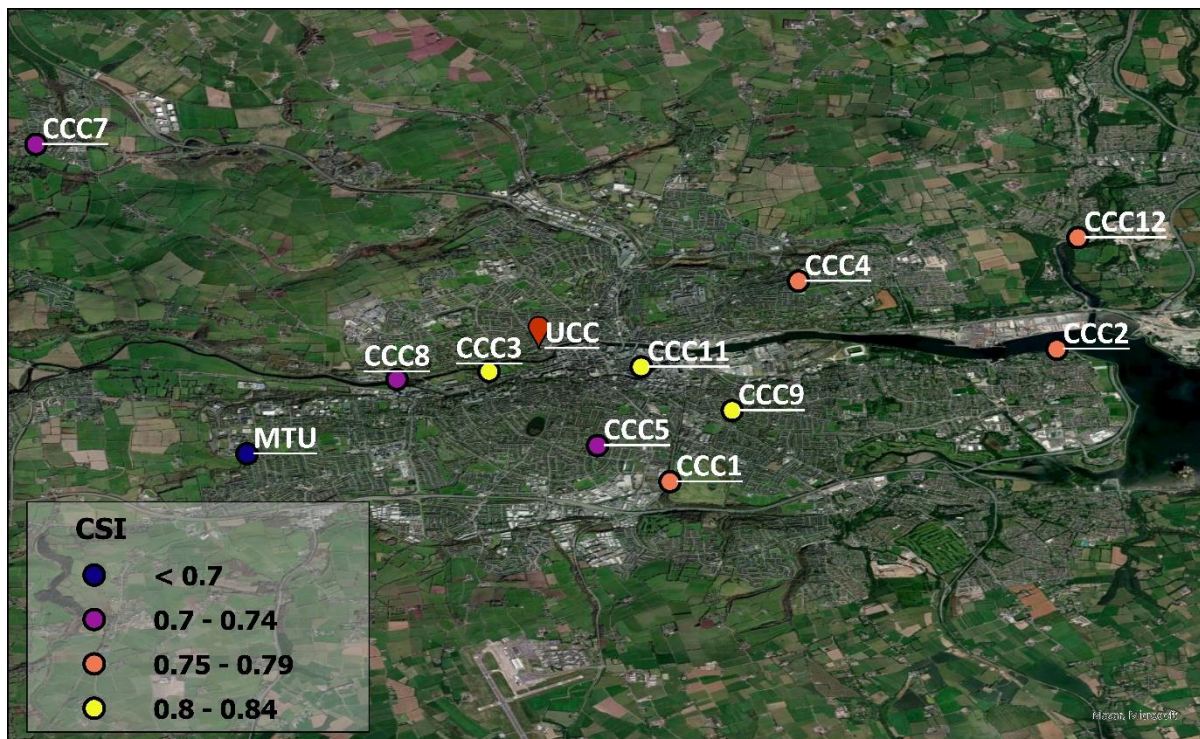
Figure 10: Cork AQS locations with UCC CSI results indicated in coloured circles (blue = lowest CSI, yellow = highest CSI), and UCC location indicated with red pin marker. (Map obtained from Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community)

4 Conclusion

A robust framework for comparing data series from individual air quality sensors in a network has been established and a new metric, the Concentration Similarity Index (CSI), has been developed, optimised, and tested on a co-location dataset. The CSI allows one to consider the monitoring network in terms of the similarity of the concentration-time profile of PM$_{2.5}$ at one location to those at the other locations in the same network. The harmonised dataset with minimal unexplained inter-sensor variation underpins the development of the CSI method, along with robust tests to ensure that the function represents an unbiased and fair depiction of the inter-sensor relationships after deployment in a monitoring network.

