European Journal of Geography

The publication of the European Journal of Geography (EJG) is based on the European Association of Geographers' goal to make European Geography a worldwide reference and standard. Thus, the scope of the EJG is to publish original and innovative papers that will substantially improve, in a theoretical, conceptual, or empirical way the quality of research, learning, teaching, and applying geography, as well as in promoting the significance of geography as a discipline. Submissions are encouraged to have a European dimension. The European Journal of Geography is a peer-reviewed open access journal and is published quarterly.

Received: 19/01/2026
Revised: 17/04/2026
Revised: 05/06/2026
Accepted: 07/06/2026
Published: 10/06/2026

Editor:
Dr Alexandros Bartzokas-Tsiompras

DOI: 10.48088/ejg.s.zaf.17.1.212.230
ISSN: 1792-1341
E-ISSN: 2410-7433

Research Article

Distinguishing Sensor Errors from Environmental Events: A Spatio-temporal Analysis of Outlier Detection During Wildfire Pollution in Athens

Sofia Zafeirelli [1]✉, Marios Batsaris [1], Olga Roussou [1], Javier Sigró [2,3] & Dimitris Kavroudakis [1]

[1] Department of Geography, University of the Aegean, Mytilene, Greece
[2] Centre for Climate Change (C3), Department of Geografia, Universitat Rovira i Virgili (URV), Tarragona, Spain
[3] University Institute for Research in Sustainability, Climate Change, and Energy Transition (IU-RESCAT), Universitat Rovira i Virgili (URV), Tarragona, Spain

✉ Correspondence: s.zafeirelli@aegean.gr

Abstract: Low-cost environmental sensors in smart cities play a critical role in monitoring the environment, offering real-time information for urban management. The reliability of smart sensors remains uncertain, since sensors may report outliers when malfunctioning, or due to anomalies in the environment or extreme occurrences, which might skew the analysis if not treated carefully. The study seeks to support the distinction between likely sensor anomalies and spatially coherent environmental events by comparing different outlier detection methods: Interquartile Range (IQR), Local Outlier, and Global Outlier. The IQR method identifies temporal outliers based on historical data, whereas Local and Global methods use the spatial dimension, calculating the deviations from the local and global averages, respectively. During wildfire incidents near Athens, Greece, in August 2021, the methods were applied on environmental data from the PurpleAir sensor platform, which measures $PM_{1.0}$, $PM_{2.5}$, $PM_{10.0}$, temperature and relative humidity. The IQR approach performed well in depicting short-term pollution peaks temporally associated with the wildfire period. The Local Outlier approach identifies a higher rate of local extreme values, thus suggesting sensitivity to localized environmental variability, while the Global Outlier method is more appropriate for widespread events.

Keywords: Outlier detection; environmental events; low-cost sensors; spatio-temporal data; wildfire; smart cities

Highlights:

- Evaluates outlier detection approaches applied to low-cost environmental sensor data in a wildfire pollution context.
- Three outlier detection methods are compared, IQR, Local Outlier and Global Outlier.
- Application of these methods to PurpleAir sensor data during August 2021 wildfires in Athens, Greece.
- IQR captured pollution-related spikes; Local Outlier indicated localized deviations; Global Outlier highlighted broader spatial patterns.

1. Introduction

Over the last decade, smart sensors have increasingly been used to record environmental conditions in cities. Most studies have been conducted using low-cost smart sensors due to their affordability, allowing broader spatial coverage and more detailed monitoring. These sensors, in addition to their affordability, can transmit their data in real time on various aspects of the urban environment, enhancing timely decision-making for environmental management. Thanks to increasingly integrated Internet of Things (IoT) networks, sensors can transmit and exchange data in real time, enabling more effective monitoring, analysis and control of urban systems and processes. In addition, community engagement, citizen science, and public awareness and participation in environmental issues are enhanced. The need for affordable and reliable solutions to monitor sensor data in many applications has increased research regarding the improvement of the accuracy of low-cost sensors. These sensors are often used as an alternative to traditional instruments for environmental monitoring, healthcare, smart agriculture and smart cities. Nevertheless, their precision and reliability are often questioned, increasing the research into their improvement (Munir et al., 2019; Slongo et al., 2024; Barkjohn et al., 2025; Hayward et al., 2025).

It is very important to monitor the environmental conditions at regular intervals in order to understand the spatio-temporal variations in urban environmental quality. The data collected from the sensors, apart from immediate monitoring and problem solving, can also be used for long-term policy making and urban planning. However, sensor data often contains outliers, which can skew analysis results if not handled appropriately, particularly due to environmental influences and sensor limitations (Hayward et al., 2025). In environmental monitoring, it is important to distinguish between outliers and events. Both categories have a variety of different approaches and definitions. Outliers are data values that deviate significantly from the rest of the data in a dataset, which can be either anomalies, errors, or very extreme conditions. Events are observations which indicate changes in their historical pattern due to their environment and usually last for a reasonably long period of time (Zhang et al., 2010). Spatiotemporal events in environmental sensor data occur when multiple extreme values appear, which are spatially close and in consecutive time intervals (Janeja et al., 2010). Both concepts involve deviations from the expected behavior of the data. However, sensor-related outliers are often isolated in space or time, whereas environmental events tend to exhibit spatial and temporal coherence across multiple observations (Janeja et al., 2010; Zhang et al., 2010). In previous research, various outlier detection methods were applied in environmental data (Poornima & Paramasivan, 2020; El-Shafeiy et al., 2023; Sánchez-Lasheras et al., 2020), as well as pollution event Functional Profiles (Papayiannis et al., 2023). They often focus on statistical measures (Zhang et al., 2012), spatial analysis (Liang et al., 2021), machine learning techniques (Anggraini et al., 2024) and temporal patterns (Graça et al., 2023) to identify outliers.

One of the most widely monitored environmental parameters is particulate matter (PM), which is commonly used as an indicator of air quality. For instance, Bulot et al. (2019), Levy Zamora et al. (2019), Zusman et al. (2020) and Tagle et al. (2020) have evaluated the performance of low-cost sensors, reporting moderate to high correlation with reference instruments, while also highlighting significant variability in accuracy depending on environmental conditions such as humidity, temperature, and particle composition. Miao et al. (2022) studied the spatiotemporal patterns of Air Quality Index (AQI) in the Yangtze River Delta, China, identifying significant spatial heterogeneity and highlighting $PM_{2.5}$ as the dominant pollutant across most metropolitan areas. Similarly, Keshtkar et al. (2022) studied changes in air pollution patterns and demonstrated significant spatial autocorrelation in AQI, as well as reductions in pollution levels during COVID-19 restrictions. Despite advances in sensor technology and sensor data analysis, there is a significant gap in accurately distinguishing between erroneous outliers and significant environmental events using real-time datasets. Many studies have explored outlier detection techniques in environmental monitoring (e.g., Poornima & Paramasivan, 2020), but few focused on spatio-temporal patterns in order to separate local outliers from global events.

This study seeks to contribute to addressing the gap of comparative assessments of temporal and spatial outlier detection techniques with high-frequency sensor observations in real-world environmental events, such as urban wildfire smoke events. Although there is extensive literature on low-cost air quality sensor networks and outlier detection, most studies focus either on improving sensor accuracy (Bulot et al., 2019; Levy Zamora et al., 2019; Zusman et al., 2020), or on applying a single outlier detection approach, or on using advanced machine learning techniques (Ottosen and Kumar, 2019; Zhang et al., 2012; Graça et al., 2023), often without clearly addressing the distinction between sensor-related errors and real environmental events. In addition, many existing approaches concentrate on a single pollutant or rely on long-term aggregated data, which limits their ability to capture short-lived and spatially variable extreme pollution episodes (Miao et al., 2022; Keshtkar et al., 2022).

In this study, the terms outliers, anomalies, and events are used in slightly different ways, even though they are often treated as interchangeable. Outliers represent the broadest category and refer to data points that fall well outside expected values. Some of these can be attributed to sensor-related issues, such as noise measurement or device malfunction, and are referred to here as anomalies. In contrast, events are understood as deviations that are spatially and/or temporally consistent, reflecting real environmental processes rather than sensor error. This distinction is applied consistently throughout the analysis, where isolated or inconsistent deviations are interpreted as potential anomalies, while spatio-temporally coherent patterns are treated as indications of environmental events.

The study is framed as a comparative evaluation of temporal and spatial statistical outlier detection methods: Interquartile Range (IQR), Local Outlier, and Global Outlier, across multiple environmental parameters ($PM_{1.0}$, $PM_{2.5}$, $PM_{10.0}$, temperature, and relative humidity) in a real urban wildfire pollution setting. Rather than focusing on a single detection technique, this study explores how different methods respond to localized versus widespread pollution patterns, supporting the interpretation of likely sensor anomalies and spatially coherent environmental events. The analysis emphasizes interpretability and spatio-temporal consistency rather than employing intricate black-box models and provides practical guidance for the deployment of dense low-cost sensor networks for monitoring extreme air pollution events.

From a practical perspective, this study provides guidance for geographers, urban planners, and environmental managers on how different outlier detection approaches can be applied to low-cost sensor networks for monitoring extreme air pollution events. The findings highlight the importance of combining temporal and spatial methods to better interpret sensor data in real-world urban environments.

2. Related Works

Several studies have examined outlier detection and sensor data quality improvement, particularly in air pollution. Liang et al. (2021) applied spatial interpolation techniques to $PM_{2.5}$ data to improve data accuracy. In line with these efforts, Chen et al. (2018) presented an Anomaly Detection Framework (ADF), aimed at improving the data quality in large-scale $PM_{2.5}$ smart city sensing systems. It implements four modules: Time-Sliced Anomaly Detection (TSAD), Real-Time Emission Detection (RED), Device Ranking (DR), and Malfunction Detection (MD).

