



## **Harmonization and Visualization of Data from a Transnational Multi-Sensor Personal Exposure Campaign**

Rok Novak, Ioannis Petridis, David Kocman, Johanna Amalia Robinson, Tjaša Kanduč, Dimitris Chapizanis, Spyros Karakitsios, Benjamin Flückiger, Danielle Vienneau, Ondřej Mikeš, et al.

### **► To cite this version:**

Rok Novak, Ioannis Petridis, David Kocman, Johanna Amalia Robinson, Tjaša Kanduč, et al.. Harmonization and Visualization of Data from a Transnational Multi-Sensor Personal Exposure Campaign. International Journal of Environmental Research and Public Health, 2021, 18, <10.3390/ijerph18211614>. <hal-03415993>

**HAL Id: hal-03415993**

**<https://hal.science/hal-03415993v1>**

Submitted on 5 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Article

# Harmonization and Visualization of Data from a Transnational Multi-Sensor Personal Exposure Campaign

Rok Novak <sup>1,2,\*</sup>, Ioannis Petridis <sup>3</sup>, David Kocman <sup>1</sup>, Johanna Amalia Robinson <sup>1,2</sup>, Tjaša Kanduč <sup>1</sup>, Dimitris Chapizanis <sup>3</sup>, Spyros Karakitsios <sup>3,4</sup>, Benjamin Flückiger <sup>5,6</sup>, Danielle Vienneau <sup>5,6</sup>, Ondřej Mikeš <sup>7</sup>, Céline Degrendele <sup>7,8</sup>, Ondřej Sánka <sup>7</sup>, Saul García Dos Santos-Alves <sup>9</sup>, Thomas Maggos <sup>10</sup>, Demetra Pardali <sup>10</sup>, Asimina Stamatelopoulou <sup>10</sup>, Dikaia Saraga <sup>10</sup>, Marco Giovanni Persico <sup>11,12</sup>, Jaideep Visave <sup>11,12</sup>, Alberto Gotti <sup>12</sup> and Dimosthenis Sarigiannis <sup>3,4,11</sup>

- <sup>1</sup> Department of Environmental Sciences, Jožef Stefan Institute, 1000 Ljubljana, Slovenia; david.kocman@ijs.si (D.K.); johanna.robinson@ijs.si (J.A.R.); tjas.kanduc@ijs.si (T.K.)
- <sup>2</sup> Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia
- <sup>3</sup> Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; ioannis.petridis89@gmail.com (I.P.); dimitris.chapizanis@gmail.com (D.C.); spyros.karakitsios@gmail.com (S.K.); denis@eng.auth.gr (D.S.)
- <sup>4</sup> HERACLES Research Centre on the Exposome and Health, Center for Interdisciplinary Research and Innovation, 54124 Thessaloniki, Greece
- <sup>5</sup> Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, CH-4051 Basel, Switzerland; benjamin.flueckiger@swisstph.ch (B.F.); danielle.vienneau@swisstph.ch (D.V.)
- <sup>6</sup> University of Basel, CH-4001 Basel, Switzerland
- <sup>7</sup> RECETOX, Faculty of Science, Masaryk University, 62500 Brno, Czech Republic; ondrej.mikes@recetox.muni.cz (O.M.); celine.DEGRENDELE@univ-amu.fr (C.D.); ondrej.sanka@recetox.muni.cz (O.S.)
- <sup>8</sup> LCE, CNRS, Aix-Marseille University, 13003 Marseille, France
- <sup>9</sup> Department of Atmospheric Pollution, Instituto de Salud Carlos III, Instituto de Salud Carlos III (ISCIII), 28220 Madrid, Spain; sgarcia@isciii.es
- <sup>10</sup> Atmospheric Chemistry and Innovative Technologies Laboratory, INRASTES, NCSR “Demokritos”, Aghia Paraskevi, 15310 Athens, Greece; tmaggos@ipta.demokritos.gr (T.M.); demetra.pard@gmail.com (D.P.); mina.stam@ipta.demokritos.gr (A.S.); dsaraga@ipta.demokritos.gr (D.S.)
- <sup>11</sup> Department of Science, Technology and Society, University School of Advanced Study IUSS, 27100 Pavia, Italy; marco.persico@iusspavia.it (M.G.P.); jaideep.visave@eucentre.it (J.V.)
- <sup>12</sup> Eucentre Foundation, Via A. Ferrata, 1, 27100 Pavia, Italy; gottial@gmail.com
- \* Correspondence: rok.novak@ijs.si

**Citation:** Novak, R.; Petridis, I.; Kocman, D.; Robinson, J.A.; Kanduč, T.; Chapizanis, D.; Karakitsios, S.; Flückiger, B.; Vienneau, D.; Mikeš, O.; et al. Harmonization and Visualization of Data from a Transnational Multi-Sensor Personal Exposure Campaign. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11614. <https://doi.org/10.3390/ijerph182111614>

Academic Editors: Tareq Hussein, Martha A. Zaidan and Paul B. Tchounwou

Received: 28 September 2021

Accepted: 1 November 2021

Published: 4 November 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Use of a multi-sensor approach can provide citizens with holistic insights into the air quality of their immediate surroundings and their personal exposure to urban stressors. Our work, as part of the ICARUS H2020 project, which included over 600 participants from seven European cities, discusses the data fusion and harmonization of a diverse set of multi-sensor data streams to provide a comprehensive and understandable report for participants. Harmonizing the data streams identified issues with the sensor devices and protocols, such as non-uniform timestamps, data gaps, difficult data retrieval from commercial devices, and coarse activity data logging. Our process of data fusion and harmonization allowed us to automate visualizations and reports, and consequently provide each participant with a detailed individualized report. Results showed that a key solution was to streamline the code and speed up the process, which necessitated certain compromises in visualizing the data. A thought-out process of data fusion and harmonization of a diverse set of multi-sensor data streams considerably improved the quality and quantity of distilled data that a research participant received. Though automation considerably accelerated the production of the reports, manual and structured double checks are strongly recommended.