The CSI method has been used to analyse data generated by PM$_{2.5}$ sensor networks in two locations in Ireland, the coastal town of Dungarvan and the city of Cork. Clustering techniques are applied to the CSI matrix and comparable similarity trends between locations drives the distinctions made with the clustering algorithm. The resulting groupings can provide several insights into the PM$_{2.5}$ profile at each location, including the likelihood of similarity in pollution sources, spatial patterns, and temporal trends. An interesting contrast in the CSI results from the two monitoring networks was obtained from the clustering analysis. In Dungarvan, the locations generated clusters that were well reflected when comparing the individual diurnal profiles and specifically the diurnal maximum values, indicating that this factor has a major influence

470

475

480

25

485 when relating the concentration-time profiles at each location to one another in this network. However, for the Cork City network results, this was not as apparent. The clusters were not aligned based on diurnal peaks but rather the differentiating factor was more nuanced. Both clusters contained locations with a mix of higher and lower diurnal maxima and overall maxima. However, both network groupings reflect that devices may report dissimilar CSI results to other devices located nearby, and that considering location specifications or types, such as residential areas, is more important than physical

490 proximity when it comes to understanding and quantifying the similarities between locations.

The CSI function was also applied to two separate months in the network datasets, with January chosen to represent a period of higher $PM_{2.5}$ levels due to solid fuel burning emissions, and May chosen to represent a period with lower $PM_{2.5}$ concentrations due to reduced solid fuel burning. In both locations, the mean CSI for the network comparisons was higher in May than in January, indicating that higher $PM_{2.5}$ levels is a major driver for lower similarity indices between sensor

495 locations. Combining this with the findings of our previous study, we provide further evidence that high levels of localised $PM_{2.5}$ cause distinct disparities in exposure to poor air quality in different locations. Furthermore, to properly assess the burden of $PM_{2.5}$ experienced by a population and to accurately compare the measurements at two locations, the wintertime PM data must be included in the assessment.

The similarity of $PM_{2.5}$ measured at designated sites in the national air quality monitoring network compared to the rest of

500 locations in the sensor networks was analysed to give an estimation of the representativeness of the air pollution measured at the designated monitoring site. The national monitoring site location in Dungarvan was shown to be moderately representative of the other AQS network locations in the town, with CSI values ranging from 0.53 to 0.72. The CSI values for the Cork City comparison ranged from 0.62 to 0.84, also showing a fair representation of the air pollution experienced in the rest of the network. The CSI function was also tested via synthetic datasets which showed that a positive offset of just 5

505 µg m^{-3} resulted in almost halving the CSI, which was a lower CSI than most of the sensor comparisons in both network locations. So, while a CSI of 0.85 was used as a limit for two sensor measurement sets being very similar, CSI values between 0.6 and 0.7 are still moderately similar. In general, the CSI values in Cork City for the reference site comparison were higher (mean = 0.75) than that of Dungarvan (mean = 0.63), indicating less similarity between the reference site and devices in the Dungarvan network compared to Cork City. This analysis and application of the CSI function displays the

510 potential for AQS networks to be used in conjunction with a regulatory monitoring system. There is potential for the application of sensor networks to assess the need for more regulatory monitoring in an area, and to identify locations that are being poorly represented by the current system. Furthermore, the CSI method can be used to optimise a sensor network by carrying out short term sensor deployments and identifying areas of similarity or dissimilarity and thus assessing where the best locations for sensors are based on the similarity in exposure to air pollution.

515

Atmospheric
Measurement
Techniques
Discussions

Open Access

EGU

Data Availability

All raw data are available upon request.

Author Contributions

RB and SH conceptualised the project and the methodology. RB carried out the formal data analysis, investigation, visualisation, and wrote the manuscript with supervision and contributions from SH and JCW.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

Atmospheric
Measurement
Techniques
Discussions

References

Austin, E., Coull, B., Thomas, D., and Koutrakis, P.: A Framework for Identifying Distinct Multipollutant Profiles in Air Pollution Data, Environ Int, 45, 112, https://doi.org/10.1016/J.ENVINT.2012.04.003, 2012.

535 Austin, E., Coull, B. A., Zanobetti, A., and Koutrakis, P.: A framework to spatially cluster air pollution monitoring sites in US based on the PM2.5 composition, Environ Int, 59, 244–254, https://doi.org/10.1016/J.ENVINT.2013.06.003, 2013.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B.: Julia: A fresh approach to numerical computing, SIAM Review, 59, 65–98, https://doi.org/10.1137/141000671, 2017.