An attempt to increase the accuracy for a low-cost sensor is done by Ottosen and Kumar (2019). They applied k-Nearest Neighbour (k-NN) and AutoRegressive Integrated Moving Average (ARIMA) outlier detection techniques to a single sensor of different air quality indicators, questioning whether outliers occur during some errors or happen under some unusual conditions. Another question raised was how spatial outliers can be managed in networks of low-cost environmental sensors. A related study by Bobbia et al. (2015) focused on detecting spatial outliers of hourly $PM_{10.0}$ concentration in Normandy, France. They evaluated two methods, a nearest-neighbors weighted median and a kriging-based method, able to detect spatial outliers. Similar approaches combining spatial information for environmental sensor data quality assessment have been reported by Janeja et al. (2010) and Liang et al. (2021). Outlier detection has also been applied to indoor environmental monitoring applications, such as in Wei et al. (2020), where they performed outlier detection in $PM_{10.0}$ data from low-cost sensors.

In de Azevedo et al. (2022) the use of IoT for environmental monitoring is discussed. Their research proposes data improvement through spatial statistics and geostatistics. In Wireless Sensor Networks (WSNs), the detection of outliers is challenging, and studies have been conducted on the detection of extreme values and more specifically on the classification of outlier points as errors or as events with high accuracy. Kamal et al. (2016) attempted this classification, using a combination of fuzzy logic and spatiotemporal similarity.

Research considering spatial outlier detection has also been done for environmental parameters, such as CO2 using Spatial Local Outlier Factor (Xin et al., 2015) and $PM_{10.0}$ with jackknife and kriging approaches (Bobbia et al., 2015).

Three different outlier detection methods, namely, Interquartile Range (IQR), Local Outlier and Global Outlier, were applied in this research. These methods were chosen to be applied to analyze temporal and spatial variations within environmental sensor measurement data. IQR is a statistical method detecting temporal outliers based upon the interquartile range of historic measurements and has application to detecting deviations over the short term. Local Outlier and Global Outlier integrate spatial data and provide information on local environmental changes and broader trends, respectively. The comparison of these approaches aims to improve the understanding of outlier detection in real-time spatio-temporal environmental management.

These methods are applied to environmental data from PurpleAir smart sensors, recording temperature, relative humidity and particulate matter in Athens, Greece. There are three research questions addressed here: (a) How can the spatial dimension of the data be used to support the distinction between outliers and potential environmental events?, (b) How do these different methods for outlier detection compare in terms of behavior, sensitivity and practical applicability?, and (c) Does the temporal scale of data affect the efficacy of outlier detection methods?.

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

Similar wildfire-related studies have been conducted by Masoom et al. (2023) and Giannaros et al. (2022). Masoom et al. (2023) examined the effects of Greek wildfires on aerosol optical properties and solar irradiance, while Giannaros et al. (2022) analyzed meteorological conditions and pyroconvection potential during the 2021 fire events. However, these studies do not address sensor-based anomaly detection or the reliability of low-cost environmental data.

In this paper, we examine the reliability of real-time environmental sensor networks during wildfire events using outlier detection methods. This assesses environmental impacts and proposes a data-driven framework to support the interpretation of sensor anomalies and possible pollution-related events, which few of the above studies addressed. This paper contributes to the discussion by focusing on urban sensor data quality and anomaly identification in a spatio-temporal context.

3. Method

To explore whether a sensor record represents an actual event or an outlier value, this paper uses three methods to analyze environmental sensor data during an actual event. Three outlier detection techniques are used, both statistical and spatial, Interquartile Range (IQR) method, Local Outlier and Global Outlier, to analyze environmental data from PurpleAir sensors in the city of Athens, the capital of Greece. The aim was to examine whether different outlier detection methods can support the interpretation of statistical outliers and patterns potentially associated with environmental events. In this case the event was wildfires near Athens in August 2021.

The selected methods were chosen based on how they handle spatial and temporal variation of sensor readings differently, and how these could be compared from different anomaly detection frameworks. The Interquartile Range (IQR), which is a widely used, noise-resistant, and computationally simple method, is appropriate for detecting abrupt temporal anomalies from each sensor's historic records, but it doesn't account for spatial dimension which is important for environmental monitoring (Ottosen and Kumar, 2019; Zhang et al., 2012). In contrast, Local and Global Outlier methods consider spatial dimension, as Local Outlier looks for anomalies based on local neighborhood behavior and Global Outlier tests against the overall sensor network. Earlier in Zafeirelli et al. (2024), methods based on Interquartile Range (IQR), and Generalized Extreme Studentized Deviate (GESD) did not account for spatial organization.

In this work, the contemporary urban-wildfire pollution context demands responsiveness to local anomalies and pervasive environmental alterations, which are valid grounds for using spatial-aware methods.

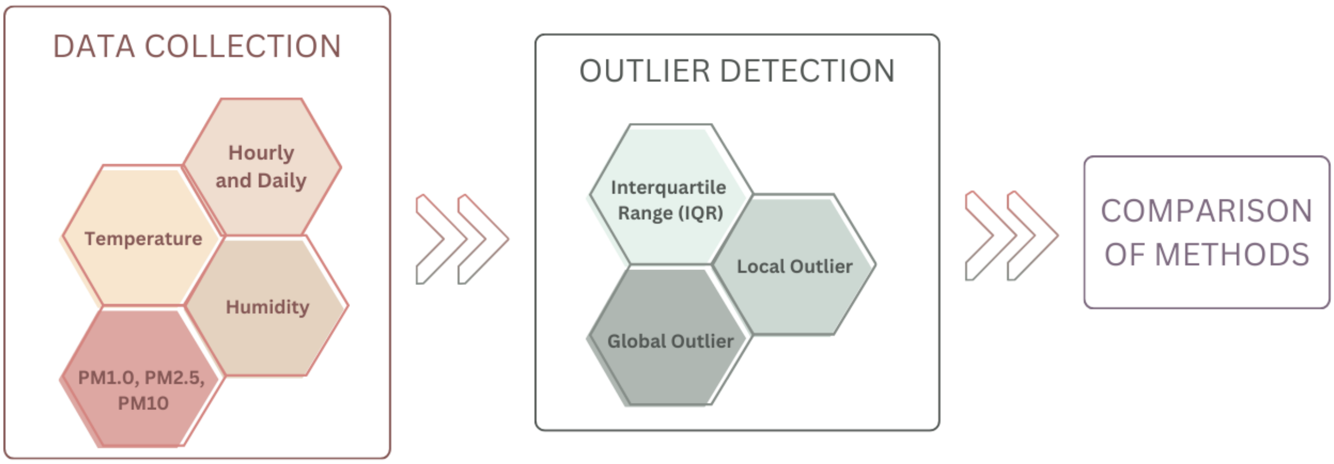


Figure 1. Methodology followed. First Data Collection, then Outlier Detection and finally Comparison of the Outlier Detection methods

The Local Outlier method is based on neighborhood-based statistical outlier detection, which checks deviations against the local mean and the variance of nearby sensors. This method was introduced in spatial anomaly detection frameworks for sensor networks (Janeja et al., 2010; Zhang et al., 2012). The Global Outlier method refers to the classical statistical outlier detection which is based on the standardized deviations from the overall mean of the sensor network (Barnett & Lewis, 1994; Aggarwal, 2017).

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

The mean and the standard deviation of the *k* nearest neighbors for each data point are calculated in Local Outlier and in Global Outlier the mean and standard deviation of the entire dataset at a point in time is calculated. Then for both methods, z-scores are calculated to identify outliers (Alghushairy et al., 2020). Local Outlier considers the immediate neighborhood of each data point, making it highly sensitive to local variations (Ayadi et al., 2017). This can be beneficial for detecting slight changes, but it can also result in false positives, which is possible in environments with localized environmental variability (Janeja et al., 2010). Global Outlier method can sometimes overlook localized events, missing spatial outliers.

Data were collected in aggregated hourly and daily temporal intervals, allowing multi-scale analysis capturing short-term and long-term fluctuations and trends in air quality parameters during the period from August 1 to August 31, 2021. The methodology followed (Figure 1) enables the comparison of outlier detection methods and their effectiveness in indicating patterns potentially associated with environmental events in urban low-cost sensor networks.

The data for the IQR method were taken from Zafeirelli et al. (2024). The IQR method determines the first (Q1) and third quartile (Q3) and then calculates the range between Q3 and Q1, IQR = Q3 – Q1 (Tukey, 1977). Outliers were defined using the more conservative thresholds $Q1 - 3 \times \mathrm{IQR}$ and $Q3 + 3 \times \mathrm{IQR}$, instead of the standard Tukey fences of $Q1 - 1.5 \times \mathrm{IQR}$ and $Q3 + 1.5 \times \mathrm{IQR}$, in order to focus on more extreme deviations and reduce false positives caused by short-term sensor variability.

The outlier detection analysis is performed independently for each time step to ensure that it reflects instantaneous spatial deviations instead of temporal aggregation effects.

3.1. For the Local Mean and Standard Deviation method:

For each data point *i*, the local mean and standard deviation are computed using its *k* nearest neighboring sensors. In this study, *k = 3* nearest neighbors were considered. This value was selected as a practical methodological choice given the spatial density and distribution of the available sensors in Athens, allowing the Local Outlier method to reflect nearby conditions while maintaining sensitivity to localized deviations. No formal sensitivity analysis was conducted to optimize the value of *k*.

$$\mu_{\mathrm{local}}(i) = \frac{1}{k}\sum_{j=1}^{k} x_j$$

$$\sigma_{\mathrm{local}}(i) = \sqrt{\frac{1}{k-1}\sum_{j=1}^{k}\left(x_j - \mu_{\mathrm{local}}(i)\right)^2}$$

where *x_j* is the measurement value of the *j*-th nearest neighboring sensor.