**Keywords:** data fusion; multi-sensor; data visualization; data treatment; participant reports; air quality; exposure assessment

## 1. Introduction

The health impacts of poor air quality have become a central point of discussion in policy development and in personal exposure studies [1–3]. A growing selection of low-cost sensors (LCSs) that measure environmental conditions allow individuals to collect data about their own living environment and estimate their exposure to different stressors [4–6]. Several issues remain regarding bulkiness, design, power consumption, data loss [7], unreliable and (unintentionally) misleading data, lack of quality control, validation and calibration [8], and user experience [9]. Providing meaningful information to individuals about their environment and related stressors is in line with the United Nations Sustainable Development Goals (SDGs) calling for participatory, integrated, and sustainable human settlement planning (Target 11.3 [10]), which can only be achieved if the public is well-informed. Several goals and targets in the SDGs are assessed based on the “Mean urban air pollution of particulate matter (PM) of different sizes” indicator [11]. Considering the often-low spatial resolution of PM measurements (i.e., at a city level), typically only sampling outdoor air pollution, the use of individual low-cost PM sensors could be useful in estimating human exposure to PM.

Airborne particulate matter concentration is only one facet of air quality, and when assessing the impact of air quality on human health, pollutants such as nitrogen dioxide (NO<sub>2</sub>) [12], ozone (O<sub>3</sub>) [13], and volatile organic compounds (VOCs) [14,15] should be considered. Elevated levels of indoor carbon dioxide (CO<sub>2</sub>) concentrations can also pose health risks [16].

Data fusion techniques combine data from multiple source (i.e., sensors) and related information from databases to obtain more consistent, accurate, and useful information than can be obtained by the use of a single sensor alone, including fusing features and data to support decisions [17]. Fusing data from different low-cost sensors has previously been employed to supplement existing datasets from environmental monitoring networks with high-resolution spatiotemporal measurements from LCSs [18,19], by using mobile LCSs for air quality mapping in combination with dispersion model calculations [20] or by using stationary data with transport model results [21]. This enables the efficient integration of data derived from multiple sources at different stages of analysis and visualization.

An increase in the availability of devices with very diverse input parameters and data collection protocols poses some unique data fusion and visualization challenges, including non-standard timestamps, data gaps, different classifications, a multitude of data logging processes, etc. While LCSs generally provide a larger quantity of data, there is a lack of data on comparability from one device to another. Good metadata and documentation on how data are recorded and presented can help researchers make informed decisions and better comprehend potential issues prior to using the sensor [22]. The reliability and accuracy of LCSs may necessitate validation/calibration prior to use. Such processes are not standardized and can vary from device to device. The results are usually presented using the correlation coefficient, root mean square error, and mean absolute error, which, while useful, must be accompanied with information regarding the conditions under which the validation/calibration was performed [23]. In turn, this makes the process of data fusion and visualization more straightforward.

To facilitate data fusion and visualization, where one of the goals is to provide meaningful information to participants, there should be a greater focus on assessing the characteristics of the sensor itself, providing more context and associated uncertainties (where available) [22]. A benefit of participatory approaches, where citizens use LCSs, is the ability to gain additional (qualitative) information from the user through interviews or smartphone surveys [24] about specific environmental conditions to inform data fusion. Another benefit is the ability to obtain information about how well the sensors function.

Preparing visualizations of data for lay end users requires a balance in providing the most relevant data in an understandable way. Selecting the proper type of visualization can have a meaningful impact on the perception of the end user and the information that

they are able to extract [25,26], and promote better risk assessment and reduction in exposure due to personal decision making [27]. An improvement, which is already being employed in some visualization efforts, is the ability of users to interact with the final dataset and make their own adjustments [28].

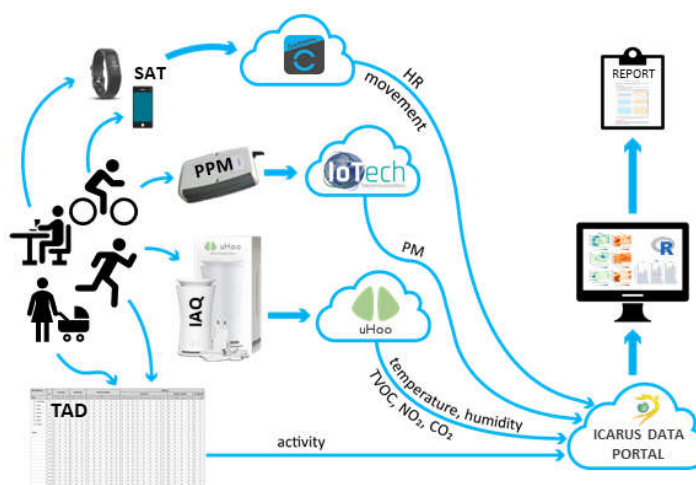
Collecting data from multi-sensor and multi-parameter data flows from hundreds of individuals involved in an exposure campaign produced unique issues and challenges which this paper specifically addresses. A key objective was to produce an aggregated and harmonized dataset that allowed for an efficient way of visualizing data via data fusion. Additionally, it provides a starting point for numerous individual-level and community-level exposure assessments and further data analysis, which will be explored in future research. An algorithm was developed that would clean, fuse, and visualize the collected data and present them to the participants in a straightforward and understandable report. This “final report” for the participants was generated in their respective local language, and included as much data as possible without making the report too long and complicated. The report aimed to provide enough details for participants to discern relevant information related to their local air quality, living environment, and behavior, with a view to eventually promote more environmental conscious lifestyles.

Specifically, the objectives of this study were to provide insights and specifics on:

- Outputs resulting from multi-sensor and multi-parameter data flows;
- Aggregation and harmonization of data collected;
- Production of tailored visualizations by fusing data from multiple sources, and automated compilation of individualized final reports.

## 2. Materials and Methods

Input data for data fusion and visualization were obtained from three sensor devices, data collected through questionnaires for households and individuals, and time activity diaries (TADs). They were part of the Integrated Climate forcing and Air pollution Reduction in Urban Systems (ICARUS) H2020 project, which applied integrated tools and strategies for urban impact assessment in support of air quality and climate change governance [29,30]. For this purpose, about 100 participants were recruited in each of the seven selected European cities—Athens, Basel, Brno, Ljubljana, Madrid, Milano, and Thessaloniki—and were provided with all the tools required to collect the necessary data. The data were collected in two seasons in 2019—heating (winter) and non-heating (summer)—to observe any differences between the seasons, as the use of heating devices might influence air quality [31,32]. Two of the sensor devices were commercial: a smart activity tracker (SAT) and an indoor air quality (IAQ) sensing station. The third, called a personal particulate matter (PPM) sensing device, was specifically constructed for the purposes of the research project using the Arduino platform. A schematic representation of the devices and protocols used is shown in Figure 1. A detailed description of the campaign and its goals can be found in Robinson et al. [33]. All data cleaning, harmonization, fusion, visualization, and report compilation and output were done in R [34] with support from different R packages, e.g., ggplot2 [35], dplyr [36], knitr [37], and rmarkdown [38].