Byrne, R., Ryan, K., Venables, D. S., Wenger, J. C., and Hellebust, S.: Highly local sources and large spatial variations in

540 PM2.5 across a city: evidence from a city-wide sensor network in Cork, Ireland, Environmental Science: Atmospheres, 3, 919–930, https://doi.org/10.1039/D2EA00177B, 2023.

Caliñski, T. and Harabasz, J.: A Dendrite Method For Cluster Analysis, Communications in Statistics, 3, 1–27, https://doi.org/10.1080/03610927408827101, 1974.

Cesaroni, G., Badaloni, C., Gariazzo, C., Stafoggia, M., Sozzi, R., Davoli, M., and Forastiere, F.: Long-term exposure to

545 urban air pollution and mortality in a cohort of more than a million adults in Rome, Environ Health Perspect, 121, 324–331, https://doi.org/10.1289/EHP.1205862, 2013.

Clarity Movement Co.: https://www.clarity.io/, last access: 29 August 2023.

Clarity Movement Co. Node-S technical sheet: https://click.clarity.io/hubfs/Marketing%20Assets%20-%20PDFs/Product%20and%20Specification%20Sheets/Node%20S%20Specifications%20Sheet.pdf, last access: 29 August

550 2023.

Crawford, B., Hagan, D. H., Grossman, I., Cole, E., Holland, L., Heald, C. L., and Kroll, J. H.: Mapping pollution exposure and chemistry during an extreme air quality event (the 2018 Kīlauea eruption) using a low-cost sensor network, Proc Natl Acad Sci U S A, 118, e2025540118, https://doi.org/10.1073/pnas.2025540118, 2021.

Dall'osto, M., Ovadnevaite, J., Ceburnis, D., Martin, D., Healy, R. M., O'Connor, I. P., Kourtchev, I., Sodeau, J. R.,

555 Wenger, J. C., and O'Dowd, C.: Characterization of urban aerosol in Cork city (Ireland) using aerosol mass spectrometry, Atmos. Chem. Phys, 13, 4997–5015, https://doi.org/10.5194/acp-13-4997-2013, 2013.

Dall'Osto, M., Hellebust, S., Healy, R. M., Connor, I. P., Kourtchev, I., Sodeau, J. R., Ovadnevaite, J., Ceburnis, D., O'Dowd, C. D., and Wenger, J. C.: Apportionment of urban aerosol sources in Cork (Ireland) by synergistic measurement techniques, Science of The Total Environment, 493, 197–208, https://doi.org/10.1016/J.SCITOTENV.2014.05.027, 2014.

560 Diez, S., Lacy, S., Bannan, T. J., Flynn, M., Gardiner, T., Marsden, N., Martin, N., Read, K., and Edwards, P. M.: Air pollution measurement errors: Is your data fit for purpose?, Atmos Meas Tech, 4091–4105, https://doi.org/10.5194/amt-2022-58, 2022.

Environmental Protection Agency (EPA): Air Quality in Ireland 2020, 2020.

National Ambient Air Quality Monitoring Network : https://airquality.ie/, last access: 26 October 2023.

565   Everitt, B. S., Landau, S., Leese, M., and Stahl, D.: Cluster analysis: Fifth edition, Cluster Analysis: Fifth Edition, 1–330, https://doi.org/10.1002/9780470977811, 2011.

Flemming, J., Stern, R., and Yamartino, R. J.: A new air quality regime classification scheme for O3, NO2, SO2 and PM10 observations sites, Atmos Environ, 39, 6121–6129, https://doi.org/10.1016/J.ATMOSENV.2005.06.039, 2005.

Frederickson, L. B., Sidaraviciute, R., Schmidt, J. A., Hertel, O., and Johnson, M. S.: Are dense networks of low-cost nodes
570   really useful for monitoring air pollution? A case study in Staffordshire, Atmos Chem Phys, 22, 13949–13965, https://doi.org/10.5194/acp-22-13949-2022, 2022.