The local *z*-score is computed as:

$$z_{\mathrm{local}}(i) = \frac{x_i - \mu_{\mathrm{local}}(i)}{\sigma_{\mathrm{local}}(i)}$$

3.2. For the Global Mean and Standard Deviation method:

The mean and standard deviation of the variable across all *n* sensors at a given time are computed.

$$\mu_{\mathrm{global}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma_{\mathrm{global}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \mu_{\mathrm{global}}\right)^2}$$

The global *z*-score is computed as:

$$z_{\text{global}}(i) = \frac{x_i - \mu_{\text{global}}}{\sigma_{\text{global}}}$$

The main distinction between the Local and Global Outlier methods is the spatial scale of the reference that is used for comparison. The Local Outlier technique assesses each sensor by looking at its closest neighbors, so it can identify localized deviations relative to nearby sensors. On the other hand, the Global Outlier method looks at each observation in the context of the entire set of sensors, thereby revealing the larger scale spatial patterns and environmental changes that are pervasive.

When $|z| > 1.96$ a data point is classified as an outlier. The threshold of ± 1.96 corresponds to the *95%* confidence interval of a standard normal distribution under the assumption of approximate normality. This approach assumes approximate normality of the data distributions; however, no formal normality tests were performed, and this assumption is acknowledged as a limitation of the method.

The final step was the comparison of the results of the three methods used for hourly and daily data. For each data set, records were grouped by outlier detection method (IQR, Local outlier, Global Outlier) and date-time, or date, and the average of each environmental measurement ($PM_{1.0}$, $PM_{2.5}$, $PM_{10.0}$, temperature, humidity). Mean values of each measurement for inliers and outliers for each outlier detection technique were calculated. These values were used to calculate the Difference (Outliers - Inliers), which is the absolute difference between the average values of Outliers and Inliers for each method, and the Percentage Difference which is the percentage increase or decrease in the average value of outliers compared to inliers, expressed as a percentage of inliers, quantifying the size of deviation for each method.

4. Data

The data used in this study were downloaded from the PurpleAir platform. This platform interconnects low-cost sensors that measure environmental parameters. Sensors are widely used in different sectors like citizens, researchers or organizations, ensuring real-time observation of environmental conditions in locations worldwide. The dataset used in this study consists of hourly and daily air quality sensor measurements in the capital of Greece, Athens. The sensors cover the urban municipalities with the most population of Athens and the sensor network is denser in the center of the city. This sensor distribution is suitable due to the morphology of the city's relief.

The variables used consist of Temperature (°C), Relative Humidity (%) and air quality with measurements of fine particles, Particulate Matter (PM), of sizes 1.0, 2.5 and 10.0 micrometers. Particulate Matter sensors use laser counters, which draw air into a small chamber and then the laser detects the density of particles. For the present study, PM values were obtained using the ATM (Atmospheric) correction column made available by the PurpleAir platform. This correction adjusts raw laser particle count data by applying a calibration factor to bring it closer to reference-grade monitors in standard outside conditions. The ATM correction assumes standard relative humidity and atmospheric pressure and adjusts $PM_{2.5}$ and $PM_{10.0}$ readings through empirically based linear regression equations (e.g., PM2.5_ATM = 0.778 × CF1 + 2.65, where CF1 is raw concentration). The absolute levels and frequency of detected outliers can be influenced by this correction. For instance, what would be below an outlier threshold raw would be above it after correction, especially under conditions of high humidity or smoke, which bias laser scattering. Therefore, resulting outlier detections, particularly for the IQR and Global methods which are magnitude-sensitive, could be affected by real-world changes and by the effect of this correction model.

PurpleAir sensors record data in real time, usually presenting the summarized data every 2 minutes, as well as the facility to get summaries for a period other than this as well. For the purposes of this study, hourly and daily summaries were downloaded. This choice was made to balance manageability, since the original 2-minute dataset presented computation and storage issues. While potentially capturing more transient phenomena using higher-frequency data, aggregated summaries represent a balance of detail and tractability to analysis. Noted is that PurpleAir sensor specific data utilized herein are no longer publicly available for free, but at the time of data collection were freely accessible. Access to historic sensor data on PurpleAir is, as of the publication dates, paid-only, which constrains full reproducibility. Methodological inputs and parameters are given, though, to aid possible reproduction upon similar datasets. The data was collected for August 2021, when the fire events occurred. There were 36 sensors for PM data and 35 sensors for temperature and humidity data available for this period.

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

Data quality is affected by several factors, such as the environment and the maintenance of the sensors. PurpleAir sensors are known for their high sensitivity to data accuracy. Their accuracy can be influenced by environmental factors like humidity and temperature, and they might require frequent calibration. The accuracy of the measurements has been the subject of study and comparison with other more reliable monitoring devices (Couzo et al., 2024). Usually, PurpleAir sensors have good data correlation with reference-grade instruments, but sometimes deviations occur under different environmental conditions and improvement of data accuracy is needed through calibration and correction.

5. Case study: Wildfire events near the city of Athens, Greece in August 2021

Greece has a Mediterranean climate with seasonal variations. During the summer season, which is from June to August, the weather conditions are typically hot and dry, while the average temperatures often exceed 30°C. Summer season has minimal rainfall and it's usually less than 5 mm per month, increasing dry conditions and the risk of wildfires. In Athens, the capital of Greece, the daytime temperatures often exceed 40°C and the solar radiation is strong with high levels of UV radiation.

In August 2021, Attica and broader areas in Greece experienced extreme wildfire events (Figure 2) during a historic heatwave, which significantly affected air quality over the city of Athens. The wildfires coexisted with a Saharan dust event over Athens due to southern winds, which combined with the smoke from the fire events. This mixture resulted in increased levels of air pollution and solar radiation (Masoom et al., 2023). The region had a severe heatwave that lasted a long time and experienced, during this time, very high temperatures, very dry air, quiet winds, and a significant shortage of rainfall. The combination of these weather conditions offered the best opportunity to create an ideal setting for wildfires to ignite and spread rapidly.

A wildfire broke out on the 3rd of August 2021, at 13:22, at the foot of Mount Parnitha (location in Figure 2) in a settlement named Varympompi, located 18 km north of the center of Athens, within one of the wildland-urban interface (WUI) areas surrounding the city. According to the Fire Service records, the fire was fully controlled on 7th August 2021 and on 11th of September 2021 was officially extinguished.

The Varympompi fire quickly spread due to the wind conditions in the area in combination with flammable forest fuels (*Pinus halepensis*). During the two hours following ignition, wind conditions were light to moderate, averaging from 2 to 4 Bf (Beaufort scale) with gusts going as high as 5 Bf. These meteorological conditions contributed to the rapid spread of the fire, consistent with the analyses reported by Giannaros et al. (2022). The fire, even though was brought under partial control on 5th of August early in the morning, reignited in the afternoon in Vasilika Parnithas, worsening the situation by extending to Afnides and continued to burn heading towards Agios Stefanos and Lake Marathon. Air quality changes observed in Athens during this period were temporally consistent with the Varympompi wildfire, which burned approximately 8,377 ha. There were also two other fire events, one on Evia Island (3rd to 15th of August 2021) which is about 200 km away from Athens, with about 51,000 ha burned, and another one between the 16th and 25th of August in Vilia (locations in Figure 2), which is closer than Evia to Athens, burning an area of 9,400 ha (Masoom et al., 2023).

The wildfire events are an important case study for the analysis of wildfire impacts on air quality. The time scale studied allows the monitoring and study of changes in microparticle concentrations, temperature and humidity, before, during and after the fire events. A multi-scale method is applied for data analysis that captures both short-term fluctuations and long-term air quality trends. The hourly data may provide a more accurate representation of the situation and the shift in patterns, whereas daily data can be analyzed to identify the overall trends. This method helps examine spatio-temporal patterns in air quality during periods coinciding with wildfire events, which may have a severe impact on air quality not only at the beginning but also in the course of time.

6. Results

Descriptive statistics reveal differences between hourly and daily data on environmental parameters (Table 1). The hourly data shows a broader range and higher variability as seen in between maximum and minimum values, as well as larger standard deviations for all metrics. For example, for daily temperature data between 27.5°C and 43.4°C, hourly measurements while covering between 23.3°C and 50.0° indicate more minute variations due to shorter time sample recording. The hourly $PM_{1.0}$, $PM_{2.5}$, and $PM_{10.0}$ measurements cover a significantly wider range and greater variability compared to daily data. For instance, even though daily $PM_{2.5}$ concentrations are between 0 and 461.4 μg/m^3, the hourly data reports values between 0 and 554.9 μg/m³, thus short-term spikes in PM are better captured in hourly measurements. Similarly, hourly $PM_{10.0}$ concentrations can be as high as 632.9 μg/m³, while daily values peak at 187.6 μg/m³. This higher variability (hourly data) implies that PM concentrations can indeed change considerably in a short span of time, reflecting short-term pollution episodes or transient increases in particulate matter that are less visible in daily averages.