**Figure 1.** Schematic representation of data collection devices and protocols, transfer paths, aggregation, visualization, and delivery protocols.

### 2.1. PPM Data

The PPM sensing device (IoTech Telecommunications, Thessaloniki, Greece) was designed for the purposes of the sampling campaign of the ICARUS project [39]. It collected PM concentration data in three class sizes ( $<1 \mu\text{m}$  ( $\text{PM}_{10}$ ),  $<2.5 \mu\text{m}$  ( $\text{PM}_{2.5}$ ), and  $<10 \mu\text{m}$  ( $\text{PM}_{10}$ )) and ambient temperature as well as relative humidity data, in addition to GPS/location coordinates (including speed and altitude). As the device did not have a real-time clock (RTC) module (e.g., [40]), the timestamp was obtained by connecting it to an online server via a SIM card. Without this connection, the device did not provide data with accurate timestamps, which in turn produced several data gaps. Timestamp logging was irregular and inconsistent, as evident in an example of the dataset in section A of the Supplementary Data (SD-A).

### 2.2. SAT Data

A commercial SAT was used (Vivosmart 3, manufactured by Garmin, Olathe, Kansas, U.S. [41]) to collect heart rate and movement data with a minute resolution (e.g., average heart rate, stress level, sleep status, calories burned, etc.). As the export of data is not freely available through the Garmin interface, an additional connection between a dedicated ICARUS data portal and the Garmin Connect portal was established to transfer the data. The SAT data had very few gaps (excluding the time while the device was charging). Issues with data capture of heart rate occurred when the user did not fasten the wrist strap tight enough.

A brief overview of the SAT data was included in the final report as a summary table.

### 2.3. IAQ Data

A “uHoo Smart Indoor Air Quality (IAQ) sensor” (uHoo Limited, Singapore) [42], a stationary device with multiple sensors, was used in every household. At every full minute, the IAQ provided data on temperature, relative humidity,  $\text{CO}_2$ , total VOCs (TVOCs),  $\text{PM}_{2.5}$ ,  $\text{NO}_2$ , carbon monoxide (CO), ozone ( $\text{O}_3$ ), and air pressure.

Visualization of AQ parameters measured by the IAQ was limited to three parameters ( $\text{CO}_2$ ,  $\text{NO}_2$ , and TVOCs) that showed the best performance during the collocation experiments with validated devices, as well as other tests. As offsets were observed for some sensors during these experiments, this specific dataset was visualized using a heatmap, focusing on relative changes in each variable over time. A heatmap, in this case, consists of tiles which are colored relative to all other tiles (lower values are lighter, higher values are darker), as implemented in Mahajan et al. [43]. Using minute values would create a

heat map with small tiles, which would obscure the relative differences within a day. To counteract this, hourly values were calculated and used in the heat map, reducing the number of tiles from approximately 10,000 to 170.

#### 2.4. ICARUS Data Portal

A dedicated data portal was constructed for the purposes of the ICARUS2020 project, and a decision support system (DSS) with it, which collected, compiled, and stored the data. The DSS additionally had a presentation tier with a user interface and a logic tier that stored the computational models and handled their execution [44]. In this study, the data portal was mainly utilized to store and obtain the PPM and SAT data in a uniform format, which allowed further manipulation and fusion of data.

#### 2.5. TAD Data

A key data input was the TADs, which allowed the participant to record their activity, location, means of transport, and other variables for each hour of the day. These data were collected from each participant, for seven days in two seasons, for all cities, accumulating up to approximately 10,000 TADs.

There were two methods of filling in the TADs: one was to select only one option for each hour (i.e., majority activity) and the other was to allow participants to select multiple options within an hour. This posed a unique challenge in selecting which data point to use, which activity was more relevant or more characteristic for each hour.

Some manual corrections of the data were necessary in the final stages after observing some obvious mistakes in the recording of activity. As these corrections were not double-checked with the participants, only the most obvious mistakes were corrected, e.g., if a non-smoking person truly smoked in just one instance the entire period.

Because the data for activities were for hourly values and the sensor data had a minute resolution, the former was repeated 60 times per hour, which proved to be a major issue when calculating averages and trying to discern if there were meaningful differences between activities [45].

The TAD dataset was used in three visualizations, in combination with PPM and SAT data:

- (a) A scatter plot was made for every PM size class and heart rate for both seasons. Additionally, the points were colored based on the activity at that minute, which allowed the reader to observe what activities took place at, for example, elevated levels of PM or elevated heart rate. Only the activities which the participant filled in were shown in the legend.
- (b) A similar scatter plot as in (a) was constructed, with an additional layer which showed vertical bands or ribbons of different colors corresponding with the participant's location and mode of transport. As this added another layer of complexity to the visualization, the decision was made to provide these plots only to specific individuals who expressed interest. Though activity information was missing in several TADs, the location and transport data were logged for almost the entire period of observation (for most participants). Consequently, participants could associate specific means of transport with elevated levels of PM, and corresponding activities with a higher heart rate.
- (c) The third plot showed the average weekly PM values for each activity. Six plots were constructed, three per season, one for each PM size class.

TAD data were not used in combination with IAQ data due to the higher uncertainty associated with absolute values of CO<sub>2</sub>, NO<sub>2</sub>, and TVOCs.

#### 2.6. Final Report Compilation and Production

The generation of final reports for participants was performed in three phases:

- (a) Generation of plots as described in points 2.1.–2.4., which was followed for all of the participants. These plots were saved locally in a jpeg format and labeled according to each participant ID.
- (b) Plots were integrated in a rmarkdown script, with the customization of each report designated in an Excel file. Each participant had a custom greeting with their name and gender-appropriate pronoun. All plots and other graphics were inserted using the `include_graphics` function in the knitr package.
- (c) Finally, the script was iterated over all participants in a separate script to allow some further customizations. Some participants had additional visualizations (see 2.5 point b), while others had some omitted due to missing data. After all the reports were generated in the participants' local language, they were manually checked for errors by local organizers in each participating city and distributed to all the participants.

In addition to the technical construction and production of the final report, the participant feedback and wishes for visualization were considered, where appropriate, by employing a user-centered approach and a structured focus group of participants [46].

### 2.7. Temporal Resolution and Data Treatment

A minute resolution of data was deemed as sufficient to provide enough detail of PM concentrations and exposure. The SAT and IAQ also logged data with a minute resolution, though these logs were at every full minute while the PPM sensing device logged the measurements at different fractions of the minute. These were later rounded to the nearest minute.