Frederickson, L. B., Russell, H. S., Fessa, D., Khan, J., Schmidt, J. A., Johnson, M. S., and Hertel, O.: Hyperlocal air pollution in an urban environment - measured with low-cost sensors, Urban Clim, 52, 101684, https://doi.org/10.1016/J.UCLIM.2023.101684, 2023.

575   Gentle, J. E., Kaufman, L., and Rousseuw, P. J.: Finding Groups in Data: An Introduction to Cluster Analysis., Biometrics, 47, 788, https://doi.org/10.2307/2532178, 1991.

Healy, R. M., Hellebust, S., Kourtchev, I., Allanic, A., O'Connor, I. P., Bell, J. M., Healy, D. A., Sodeau, J. R., and Wenger, J. C.: Source apportionment of PM2.5 in Cork Harbour, Ireland using a combination of single particle mass spectrometry and quantitative semi-continuous measurements, Atmos Chem Phys, 10, 9593–9613, https://doi.org/10.5194/ACP-10-9593-
580   2010, 2010.

Heimann, I., Bright, V. B., McLeod, M. W., Mead, M. I., Popoola, O. A. M., Stewart, G. B., and Jones, R. L.: Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors, Atmos Environ, 113, 10–19, https://doi.org/10.1016/J.ATMOSENV.2015.04.057, 2015.

Hodoli, C. G., Coulon, F., and Mead, M. I.: Source identification with high-temporal resolution data from low-cost sensors
585   using bivariate polar plots in urban areas of Ghana, Environmental Pollution, 317, 120448, https://doi.org/10.1016/J.ENVPOL.2022.120448, 2023.

Kassomenos, P. A., Vardoulakis, S., Chaloulakou, A., Paschalidou, A. K., Grivas, G., Borge, R., and Lumbreras, J.: Study of PM10 and PM2.5 levels in three European cities: Analysis of intra and inter urban variations, Atmos Environ, 87, 153–163, https://doi.org/10.1016/J.ATMOSENV.2014.01.004, 2014.

590   Kaur, K. and Kelly, K. E.: Performance evaluation of the Alphasense OPC-N3 and Plantower PMS5003 sensor in measuring dust events in the Salt Lake Valley, Utah, Atmos. Meas. Tech, 16, https://doi.org/10.5194/amt-16-2455-2023, 2023.

Kourtchev, I., Hellebust, S., Bell, J. M., O'Connor, I. P., Healy, R. M., Allanic, A., Healy, D., Wenger, J. C., and Sodeau, J. R.: The use of polar organic compounds to estimate the contribution of domestic solid fuel combustion and biogenic sources to ambient levels of organic carbon and PM2.5 in Cork Harbour, Ireland, Science of The Total Environment, 409, 2143–
595   2155, https://doi.org/10.1016/J.SCITOTENV.2011.02.027, 2011.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., and Britter, R.: The rise of low-cost sensing for managing air pollution in cities, Environ Int, 75, 199–205, https://doi.org/10.1016/J.ENVINT.2014.11.019, 2015.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to
600   premature mortality on a global scale, Nature 2015 525:7569, 525, 367–371, https://doi.org/10.1038/nature15371, 2015.

Li, H. Z., Gu, P., Ye, Q., Zimmerman, N., Robinson, E. S., Subramanian, R., Apte, J. S., Robinson, A. L., and Presto, A. A.:
Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant
monitors, Atmos Environ X, 2, 100012, https://doi.org/10.1016/J.AEAOA.2019.100012, 2019.

Lin, C., Huang, R. J., Ceburnis, D., Buckley, P., Preissler, J., Wenger, J., Rinaldi, M., Facchini, M. C., O'Dowd, C., and
605   Ovadnevaite, J.: Extreme air pollution from residential solid fuel burning, Nat Sustain, 1, 512–517,
https://doi.org/10.1038/s41893-018-0125-x, 2018.