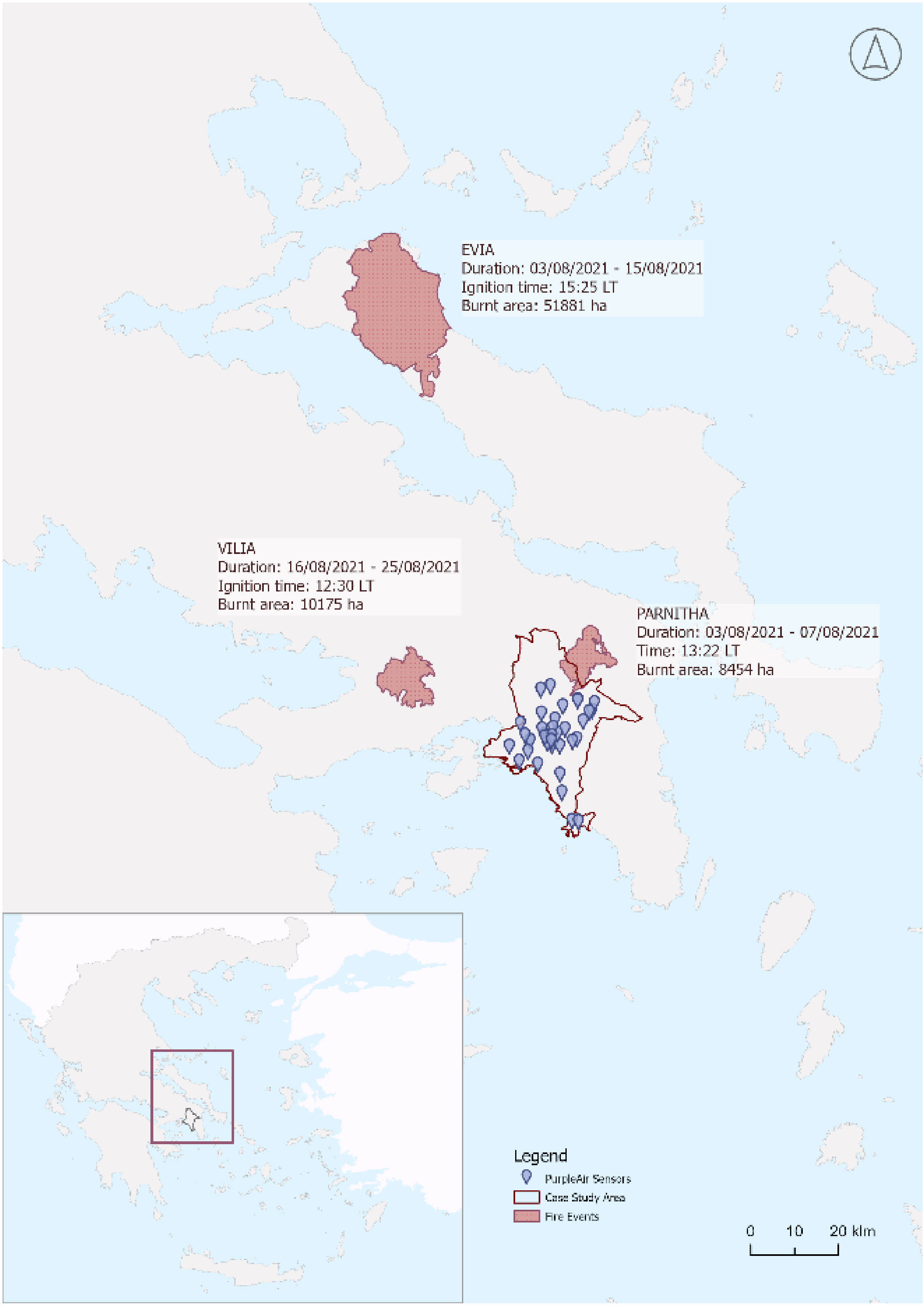


Figure 2. Wildfire events in the broader area of Athens

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

Table 1. Descriptive statistics for daily and hourly environmental metrics including Temperature (°C), Humidity (%), $PM_{1.0}$ (μg/m3), $PM_{2.5}$ (μg/m^3), and $PM_{10.0}$ (μg/m^3). Frequency is the number of observations.

Daily					
	frequency	min	max	mean	std. deviation
Temperature (°C)	971	27.5	43.4	34.4	2.7
Humidity (%)	971	11.3	84.7	29.7	8.6
$PM_{1.0}$ (μg/m^3)	999	0	94.4	13.9	9.7
$PM_{2.5}$ (μg/m^3)	999	0	461.4	22.7	28.3
$PM_{10.0}$ (μg/m^3)	999	0	187.6	21.7	16.4
Hourly					
	frequency	min	max	mean	std. deviation
Temperature (°C)	22,911	23.3	50	34.4	4.4
Humidity (%)	22,936	1.4	100	29.7	11.9
PM1.0 (μg/m3)	23,577	0	383.5	13.8	14.8
PM2.5 (μg/m3)	23,574	0	554.9	19.7	23.1
PM10.0 (μg/m3)	23,577	0	632.9	21.6	26

Analysis of daily and hourly environmental data reveals the patterns and the correlations in the occurrence of outliers in different metrics (Table 2). The percentage of outliers detected by the IQR method is low across all metrics, showing that the majority of data fall within the expected ranges. Particulate Matter data ($PM_{1.0}$, $PM_{2.5}$ and $PM_{10.0}$) show greater variability compared to temperature and humidity, which show the differences between air quality parameters and particle concentrations. This variability may be associated with localized pollution sources or varying atmospheric conditions, which could result in elevated particle concentrations that are not captured by temperature or humidity measurements. The distinction between local and global outliers is important as it enables the identification of localized deviations and broader spatial patterns.

Local outliers present higher percentages compared to IQR and Global outliers (Table 2), indicating localized anomalies, suggesting sensitivity to localized deviations, which may reflect microclimatic conditions, short-term events, or sensor-level variability. Global outliers show consistent but lower percentages reflecting a more selective approach that identifies only the most extreme deviations from the overall spatial distribution of the data.

Table 2. Outlier statistics for daily and hourly environmental metrics (Temperature (°C), Humidity (%), $PM_{1.0}$ (μg/m3), $PM_{2.5}$ (μg/m3), $PM_{10.0}$ (μg/m3)). Shows the frequency (Freq) which is the number of observations and percentage (Perc %) of data points identified as outliers by Interquartile Range (IQR), Local Outliers, and Global Outliers. "Yes" represents the Outliers and "No" represents the Inliers of each method

Daily

		Temperature (°C)		**Humidity (%)**		**$PM_{1.0}$ (μg/m^3)**		**$PM_{2.5}$ (μg/m^3)**		**$PM_{10.0}$ (μg/m^3)**	
		Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)
IQR Outlier	Yes	7	0.72	1	0.10	21	2.10	38	3.80	15	1.50
	No	964	99.28	970	99.90	978	97.90	961	96.20	984	98.50
Local Outlier	Yes	229	23.58	179	18.43	241	24.12	229	22.29	217	21.72
	No	742	76.42	792	81.57	758	75.88	770	77.08	782	78.28

		Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)
Global Outlier	Yes	70	7.21	49	5.05	43	4.30	46	4.60	44	4.40
	No	901	92.79	922	94.95	956	95.70	953	95.40	955	95.60

Hourly

		Temperature (°C)		Humidity (%)		$PM_{1.0}$ (μg/m³)		$PM_{2.5}$ (μg/m³)		$PM_{10.0}$ (μg/m³)	
		Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)	Freq	Perc (%)
IQR Outlier	Yes	241	1.05	33	0.14	465	1.97	427	1.81	474	2.01
	No	22,670	98.95	22,903	99.86	23,112	98.03	23,147	98.19	23,103	97.99
Local Outlier	Yes	5,190	22.65	4,569	19.92	5,394	22.88	5,044	21.40	4,937	20.94
	No	17,721	77.35	18,367	80.08	18,183	77.12	18,530	78.60	18,640	79.06
Global Outlier	Yes	1,486	6.49	1,201	5.24	1,181	5.01	1,213	5.15	1,220	5.17
	No	21,425	93.51	21,735	94.76	22,396	94.99	22,361	94.85	22,357	94.83

Figure 3 shows the detection of $PM_{10.0}$ outliers for each day of August using the three different methods (IQR, Local Outlier, Global Outlier). A significant spike is observed between August 4 and 7 with the IQR method, probably due to the outbreak of fires during this period. Global outlier detection method does not show any significant spike, suggesting that it may not be as sensitive to the short-term increases of $PM_{10.0}$ daily levels. This can indicate that global outliers are more suitable for detecting very extreme and widespread deviations, which have not occurred in this case. There are some lower spikes during the fire events and end after the fires. The Local outlier method does not show any significant spikes during the outbreak of the fires.

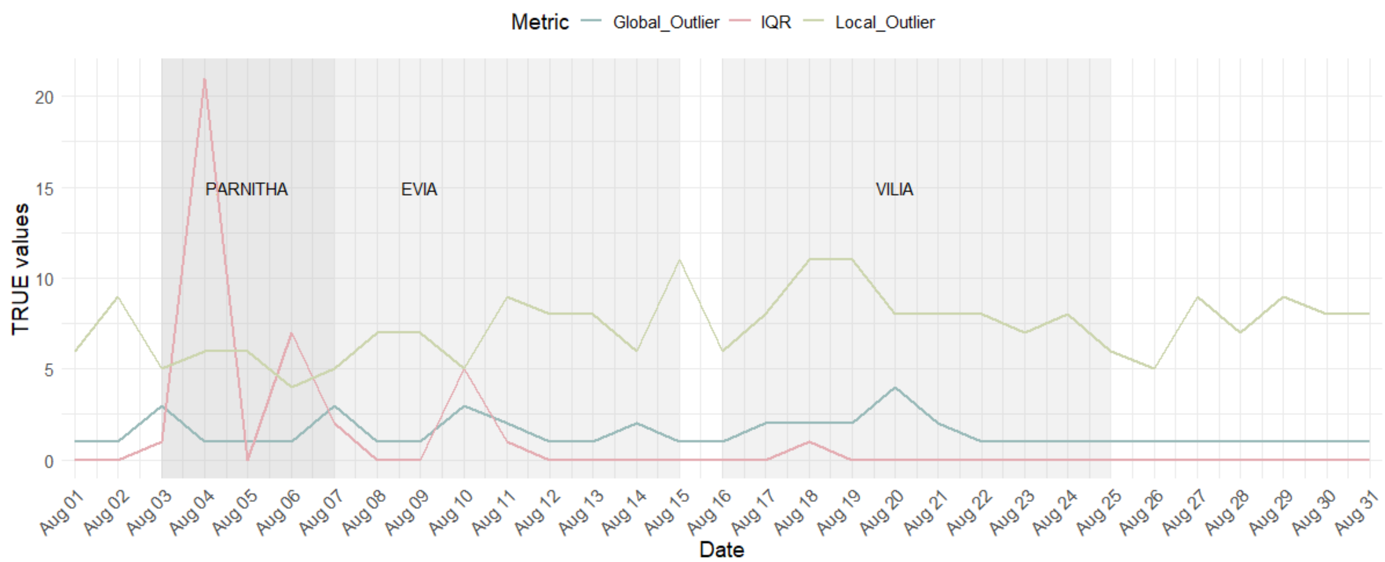


Figure 3. True values (outliers) of each outlier detection technique, Global Outlier, Local Outlier, IQR over time of daily $PM_{10.0}$ (μg/m³) data.