To compare the PM data with WHO guideline limits the minute resolution data were aggregated into daily means. More uncertainty was associated with PM daily means from the first and last day the participant was involved in the campaign, as the participants did not collect data for the entire 24 h period.

An outlier correction was performed for the PM data, where all values above 180  $\mu\text{g}/\text{m}^3$  were set to 180  $\mu\text{g}/\text{m}^3$ , based on the maximum values provided by “Air quality in Europe” as part of the CITEAIR and CITEAIR II projects [47]. This approach was used only for visualizing the data and providing the final reports to participants for clearer data representation.

The PPM data showed good agreement with absolute values measured from a reference research-grade device, a GRIMM Model 11-A, and increasingly did so with larger time-averaging intervals [39].

## 3. Results and Discussion

### 3.1. A Merged Dataset

The final merged dataset had 93 columns. Due to sensor failures, data gaps, incorrect TAD filling, etc., there were several instances of empty columns or, in some cases, completely empty datasets per participant. This was appropriately labeled in the final reports.

Section B of the Supplementary Data (SD-B) presents an example of a completed dataset, with all the data harmonized to a 1 min resolution. Each dataset includes:

- Specific characteristics for each participant (age and gender);
- PPM data (PM values, temperature, humidity, battery charge level, location coordinates, speed, and altitude);
- SAT data (where several columns proved to be somewhat redundant and were therefore removed);
- IAQ data (which proved to be easiest to handle as they had a correct timestamp for each recorded value, almost no missing values, and a simple interface to download the data);
- TAD data, presented the same way as they were recorded on the physical paper sheets: location of the participant (home, office, indoor, outdoor), transport data (bus, car, foot, etc.), indoor and outdoor activities (cooking, smoking, sports, etc.), and

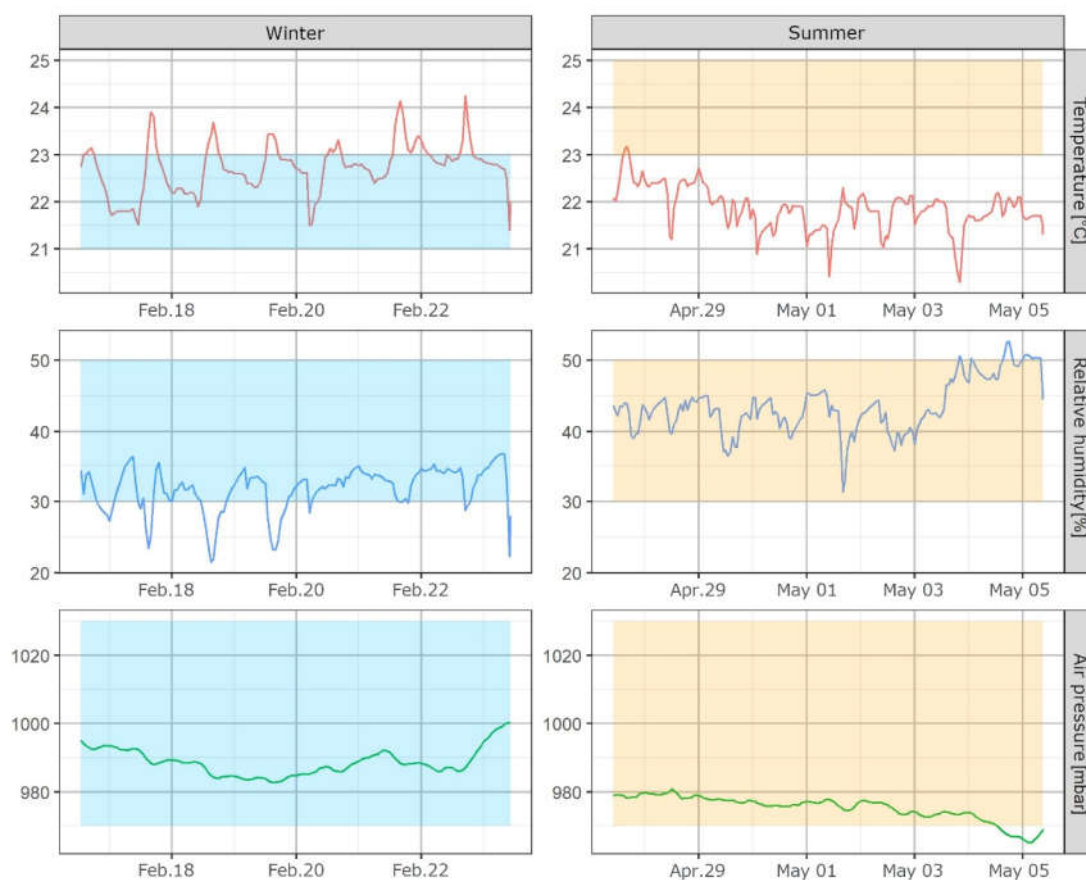
some specific conditions for the indoor space the participant was in (burning candle or fireplace, open windows, and/or AC turned on).

### 3.2. Visualizing the Data

All the visualizations are presented and described here as they were shown in the final report to participants, and are collected as examples from different participants.

Figure 2 shows the temperature, relative humidity, and air pressure during both seasons (IAQ data). Non-heating and heating seasons are indicated as “Winter” and “Summer”, respectively, as the sampling campaign for all the cities extended from January to March 2019 for the heating season and from April to July 2019 for the non-heating season. The example visualizations are only from reports to participants from the city of Ljubljana, where the non-heating sampling campaign took place earlier than that in other cities and was therefore defined as summer. Meteorological data showed the highest accuracy when compared to reference instruments, and were in turn presented with absolute values. Although the ribbons show “optimal conditions” as per the general health and comfort guidelines (modified for the appropriate climate) [36], this information is somewhat subjective and can differ from person to person. As shown in the example in Figure 2, this person had very similar indoor temperatures in both seasons, and even though the summer values are mostly outside the “optimal zone”, one could argue that a constant temperature throughout the year provides more comfort to certain individuals.