Lin, C., Ceburnis, D., Huang, R. J., Xu, W., Spohn, T., Martin, D., Buckley, P., Wenger, J., Hellebust, S., Rinaldi, M.,
Cristina Facchini, M., O'Dowd, C., and Ovadnevaite, J.: Wintertime aerosol dominated by solid-fuel-burning emissions
across Ireland: Insight into the spatial and chemical variation in submicron aerosol, Atmos Chem Phys, 19, 14091–14106,
610   https://doi.org/10.5194/ACP-19-14091-2019, 2019.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A., and Subramanian, R.: Fine particle mass
monitoring with low-cost sensors: Corrections and long-term performance evaluation, Aerosol Science and Technology, 54,
160–174, https://doi.org/10.1080/02786826.2019.1623863, 2020.

Munir, S., Mayfield, M., Coca, D., Jubb, S. A., and Osammor, O.: Analysing the performance of low-cost air quality sensors,
615   their drivers, relative benefits and calibration in cities-a case study in Sheffield, Environ Monit Assess, 191,
https://doi.org/10.1007/S10661-019-7231-8, 2019.

O'Regan, A. C., Byrne, R., Hellebust, S., and Nyhan, M. M.: Associations between Google Street View-derived urban
greenspace metrics and air pollution measured using a distributed sensor network, Sustain Cities Soc, 87, 104221,
https://doi.org/10.1016/J.SCS.2022.104221, 2022.

620   Ovadnevaite, J., Lin, C., Rinaldi, M., Ceburnis, D., Buckley, P., Coleman, L., Facchini, M. C., Wenger, J., and O'Dowd, C.:
Air Pollution Sources in Ireland, https://www.epa.ie/publications/research/air/research-385-air-pollution-sources-in-
ireland.php, last access: 28 October 2024.

Pedersen, M., Giorgis-Allemand, L., Bernard, C., Aguilera, I., Andersen, A. M. N., Ballester, F., Beelen, R. M. J., Chatzi, L.,
Cirach, M., Danileviciute, A., Dedele, A., Eijsden, M. van, Estarlich, M., Fernández-Somoano, A., Fernández, M. F.,
625   Forastiere, F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., Hoogh, K. de, van den Hooven, E. H.,
Håberg, S. E., Jaddoe, V. W. V., Klümper, C., Korek, M., Krämer, U., Lerchundi, A., Lepeule, J., Nafstad, P., Nystad, W.,
Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Sunyer, J., Stephanou, E., Sørensen, M., Thiering, E.,
Tuffnell, D., Varró, M. J., Vrijkotte, T. G. M., Wijga, A., Wilhelm, M., Wright, J., Nieuwenhuijsen, M. J., Pershagen, G.,
Brunekreef, B., Kogevinas, M., and Slama, R.: Ambient air pollution and low birthweight: A European cohort study
630   (ESCAPE), Lancet Respir Med, 1, 695–704, https://doi.org/10.1016/S2213-2600(13)70192-9, 2013.

Piersanti, A., Vitali, L., Righini, G., Cremona, G., and Ciancarella, L.: Spatial representativeness of air quality monitoring
stations: A grid model based approach, Atmos Pollut Res, 6, 953–960, https://doi.org/10.1016/J.APR.2015.04.005, 2015.

Pope, C. A. and Dockery, D. W.: Health Effects of Fine Particulate Air Pollution: Lines that Connect, https://doi.org/10.1080/10473289.2006.10464485, 56, 709–742, https://doi.org/10.1080/10473289.2006.10464485, 2012.

635 Pope, C. A., Coleman, N., Pond, Z. A., and Burnett, R. T.: Fine particulate air pollution and human mortality: 25+ years of cohort studies, Environ Res, 183, 108924, https://doi.org/10.1016/J.ENVRES.2019.108924, 2020.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., Brunekreef, B., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P. E., Nafstad, P., De Faire, U., Pedersen, N. L., Östenson, C. G., Fratiglioni, L.,

640 Penell, J., Korek, M., Pershagen, G., Eriksen, K. T., Sørensen, M., Tjønneland, A., Ellermann, T., Eeftens, M., Peeters, P. H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Key, T. J., de Hoogh, K., Concin, H., Nagel, G., Vilier, A., Grioni, S., Krogh, V., Tsai, M. Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Trichopoulou, A., Bamia, C., Vineis, P., and Hoek, G.: Air pollution and lung cancer incidence in 17 European cohorts: Prospective analyses from the European Study of Cohorts for Air Pollution Effects