Figure 4 represents outliers of the three methods applied on hourly $PM_{10.0}$ data. The IQR method appears to capture the sudden increases in fine particles near fire ignition dates of the broader area, as well as the increased amounts during the fires. These increases are temporally consistent with the wildfire period and may reflect delayed smoke transport towards Athens, depending on meteorological conditions. The Global outlier method is mostly flat with a slight increase. A similar pattern was captured in global daily outliers, where it was more obvious after 11th of August. The Local outlier method has a very active pattern, detecting variations throughout the period.

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

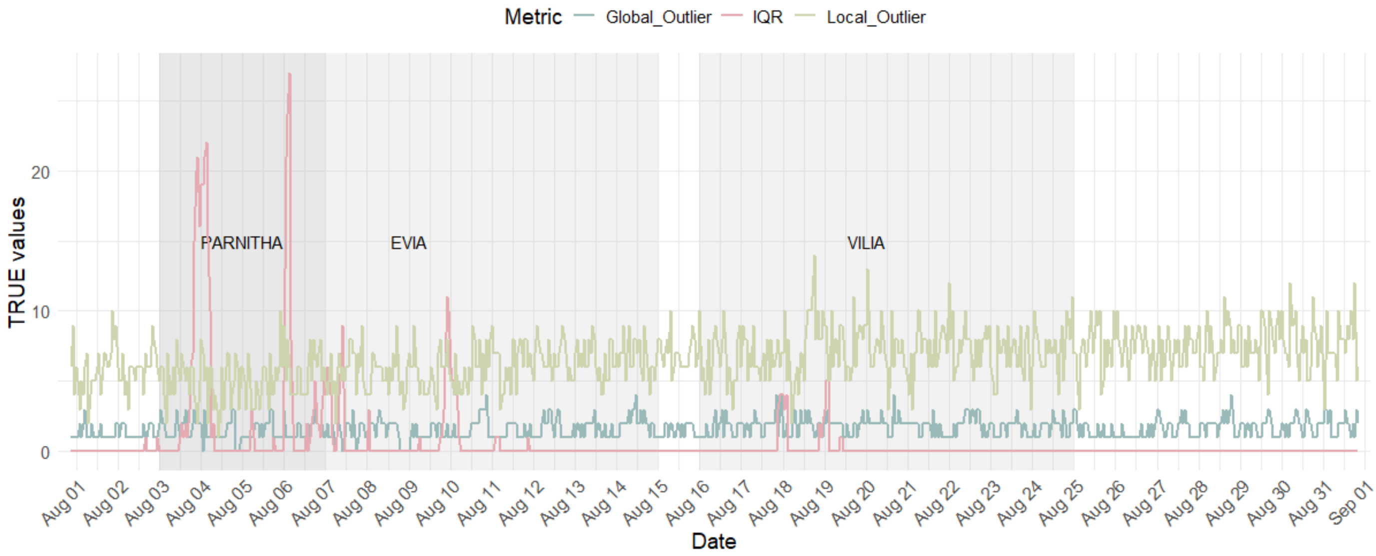


Figure 4. True values (outliers) of each outlier detection technique, Global Outlier, Local Outlier, IQR over time of hourly $PM_{10.0}$ (μg/m^3) data.

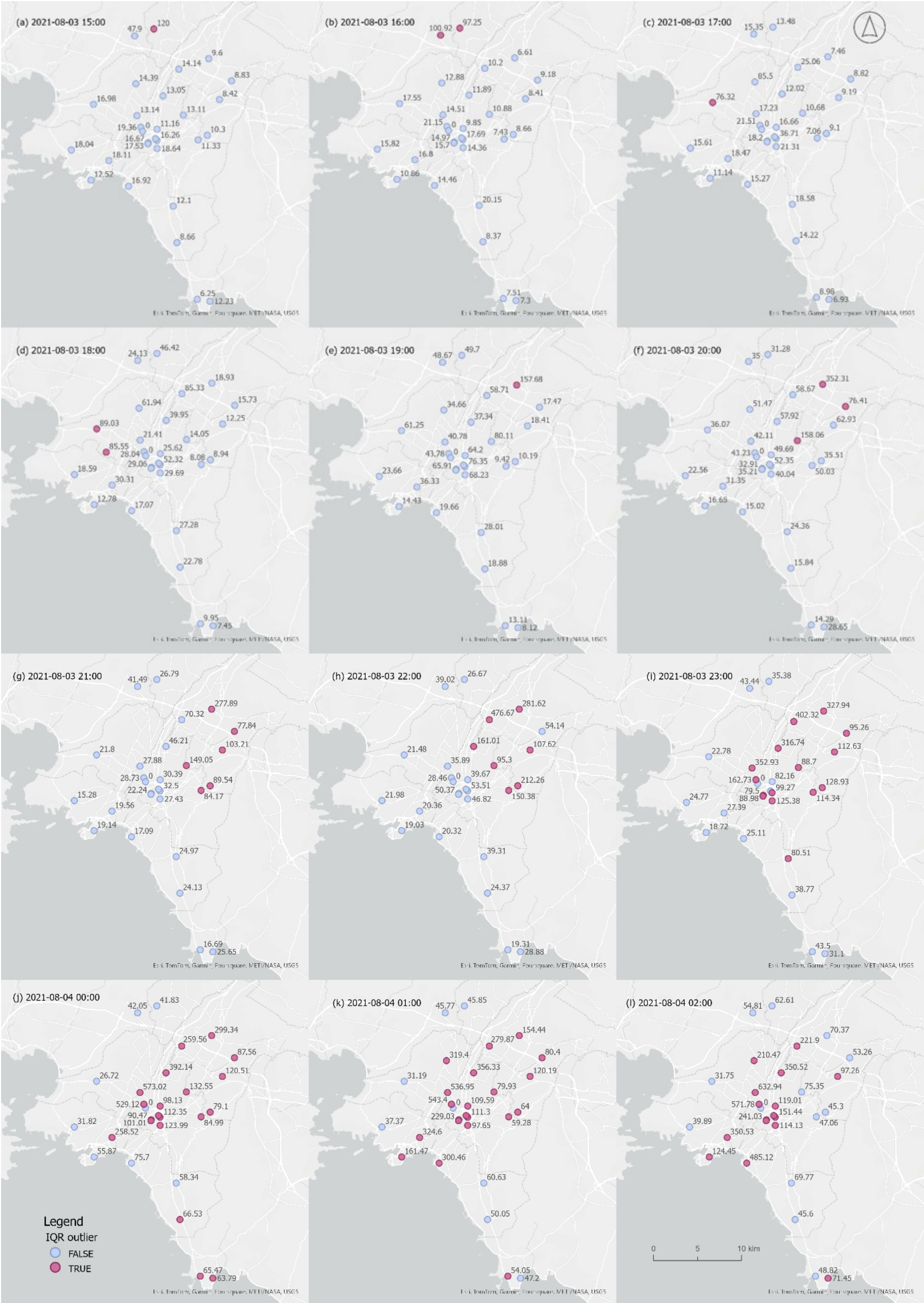
The IQR method, both in hourly and daily data, is more balanced compared to the other methods. It captures significant spikes associated with wildfire smoke events, while not being very sensitive to small fluctuations. The Local outlier method is more sensitive, since it detects even small changes. It could be useful for highly detailed monitoring, but it can also lead to false positives. Global outlier method is stable, but it cannot respond in rapid or local event detection, especially in hourly data.

The IQR method seems to capture spatio-temporal $PM_{10.0}$ outlier patterns during the wildfire period, as seen in Figures 5a- 5t. The spatial distribution of $PM_{10.0}$ outliers is consistent with possible smoke movement across the study area, highlighting some areas where the concentrations of fine particles have a large deviation in the historical data of each sensor. This phenomenon over time may reflect the movement of smoke influenced by wind conditions and local geography, which is probably due to wildfires in the broader area, throughout the urban landscape. The changing positions of the outliers may be consistent with evolving smoke distribution patterns across the city, with concentrations in different areas likely influenced by wind patterns as well as geographic features. These illustrations effectively demonstrate the ability of the IQR method to detect sudden and unusual increases in microparticles, which the other two methods are unable to detect.

The Global outlier method identifies a measurement point in the center of the city that consistently shows zero value. This persistence suggests a likely sensor malfunction rather than a true environmental anomaly (Figure 5u).

Table 3 compares the outlier detection techniques used in this paper (IQR, Local Outlier, Global Outlier) applied to hourly and daily environmental data of Particulate Matter, temperature and humidity. It presents the differences between outlier and inlier data points. These are descriptive indicators and should not be interpreted as effect sizes or accuracy metrics. The IQR method for $PM_{10.0}$ fine particles shows that the daily outliers have an average value of 111.41 μg/m^3 as opposed to the inliers, which have a value of 20.34 μg/m^3, showing a substantial descriptive difference of 447.78%. This difference indicates that outliers deviate from the expected air quality levels, which may be consistent with extreme pollution episodes during the wildfire period. On the contrary, Local Outlier and Global Outlier techniques detect smaller or negative differences, indicating that it is important to choose the appropriate method for the interpretation of environmental outliers. In addition, the IQR method produced similarly large descriptive differences for $PM_{2.5}$, reaching 551.62% in the daily data and 679.99% in the hourly data, indicating that short-term pollution peaks are more strongly expressed at finer temporal resolutions. Likewise, the percentage difference of $PM_{1.0}$ reached 577.32% in hourly data compared with 287.05% in daily data, demonstrating that more detailed temporal analysis can capture more extreme short-term variations that may be smoothed in daily averages.