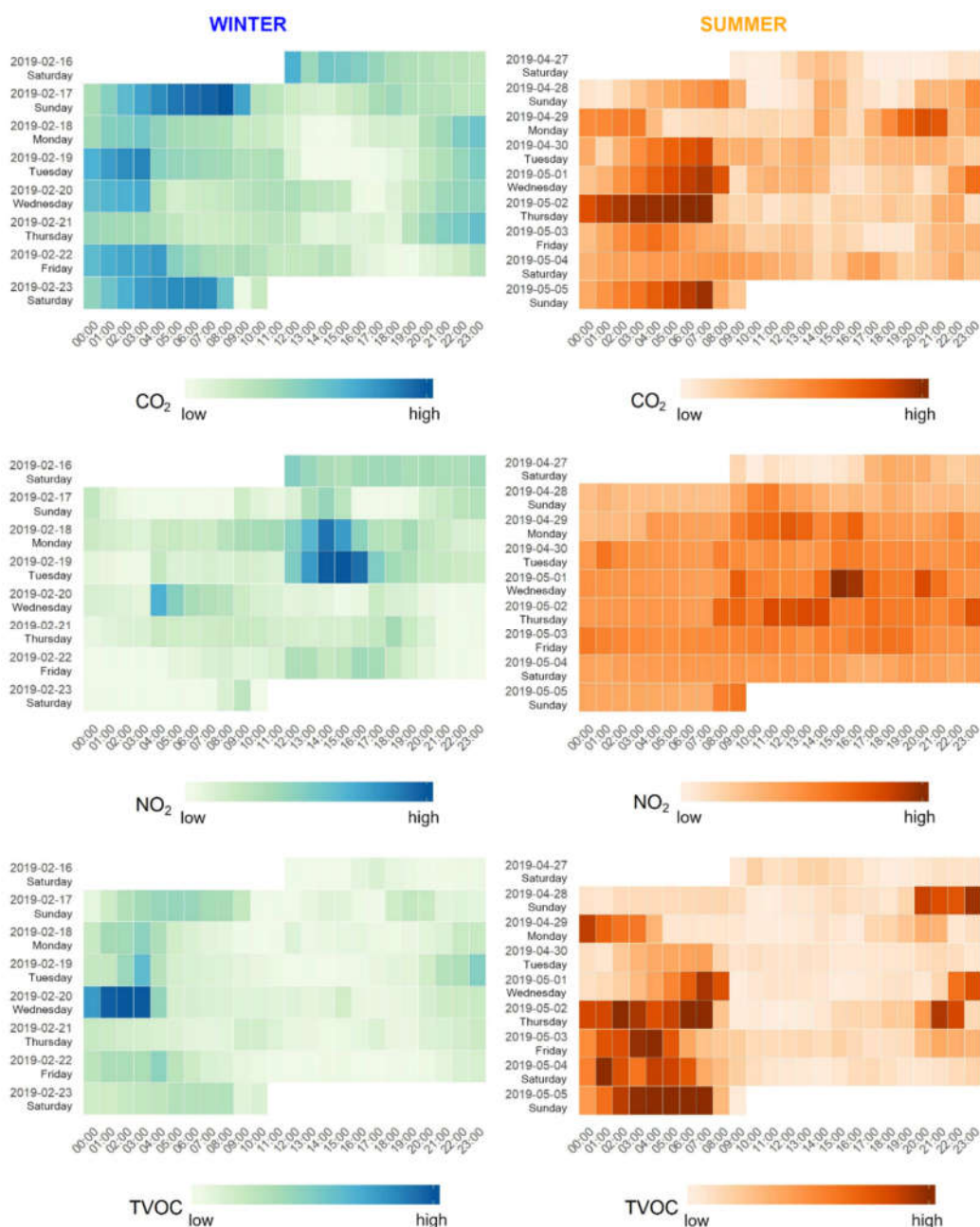
Arranging the individual plots into columns according to season makes comparisons between the seasons easier.



**Figure 2.** Faceted plots with meteorological variables—temperature, relative humidity, and air pressure; data from IAQ. Colored horizontal ribbons represent “optimal” values for each variable.



Figure 3 shows an example of the compiled visualizations of CO<sub>2</sub>, NO<sub>2</sub>, and TVOCs for this particular household. We found that these parameters typically followed expected trends, e.g., decreased values of CO<sub>2</sub> when opening a window and in turn increasing the NO<sub>2</sub> values if it was in a high-traffic area [48], as seen in Figure 3 on Tuesday the 19<sup>th</sup> of February 2019 at around 13:00, when CO<sub>2</sub> concentrations quickly fell and NO<sub>2</sub> increased rapidly at the same time. The plots allow for an intuitive way of observing relative changes in these parameters by household. We used relative values in our reports, as collocation with a reference device has previously shown that the absolute values were not accurate enough to present to participants at that time [49], though newer research shows moderate to high correlation with reference instruments in laboratory conditions [50]. These relative values still give participants an insight into their indoor air quality and possible correlations with external factors such as traffic.



**Figure 3.** Faceted heatmaps of three pollutants ( $\text{CO}_2$ ,  $\text{NO}_2$ , and TVOCs); data from IAQ.

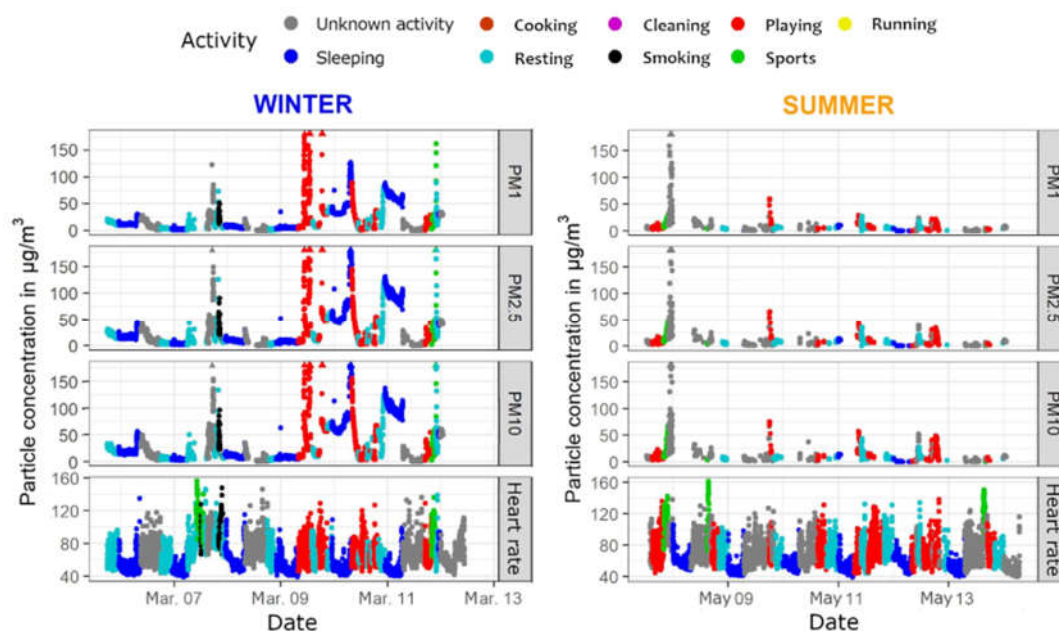
The layout of the visualization allows the reader to compare trends between seasons and between pollutants. For example, higher TVOC values during the evening and night could indicate poor ventilation in combination with a specific activity that raises the concentrations, such as cooking or smoking [51]. By putting these plots in the same figure, they can immediately observe the trends in the other two parameters and come to some conclusions.