645 (ESCAPE), Lancet Oncol, 14, 813–822, https://doi.org/10.1016/S1470-2045(13)70279-1, 2013.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J Comput Appl Math, 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Samoli, E., Stafoggia, M., Rodopoulou, S., Ostro, B., Declercq, C., Alessandrini, E., Díaz, J., Karanasiou, A., Kelessis, A. G., Tertre, A. Le, Pandolfi, P., Randi, G., Scarinzi, C., Zauli-Sajani, S., Katsouyanni, K., Forastiere, F., Alessandrini, E.,

650 Angelini, P., Berti, G., Bisanti, L., Cadum, E., Catrambone, M., Chiusolo, M., Davoli, M., de' Donato, F., Demaria, M., Gandini, M., Grosa, M., Faustini, A., Ferrari, S., Forastiere, F., Pandolfi, P., Pelosini, R., Perrino, C., Pietrodangelo, A., Pizzi, L., Poluzzi, V., Priod, G., Randi, G., Ranzi, A., Rowinski, M., Scarinzi, C., Stivanello, E., Zauli-Sajani, S., Dimakopoulou, K., Elefteriadis, K., Katsouyanni, K., G.Kelessis, A., Maggos, T., Michalopoulos, N., Pateraki, S., Petrakakis, M., Sypsa, V., Agis, D., Alguacil, J., Artiñano, B., Barrera-Gómez, J., Basagaña, X., de la Rosa, J., Diaz, J.,

655 Fernandez, R., Jacquemin, B., Linares, C., Ostro, B., Pérez, N., Pey, J., Querol, X., Sanchez, A., Sunyer, J., Tobias, A., Bidondo, M., Declercq, C., Le Tertre, A., Lozano, P., Medina, S., Pascal, L., and Pascal, M.: Associations between fine and coarse particles and mortality in Mediterranean cities: Results from the MED-PARTICLES project, Environ Health Perspect, 121, 932–938, https://doi.org/10.1289/EHP.1206124, 2013.

Sayahi, T., Butterfield, A., and Kelly, K. E.: Long-term field evaluation of the Plantower PMS low-cost particulate matter

660 sensors, Environ Pollut, 245, 932–940, https://doi.org/10.1016/J.ENVPOL.2018.11.065, 2019.

Osiris: https://turnkey-instruments.com/product/osiris/, last access: 26 February 2024.

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., and Biswas, P.: Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement, Aerosol Science and Technology, 49, 1063–1077, https://doi.org/10.1080/02786826.2015.1100710, 2015.

665   Wang, Z., Zhong, S., He, H. di, Peng, Z. R., and Cai, M.: Fine-scale variations in PM2.5 and black carbon concentrations and corresponding influential factors at an urban road intersection, Build Environ, 141, 215–225, https://doi.org/10.1016/J.BUILDENV.2018.04.042, 2018.

Wenger, J., Arndt, J., Buckley, P., Hellebust, S., Mcgillicuddy, E., O'Connor, I., Sodeau, J., and Wilson, E.: Source Apportionment of Particulate Matter in Urban and Rural Residential Areas of Ireland (SAPPHIRE),

670   https://www.epa.ie/publications/research/air/research-318.php, last access: 28 October 2024.

Zamora, M. L., Rice, J., and Koehler, K.: One year evaluation of three low-cost PM2.5 monitors, Atmos Environ, 235, https://doi.org/10.1016/j.atmosenv.2020.117615, 2020.

Zhang, Y., Shi, Z., Wang, Y., Liu, L., Zhang, J., Li, J., Xia, Y., Ding, X., Liu, D., Kong, S., Niu, H., Fu, P., Zhang, X., and Li, W.: Fine particles from village air in northern China in winter: Large contribution of primary organic aerosols from

675   residential solid fuel burning, Environmental Pollution, 272, 116420, https://doi.org/10.1016/J.ENVPOL.2020.116420, 2021.

PMS5003 series data manual: https://aqicn.org/air/sensor/spec/pms5003-english-v2.3.pdf, last access: 2 February 2022.