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230



https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

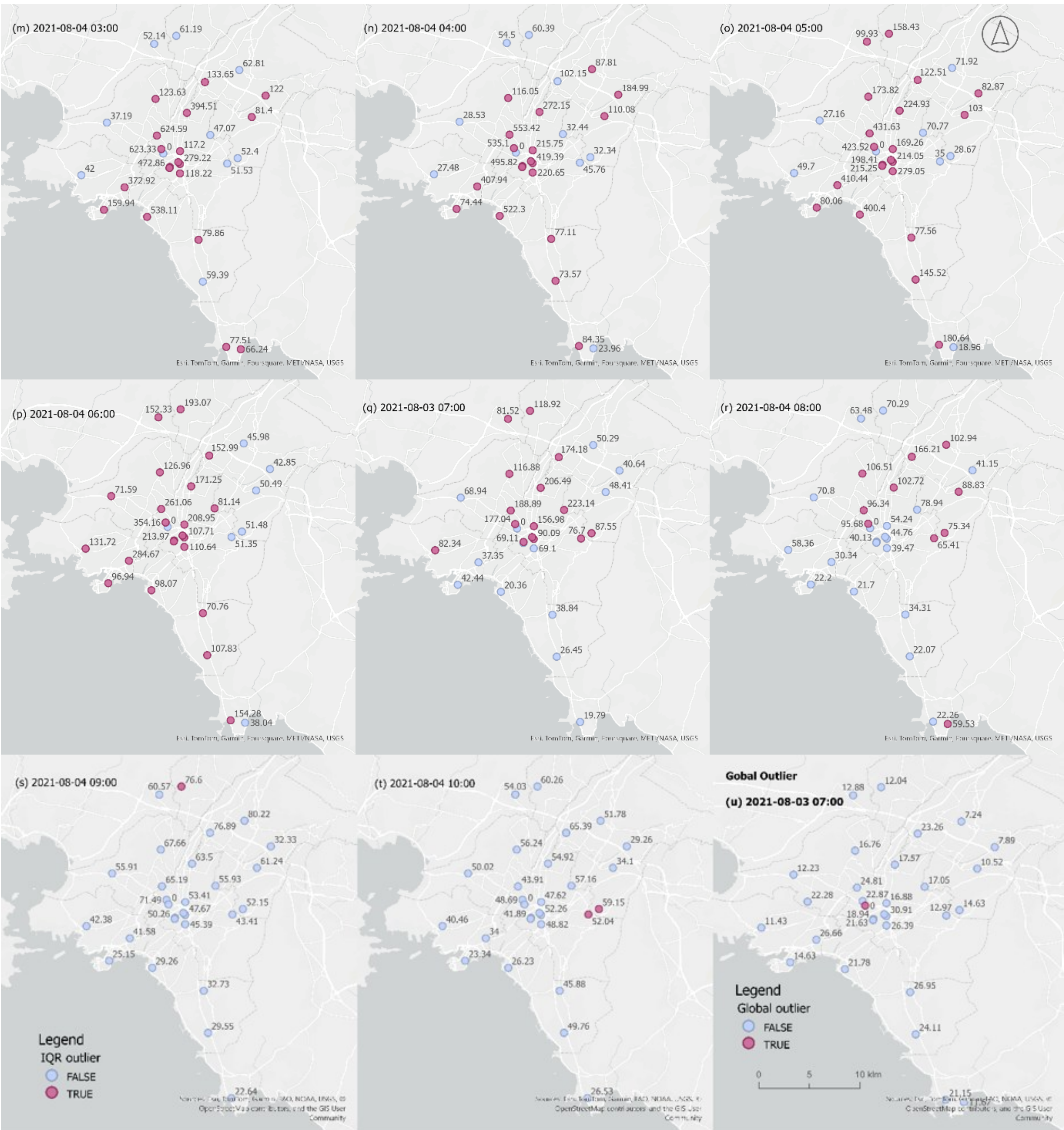


Figure 5. *Figures 5a-5t:* Outliers detected with IQR method on hourly $PM_{10.0}$ data during smoke fire event (03/08/21 15:00 - 04/08/21 10:00). *Figure 5u* represents a persistent outlier value which was also present in many datetimes using Global Outlier technique on 03/08/21 15:00. Outliers are the TRUE values and Inliers are the False values. The wildfire events occurred in the broader Attica region (northern and eastern areas), which are not explicitly shown on the map due to scale limitations.

7. Discussion

This study addresses the challenge of supporting the interpretation of likely sensor anomalies and environmentally driven patterns in urban air quality monitoring. Previous studies focus on improving the accuracy of air quality sensors in cities (Munir et al., 2019; Tagle et al., 2020). This study introduces a comparative approach to outlier detection using three methods: Interquartile Range (IQR), Local Outlier and Global Outlier. These three methods are evaluated in terms of supporting the interpretation of statistical outliers and patterns potentially associated with environmental events in the context of wildfire pollution in cities.

Table 3. Comparison of outlier detection techniques applied on all hourly and daily environmental measurements (PM, temperature, humidity). Inlier (FALSE): The average value of the metric for inlier data points. Outlier (TRUE): The average value for outlier data points. Difference (TRUE - FALSE): The difference between the outlier and the inlier mean values. Percentage Difference: The percentage change between the outlier and inlier means, measuring how much the outliers deviate from the expected range.

Metric	Method	Inlier (FALSE)	Outlier (TRUE)	Difference (TRUE - FALSE)	Percentage Difference
Daily Data					
$PM_{10.0}$ (μg/m³)	IQR	20.34	111.41	91.07	447.78
	Local Outlier	22.42	19.14	-3.28	-14.63
	Global Outlier	21.79	19.96	-1.83	-8.40
$PM_{2.5}$ (μg/m³)	IQR	18.73	122.08	103.34	551.62
	Local Outlier	22.82	22.14	-0.68	-2.99
	Global Outlier	22.45	27.08	4.63	20.62
Temperature (°C)	IQR	34.34	37.82	3.48	10.13
	Local Outlier	34.46	34.06	-0.40	-1.17
	Global Outlier	34.38	34.14	-0.24	-0.69
Humidity (%)	IQR	29.69	39.99	10.30	34.71
	Local Outlier	28.63	34.41	5.78	20.18
	Global Outlier	28.56	51.15	22.59	79.12
$PM_{1.0}$ (μg/m³)	IQR	13.16	50.92	37.76	287.05
	Local Outlier	14.55	12.07	-2.48	-17.02
	Global Outlier	14.24	7.37	-6.87	-48.24
Hourly Data					
$PM_{10.0}$ (μg/m³)	IQR	19.98	115.72	95.74	479.29
	Local Outlier	22.44	21.08	-1.36	-6.07
	Global Outlier	21.89	22.51	0.62	2.85
$PM_{2.5}$ (μg/m³)	IQR	17.54	136.80	119.26	679.99
	Local Outlier	20.09	18.28	-1.81	-9.00
	Global Outlier	19.53	22.86	3.34	17.09
Temperature (°C)	IQR	34.30	41.62	7.32	21.35
	Local Outlier	34.19	35.03	0.84	2.47
	Global Outlier	34.33	34.97	0.64	1.86
Humidity (%)	IQR	29.60	80.41	50.82	171.70
	Local Outlier	28.32	35.11	6.79	24.00
	Global Outlier	28.31	54.30	26.00	91.84
$PM_{1.0}$ (μg/m³)	IQR	12.41	84.05	71.64	577.32
	Local Outlier	14.20	12.56	-1.64	-11.52
	Global Outlier	13.81	14.10	0.29	2.11

Wildfires and other types of peri-urban fires can have far-reaching effects on the environmental conditions and public health, which are sometimes felt even in urban areas. Fires on the outskirts of cities, on a regular basis, can be one of the major causes for the deterioration of air quality, as they can significantly increase particulate matter concentrations in urban areas affected by smoke transport (Braun et al., 2025). This type of pollution poses first threats for vulnerable groups, such as children and the elderly, and people with respiratory problems. For example, the wildfires that swept through California in 2018, resulted in a significant increase in the number of hospital admissions for respiratory and cardiovascular diseases (Reid et al., 2016). Previous studies have reported that the wildfires in Greece during August 2021 were associated with substantial air quality degradation in Athens, including elevated PM concentrations following the wildfire events (Masoom et al., 2023). Furthermore, previous research has shown that long-term exposure to pollutants released during wildfire episodes may contribute to the aggravation of respiratory and cardiovascular health conditions and increased mortality risk (Masoom et al., 2023; Kaskaoutis et al., 2024). The far-reaching effects of these fires, particularly in peri-urban zones, highlight the importance of implementing efficient environmental monitoring and rapid-response systems to help protect public health and the environment.

Our results show significant variability of average and median values at the two different time scales used. Hourly data have in general a considerably wider range and higher variability than daily data, especially for the Particulate Matter, where the highest variability is observed. This aligns with findings from studies on sensor data sensitivity to short-term fluctuations (Miao et al., 2022), where the greater variability in air quality metrics was observed while analyzing data at finer temporal scales. These statistics show how very high-frequency (hourly) data could be employed to capture short-term rapid environmental changes and possible pollution episodes (Miao et al., 2022).