Each date is also labeled with the written day of the week (language-specific) to facilitate better observation of specific trends.

Figure 4 presents the concentrations of PM in three class sizes, heart rate, and designated activities for each minute during which the participant was involved in the data collection. Only the specific activities are shown; there is no additional information about the location of the participant, their mode of transport, or specific conditions in the household. Not including this information makes the visualizations less crowded and easier to read and understand, as determined by exchanges provided in a structured focus group [46]. All the values are also plotted with exact concentrations, given that the PPM device showed fairly accurate results compared to reference devices.

The participants could deduce by themselves some interpretations and extra information from the plots, e.g., a higher heart rate when running, dips during the night, a specific time of day when the PM concentrations were elevated and if they were perhaps related to a specific activity such as smoking or cooking, etc. This level of interpretation is only feasibly possible by the participants, because they would have a more complete overview of their surrounding and activities. To avoid the issue of recall, participants could be provided with their respective TADs.

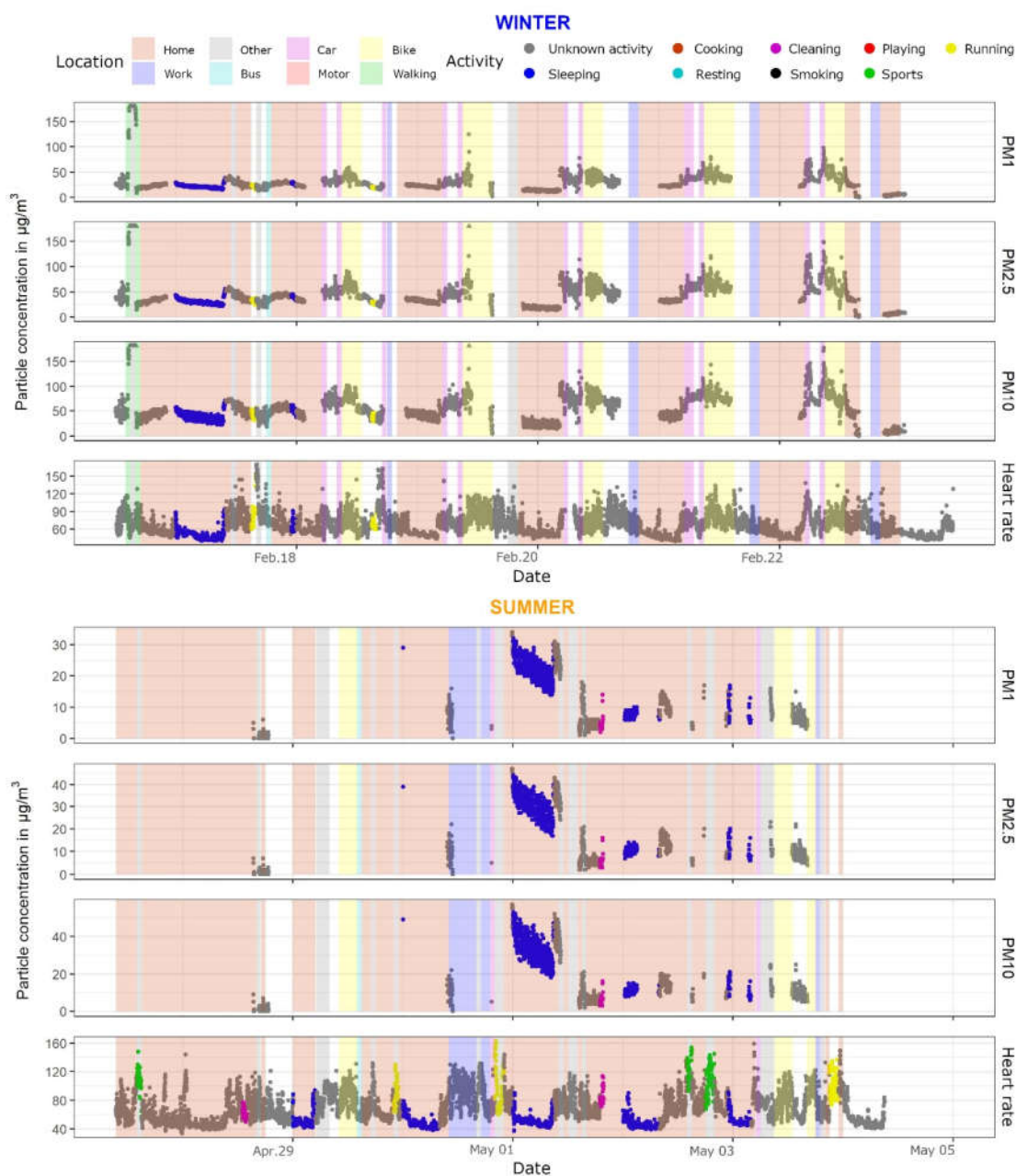
No particular difficulties were encountered with constructing this visualization, with the possible exception of some alterations to the color scale and legend to also include the activities that the participant did not perform.



**Figure 4.** Three size classes of PM ( $\text{PM}_1$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$ ) and heart rate values for both seasons with each point colored according to the associated activity; data from PPM, SAT, and TAD.

An additional figure was created to include the location and transport of the participants, in addition to PM and activity values (Figure 5).