Each of the three outlier detection methods - IQR, Local Outlier, and Global Outlier - offers distinct information for event detection. The analysis revealed notable changes in the recorded environmental parameters during the wildfire period. The comparison of the outlier detection methods revealed that the IQR method is effective in identifying temporal outliers based on historical data and outlier patterns temporally aligned with the wildfire period were observed in most cases. The number of detected outliers increased during and after the wildfire period, and the resulting patterns were consistent with possible smoke-related influences. The IQR method identified sudden increases in $PM_{10.0}$ and $PM_{2.5}$ concentrations during the wildfire period, consistent with the findings of Ottosen and Kumar (2019) that statistical outlier detection is suitable for transient events.

The spatial outlier detection techniques, Local Outlier and Global Outlier methods, provided additional insights by incorporating spatial dimensions. In contrast to the IQR method, the two spatial outlier detection techniques were less sensitive to rapidly changing spatial patterns that may be associated with smoke transport. The Local Outlier method showed higher detection rates (Table 3), which may indicate sensitivity to localized environmental changes, rather than direct evidence of localization. However, its high sensitivity can cause false positives, as shown by Bobbia et al. (2015), who reported that local outlier methods are often more sensitive to localized, transient variations. The Global Outlier method, which tracks notable changes throughout the whole dataset at any time, may not catch localized or sudden pollution spikes, rendering it less sensitive to localized deviations or short-term peaks potentially associated with smoke transport. This is consistent with previous studies on spatial anomaly detection, which reported that globally referenced approaches may be less sensitive to localized spatial anomalies and short-term local variations (Janeja et al., 2010; Ayadi et al., 2017).

Although further refined methods such as DBSCAN, Isolation Forest, or even machine learning algorithms (e.g., LSTM for pattern recognition across time) are available, these either need comprehensive parameter tuning, huge, annotated datasets, or are less interpretative when applied to real-time monitoring. The selected set of methods offers an equitable trade-off between algorithmic simplicity, interpretability, and ability to capture various aspects of the data (ranging from spatial, through temporal, to global vs. local). They are thus suitable and theory-grounded benchmarks for assessing sensor-based detection of outlying observations against which a variety of detection methods could be tested and compared.

A combination of the IQR method and the Global outlier method may be considered as a practical approach (Ottosen and Kumar, 2019; Zhang et al., 2012; Graça et al., 2023) while Global Outlier method is not able to capture sudden changes in particulate matter levels. The IQR method proved effective in identifying transient PM spikes temporally aligned with the wildfire period. The Global Outlier method could be used to detect persistent sensor malfunctions. This combination could be applied in a system where sensor malfunctions would be filtered using the Global Outlier method, while sudden environmental changes would be recorded using the IQR method. This hybrid approach may help improve the accuracy of outlier detection and support more reliable interpretation of possible technical issues and pollution-related patterns. A limitation is that there are no reference-grade monitor data available for validation of identified outlying points. The methods used show that there is internal consistency and reveal spatiotemporal patterns consistent with wildfire pollution. The results would be further validated and strengthened by comparing the outlier points with

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230

certified environmental monitor stations (e.g., National Observatory of Athens or EPA-similar equipment). Such information would verify whether outlying points are indicative of actual pollution or sensor artifacts, especially during intense periods such as wildfires. Future investigations should seek to include co-located reference monitor equipment for improved benchmarking and model assessment.

By enhancing the behavior, sensitivity, and practical applicability of outlier detection in environmental sensor data, the study contributes to more accurate monitoring of air quality, which is important for public health and urban planning, to detect anomalies and potential environmental hazards in real time (Munir et al., 2019; Graça et al., 2023). This study gives important information about the detection of outliers in environmental sensor data. However, there are some drawbacks that need to be recognized. A limitation of this study is that the analysis focuses primarily on wildfire events, without a systematic comparison to non-event periods. Therefore, the performance of the methods under typical background conditions and their ability to separate routine variability from statistically unusual patterns remain areas for further investigation. The analysis is based solely on the PurpleAir sensor data, which is one of the limitations, as -with some calibration being done- the sensor may still be inaccurate or possess some inherent biases particularly due to environmental influences such as temperature and humidity (Hayward et al., 2024; Dong et al., 2025). Moreover, the outlier detection methods employed might not catch all the complex context-related anomalies that may be present in various environmental conditions or urban areas.

8. Conclusions

Urban environments are developing rapidly and concerns about air quality are increasing, as pollutants affect public health and quality of life. These issues are often addressed using environmental sensor networks monitoring the air quality in real time. However, these systems face challenges in ensuring data accuracy and in supporting the interpretation of sensor anomalies and possible pollution-related events. Outlier detection enhances sensor data reliability and supports informed decision-making regarding health and safety of cities.

In this paper, three outlier detection methods were applied and compared to evaluate their advantages and limitations during extreme events such as wildfires. The study was conducted in the city of Athens, Greece during August of 2021, where three wildfires erupted in the broader area. Environmental sensor data was collected from PurpleAir platform including temperature, humidity and Particulate Matter. Three outlier techniques were applied to hourly and daily data, including IQR, Local Outlier, and Global Outlier techniques, where IQR is applied on historical data of each sensor and the two other methods use spatial dimension. Their performance was evaluated during the wildfire events aiming to support the interpretation of likely sensor anomalies and patterns potentially associated with environmental events. The findings may support the interpretation of environmental sensor observations and allow for more timely responses to extreme environmental conditions. It should be noted that the absence of reference-grade measurements limits the validation of the detected outliers, and therefore the distinction between sensor anomalies and environmental events is inferred rather than confirmed.

Future research could explore the integration of additional environmental parameters and the application of the three methods to different cities and environmental events.

Overall, this study demonstrates that the combined use of temporal and spatial outlier detection methods can support a more informed interpretation of low-cost sensor data, offering practical value for geographers, urban planners, and environmental decision-making in the context of extreme events.

8.1. Addressing the Research Questions

This study aims to address three questions. The first one was how the spatial dimension of the data can be used to support the distinction between outliers and potential environmental events. The Local and Global Outlier comparisons indicate that spatial information provides critical context. The Local Outlier captures fine-scale fluctuations, but Global Outlier captures broader departures, justifying the utilization of spatial measures for supporting comparison between localized and broader spatial patterns. The second question asked how various methods of outlier detection could be contrasted and compared based on accuracy and reliability. The outcomes indicate that IQR appears best suited to indicating short-term PM spikes temporally aligned with the wildfire period, particularly when examining hourly-level data. Local Outlier is very sensitive but may be prone to overflagging localized deviations, and Global Outlier is conservative and potentially misses localized phenomena. Without ground truth validation through reference sensors, "accuracy" is relative to consistency and wildfire timeliness expectations. The third question asked whether there was a temporally appropriate frame for that measurement and how that influenced the effectiveness of methods of detecting outliers. The outcomes indicate that hourly-level data captures more frequent and steep spikes, making it best for capturing transient pollution-related variations. The daily level smooths out such variability and hence reduces responsiveness. This supports selecting the temporal resolution according to the type of variation being investigated. On balance, each approach has strengths and synthesizing spatial and temporal approaches is still critical. The findings have clear implications for the design of municipal-scale air quality sensor networks looking to enhance reliability and responsiveness.

However, the distinction between likely sensor anomalies and environmentally driven patterns remains inferential and would benefit from validation using reference-grade monitoring stations.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest. Dimitris Kavroudakis is an Editorial Board Member of this journal and he was not involved in the editorial review process or the decision to publish this article.

Data Availability Statement: Historical PurpleAir data used in this study are no longer freely accessible and currently require paid access through the PurpleAir platform. Consequently, the original raw datasets cannot be redistributed by the authors. Processed datasets aggregated hourly and daily time series, outlier classification results (outlier flags), spatial summary outputs, and the analysis scripts used in this study are available from the corresponding author upon reasonable request.

References

Aggarwal, C. C. (2017). *Outlier Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-319-47578-3

Alghushairy, O., Alsini, R., Soule, T., & Ma, X. (2020). *A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams*. *Big Data and Cognitive Computing, 5*, 1. https://doi.org/10.3390/bdcc5010001

Anggraini, T. S., Irie, H., Sakti, A. D., & Wikantika, K. (2024). Machine learning-based global air quality index development using remote sensing and ground-based stations. *Environmental Advances*, *15*, 100456. https://doi.org/10.1016/j.envadv.2023.100456

Ayadi, A., Ghorbel, O., Obeid, A. M., & Abid, M. (2017). Outlier detection approaches for wireless sensor networks: A survey. *Computer Networks*, *129*, 319–333. https://doi.org/10.1016/j.comnet.2017.10.007

Barkjohn, K. K., Yaga, R., Thomas, B., Schoppman, W., Docherty, K. S., & Clements, A. L. (2025). Evaluation of Long-Term Performance of Six $PM_{2.5}$ Sensor Types. *Sensors*, 25(4), 1265. https://doi.org/10.3390/s25041265

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). John Wiley & Sons, Ltd.