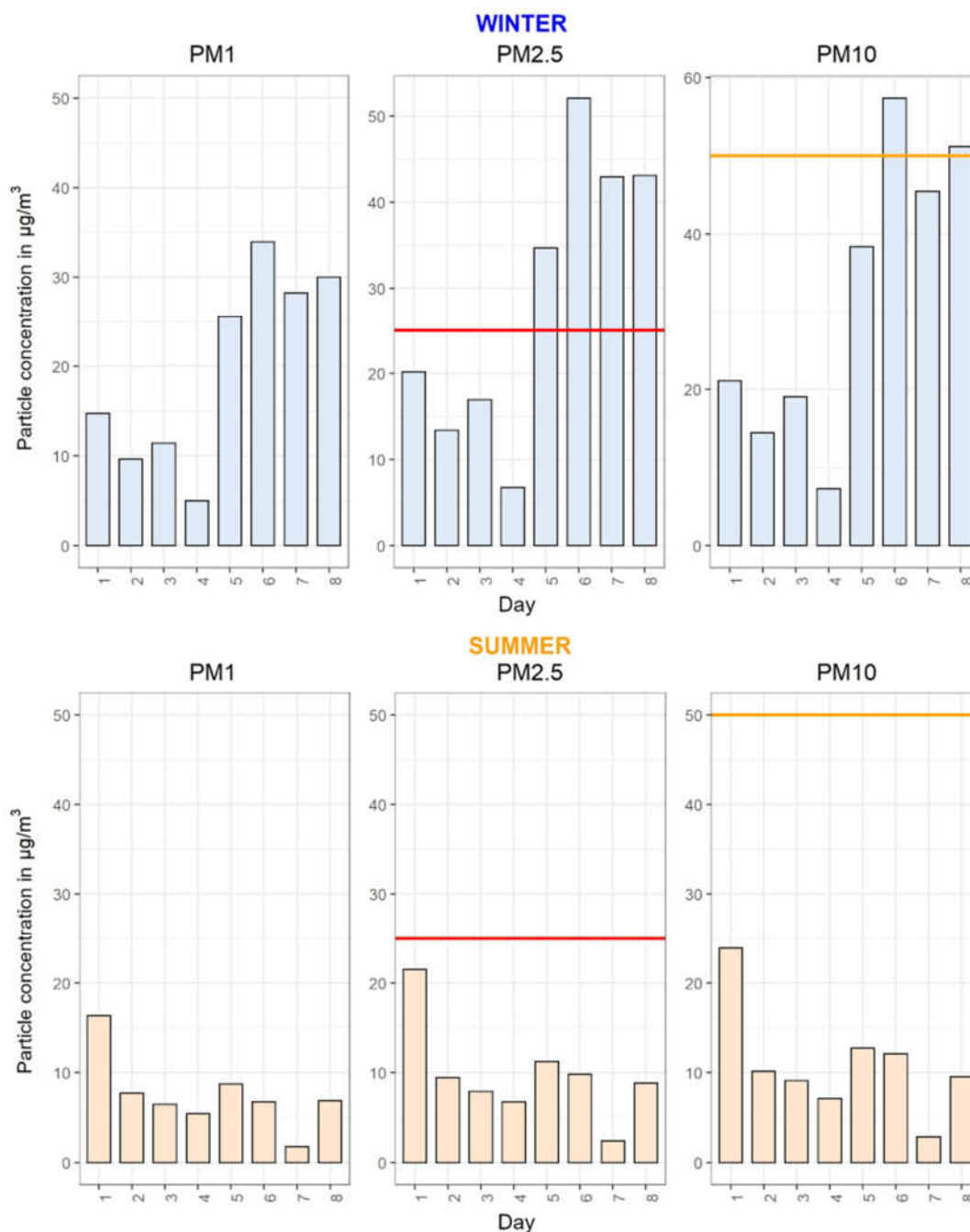
Several difficulties were encountered while constructing these plots, as the ribbons that show each activity needed a start and an end time for each location/transport at every interval. We considered only including a vertical line at each minute in the color of the location/transport, but this considerably increased the run-time per plot and was not efficient given the large number of plots to produce. An additional section of code was implemented to construct a separate data frame which had a start and an end time with a label for each location/transport. This was used in the ggplot2 geom\_rect function while compiling the plot and noticeably reduced the time it took to compile each plot.



**Figure 5.** Three size classes of PM and heart rate values for both seasons with each point colored according to the associated activity and each ribbon representing a location or means of transport for that time period; data from PPM, SAT, and TAD.

Figure 6 shows the daily average PM concentrations for both seasons, and is the only set of plots where guidelines or recommended values could be inserted. The WHO and

the EU do not have minute or hourly guidelines for concentrations of PM, though studies show that short-term exposure to elevated levels can have adverse effects on health [52,53]. The WHO does provide daily guidelines for PM<sub>2.5</sub> and PM<sub>10</sub>, which are 25  $\mu\text{g}/\text{m}^3$  and 50  $\mu\text{g}/\text{m}^3$ , respectively [54], revised in 2021 to 15  $\mu\text{g}/\text{m}^3$  and 45  $\mu\text{g}/\text{m}^3$ , respectively [55].



**Figure 6.** Faceted plots of average daily concentrations of three size classes of PM for each season, with WHO guidelines; data from PPM.

There are two important pieces of information in these plots, allowing the participant to observe (1) inter-seasonal differences and (2) day-to-day differences, while also having information about a specific size class of PM. This specific plot shows that the

concentrations are generally higher in wintertime (more indoor activities, weather patterns that trap pollution in low-lying areas, combustion of solid fuels, more use of car/buses in contrast with cycling/walking, etc.), and when there are elevated levels of PM during the summer they are still much lower than in wintertime. The participant can also observe that some particular days have elevated levels of PM, which could be associated with specific activities performed that day (or weather patterns).

Figure 7 presents two tables showing the average values for each SAT variable for each day. No additional visualizations were made for the SAT data (apart from the heart rate plots in Figures 4 and 5). There were several visualizations already available on the Garmin Connect portal for each variable.

Day	Steps	Stress level	Average heart rate	Max heart rate	Sleep[h]	Kcal
1	4,922	NaN	65	128	0	120
2	11,702	25	66	170	9	389
3	17,683	43	69	163	5	346
4	10,504	26	69	142	7	194
5	8,481	22	67	131	7	158
6	10,871	17	67	134	6	189
7	10,589	22	67	138	5	185
8	1,244	8	52	128	8	7