Bobbia, M., Misiti, M., Misiti, Y., Poggi, J. M., & Portier, B. (2015). Spatial outlier detection in the PM10 monitoring network of Normandy (France). *Atmospheric Pollution Research*, *6*(3), 476–483. https://doi.org/10.5094/APR.2015.053

Braun, R. A. & Fraser, M. P. (2025). Influence of wildfire smoke on summertime surface air quality in an urban desert region. *Atmospheric Environment*, 358. https://doi.org/10.1016/j.atmosenv.2025.121297

Bulot, F. M. J., Johnston, S. J., Basford, P. J., Easton, N. H. C., Apetroaie-Cristea, M., Foster, G. L., Morris, A. K. R., Cox, S. J., & Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Scientific Reports*, *9*(1), 7497. https://doi.org/10.1038/s41598-019-43716-3

Chen, L. J., Ho, Y. H., Hsieh, H. H., Huang, S. T., Lee, H. C., & Mahajan, S. (2018). ADF: An Anomaly Detection Framework for Large-Scale PM2.5 Sensing Systems. *IEEE Internet of Things Journal*, *5*(2), 559–570. https://doi.org/10.1109/JIOT.2017.2766085

Couzo, E., Valencia, A., & Gittis, P. (2024). Evaluation and Correction of PurpleAir Temperature and Relative Humidity Measurements. *Atmosphere*, *15*(4). https://doi.org/10.3390/atmos15040415

de Azevedo, L. J. de M., Estrella, J. C., Delbem, A. C. B., Meneguette, R. I., Reiff-Marganiec, S., & de Andrade, S. C. (2022). Analysis of Spatially Distributed Data in Internet of Things in the Environmental Context. *Sensors*, *22*(5), 1693. https://doi.org/10.3390/s22051693

Dong, J., Goodman, N., Carre, A., & Rajagopalan, P. (2025). Calibration and validation-based assessment of low-cost air quality sensors. *Science of the Total Environment*, *977*, 179364. https://doi.org/10.1016/j.scitotenv.2025.179364

EFFIS – European Forest Fire Information System. (2021). *MODIS Burnt Areas: Rapid Damage Assessment (RDA) Module of EFFIS.* Data provided on 20 June 2024 via email from: https://forest-fire.emergency.copernicus.eu/apps/data.request.form/

El-Shafeiy, E., Alsabaan, M., Ibrahem, M. I., & Elwahsh, H. (2023). Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique. *Sensors*, *23*(20), 8613. https://doi.org/10.3390/s23208613

Giannaros, T. M., Papavasileiou, G., Lagouvardos, K., Kotroni, V., Dafis, S., Karagiannidis, A., & Dragozi, E. (2022). Meteorological Analysis of the 2021 Extreme Wildfires in Greece: Lessons Learned and Implications for Early Warning of the Potential for Pyroconvection. *Atmosphere*, *13*(3). https://doi.org/10.3390/atmos13030475

Graça, D., Reis, J., Gama, C., Monteiro, A., Rodrigues, V., Rebelo, M., Borrego, C., Lopes, M., & Miranda, A. I. (2023). Sensors Network as an Added Value for the Characterization of Spatial and Temporal Air Quality Patterns at the Urban Scale. *Sensors*, *23*(4). https://doi.org/10.3390/s23041859

Hayward, I., Martin, N. A., Ferracci, V., Kazemimanesh, M., Jude, S., Walton, C., Nasir, Z. A., Kumar, P. (2025). Comprehensive comparison of correction techniques for low-cost air quality sensors: the impact of device type and deployment environment. *npj Climate and Atmospheric Science,* 8, 389. https://doi.org/10.1038/s41612-025-01231-5

Hayward, I., Martin, N. A., Ferracci, V., Kazemimanesh, M., & Kumar, P. (2024). Low-Cost Air Quality Sensors: Biases, Corrections and Challenges in Their Comparability. *Atmosphere*, *15*(12), 1523. https://doi.org/10.3390/atmos15121523

Janeja, V. P., Adam, N. R., Atluri, V., & Vaidya, J. (2010). Spatial neighborhood based anomaly detection in sensor datasets. *Data Mining and Knowledge Discovery*, *20*(2), 221–258. https://doi.org/10.1007/s10618-009-0147-0

Kamal, S., Ramadan, R., & El-Refai, F. (2016). Smart outlier detection of wireless sensor network. *Facta Universitatis - Series: Electronics and Energetics*, *29*(3), 383–393. https://doi.org/10.2298/FUEE1603383K

Kaskaoutis, D. G., Petrinoli, K., Grivas, G., Kalkavouras, P., Tsagkaraki, M., Tavernaraki, K., Papoutsidaki, K., Stavroulas, I., Paraskevopoulou, D., Bougiatioti, A., Liakakou, E., Rashki, A., Sotiropoulou, R. E. P., Tagaris, E., Gerasopoulos, E., & Mihalopoulos, N. (2024). Impact of peri-urban forest fires on air quality and aerosol optical and chemical properties: The case of the August 2021 wildfires in Athens, Greece. *Science of The Total Environment*, *907*, 168028. https://doi.org/10.1016/j.scitotenv.2023.168028

Keshtkar, M., Heidari, H., Moazzeni, N., & Azadi, H. (2022). Analysis of changes in air pollution quality and impact of COVID-19 on environmental health in Iran: application of interpolation models and spatial autocorrelation. *Environmental Science and Pollution Research*, *29*(25), 38505–38526. https://doi.org/10.1007/s11356-021-17955-9

Levy Zamora, M., Xiong, F., Gentner, D., Kerkez, B., Kohrman-Glaser, J., & Koehler, K. (2019). Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environmental Science and Technology*, *53*(2), 838–849. https://doi.org/10.1021/acs.est.8b05174

Liang, C.-J., & Yu, P.-R. (2021). Assessment and Improvement of Two Low-Cost Particulate Matter Sensor Systems by Using Spatial Interpolation Data from Air Quality Monitoring Stations. *Atmosphere*, *12*(3), 300. https://doi.org/10.3390/atmos12030300

Masoom, A., Fountoulakis, I., Kazadzis, S., Raptis, I.-P., Kampouri, A., Psiloglou, B. E., Kouklaki, D., Papachristopoulou, K., Marinou, E., Solomos, S., Gialitaki, A., Founda, D., Salamalikis, V., Kaskaoutis, D., Kouremeti, N., Mihalopoulos, N., Amiridis, V., Kazantzidis, A., Papayannis, A., Zerefos, C. S., & Eleftheratos, K. (2023). Investigation of the effects of the Greek extreme wildfires of August 2021 on air quality and spectral solar irradiance. *Atmospheric Chemistry and Physics, 23*(14), 8487–8514. https://doi.org/10.5194/acp-23-8487-2023

Miao, L., Liu, C., Yang, X., Kwan, M. P., & Zhang, K. (2022). Spatiotemporal heterogeneity analysis of air quality in the Yangtze River Delta, China. *Sustainable Cities and Society*, *78*. https://doi.org/10.1016/j.scs.2021.103603

Munir, S., Mayfield, M., Coca, D., Jubb, S. A., & Osammor, O. (2019). Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case study in Sheffield. *Environmental Monitoring and Assessment*, *191*(2), 94. https://doi.org/10.1007/s10661-019-7231-8

Ottosen, T.-B., & Kumar, P. (2019). Outlier detection and gap filling methodologies for low-cost air quality measurements. *Environmental Science: Processes & Impacts*, *21*(4), 701–713. https://doi.org/10.1039/C8EM00593A

Papayiannis, G. I., Psarakis, S., & Yannacopoulos, A. N. (2023). Modelling of Functional Profiles and Explainable Shape Shifts Detection: An Approach Combining the Notion of the Fréchet Mean with the Shape-Invariant Model. *Mathematics*, *11*(21), 4466. https://doi.org/10.3390/math11214466
Poornima, I. G. A., & Paramasivan, B. (2020). Anomaly detection in wireless sensor network using machine learning algorithm. *Computer Communications*, *151*, 331–337. https://doi.org/10.1016/j.comcom.2020.01.005

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. *Environmental Health Perspectives*, *124*(9), 1334–1343. https://doi.org/10.1289/ehp.1409277
Sánchez-Lasheras, F., Ordóñez-Galán, C., García-Nieto, P. J., & García-Gonzalo, E. (2020). Detection of outliers in pollutant emissions from the Soto de Ribera coal-fired power plant using functional data analysis: a case study in northern Spain. *Environmental Science and Pollution Research*, *27*(1), 8–20. https://doi.org/10.1007/s11356-019-04435-4
Slongo, J., Lindino, C., Martins, L. D., Spanhol, F. A., Carneiro, E., & Camargo, E. T. (2024). Evaluation of low-cost sensors to integrate in a water quality monitor for real-time measurements. *Environmental Monitoring and Assessment*, *196*(8), 716. https://doi.org/10.1007/s10661-024-12884-9
Tagle, M., Rojas, F., Reyes, F., Vásquez, Y., Hallgren, F., Lindén, J., Kolev, D., Watne, Å. K., & Oyola, P. (2020). Field performance of a low-cost sensor in the monitoring of particulate matter in Santiago, Chile. *Environmental Monitoring and Assessment*, *192*(3). https://doi.org/10.1007/s10661-020-8118-4
Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.
Wei, Y., Jang-Jaccard, J., Sabrina, F., & Alavizadeh, H. (2020). Large-Scale Outlier Detection for Low-Cost PM_{10} Sensors. *IEEE Access*, *8*, 229033–229042. https://doi.org/10.1109/ACCESS.2020.3043421
Xin, L., Shaoliang, Z., & Pulin, Z. (2015). Spatial Outlier Detection of CO_2 Monitoring Data Based on Spatial Local Outlier Factor. *Journal of Engineering Science and Technology Review, 8*(5), 110-116. https://doi.org/10.25103/jestr.085.15
Zafeirelli, S., & Kavroudakis, D. (2024). Comparison of outlier detection approaches in a Smart Cities sensor data context. *International Journal on Smart Sensing and Intelligent Systems*, *17*(1). https://doi.org/10.2478/ijssis-2024-0004
Zhang, Y., Hamm, N. A. S., Meratnia, N., Stein, A., van de Voort, M., & Havinga, P. J. M. (2012). Statistics-based outlier detection for wireless sensor networks. *International Journal of Geographical Information Science*, *26*(8), 1373–1392. https://doi.org/10.1080/13658816.2012.654493
Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: A survey. *IEEE Communications Surveys and Tutorials*, *12*(2), 159–170. https://doi.org/10.1109/SURV.2010.021510.00088
Zusman, M., Schumacher, C. S., Gassett, A. J., Spalt, E. W., Austin, E., Larson, T. v., Carvlin, G., Seto, E., Kaufman, J. D., & Sheppard, L. (2020). Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment International*, *134*. https://doi.org/10.1016/j.envint.2019.105329

https://doi.org/10.48088/ejg.s.zaf.17.1.212.230