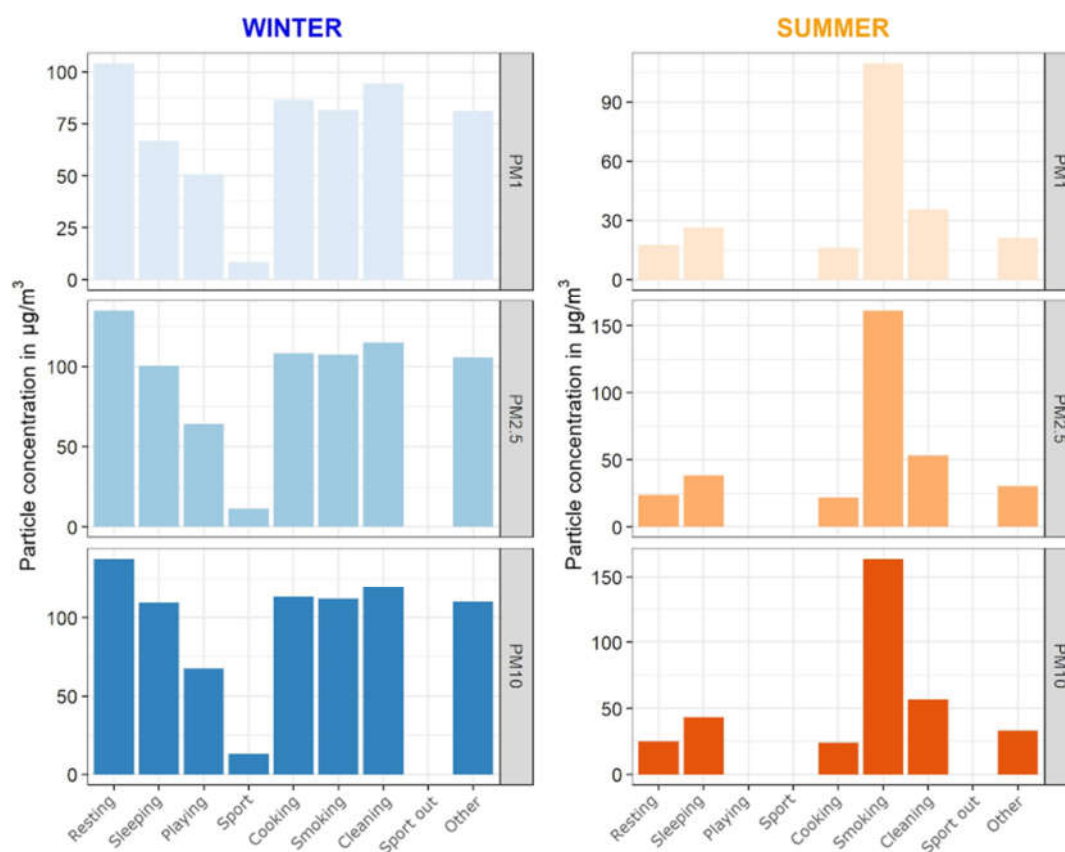
Day	Steps	Stress level	Average heart rate	Max heart rate	Sleep[h]	Kcal
1	6,240	7	63	148	0	186
2	8,000	20	60	144	8	172
3	19,377	16	70	132	4	312
4	16,041	28	73	163	7	488
5	17,717	14	63	118	8	407
6	19,561	25	69	154	7	592
7	30,675	26	74	159	4	443
8	520	12	52	102	7	9

**Figure 7.** Aggregated data from SAT.

Figure 8 shows the average PM values for each activity as indicated by the participant in the TADs. There are certain shortcomings to this visualization as it does not provide any data about the number of instances for each activity, e.g., in this example there is only one hour of smoking indoors in the entire week during the summer season, but over 50 h of sleeping. Although the caption under the plots clearly states that the empty columns mean that there were no recorded instances of that specific activity, there can still be some confusion where the reader might assume that the average concentration is 0  $\mu\text{g}/\text{m}^3$ .

Primarily this plot should communicate differences between the activities in each respective reason. In the example provided in Figure 8, the PM values for smoking are higher than all other activities during the summer season, but not that different from all other activities during winter. A possible explanation would be that there is less natural air circulation during the winter (opening windows or doors), though there could be other explanations. This is another prime example where detailed information about their surroundings would give the individual the most accurate assessment of what the source of the elevated concentrations of PM could be.





**Figure 8.** Faceted plots of average values of three size classes of PM (PM<sub>1</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>) for each specific activity and each season; data from PPM and TAD.

### 3.3. The Final Report

An example of the final report provided to all participants is shown in section C of the Supplementary Data (SD-C). The report began with a personalized greeting, a general description of the project, and the contents of the report. There were also disclaimers about the nature of the low-cost sensors and the uncertainty associated with them. The next page (“Part A”) had a more detailed description of the study, the devices and approaches that were used, and what the reader should specifically focus on. “Part B” described the household conditions, focusing mainly on the data from the IAQ with Figures 2 and 3, accompanied with appropriate captions.

“Part C” contained the plots concerning personal exposure to air pollutants, beginning with PM data, shown in Figures 4–6. Figure 5 was provided only to a handful of participants who had more recorded data and requested a more thorough overview for the entire duration of their involvement. The physical activity information, shown in Figure 7, was presented next and accompanied with a more detailed description of each variable, including some measures for low, average, and elevated heart rate to aid the reader in their interpretation of the data. Figure 8, showing the average PM values for each activity, was included last, with a specific disclaimer that the scales on the y-axis are free.

Some general recommendations on “How to improve indoor air quality” were provided at the end of each report together with two tables extracted from the uHoo sensor device recommendations and descriptions [42].

### 3.4. Issues Faced and Recommendations for Future Studies

Several issues were encountered while compiling, cleaning, and visualizing the data collected from LCSs. While the PPM device proved to be the most accurate when compared with a reference instrument, it also had the most issues regarding data gaps and inconsistent timestamps. Two relatively small improvements to the device would have helped, as they would make the device independent of the GPRS signal: (a) installing a real-time clock (RTC) module which would provide consistent timestamps, and (b) increasing the internal storage and buffer to record PM values without a connection to the server. Several optimizations to reduce energy usage would be possible, e.g., less frequent GPS recordings while stationary, option to only upload the data when the device is charging, etc.

On the other hand, the IAQ had very consistent data streams, accurate timestamps, and a very intuitive interface. Two improvements would make the device function more independently: (a) a small internal storage for times when there was no wi-fi signal, which would allow the device to store the data in an internal buffer and upload it when the connection was re-established, and (b) a small battery to allow the device to function during power outages.

The SAT was very reliable, had an internal storage capacity for 14 days of data, and had a battery that lasted between 5 and 7 days. An improvement would be to provide a way to verify if the data is being logged correctly. At times, the device was not placed properly on the wrist or had some other error with data logging, and this was only observable at the end of the sampling campaign. Though the SAT did provide a uniform dataset it had to be extracted by a separate process in collaboration with the company that produced the device. Accessing data from commercial devices proved complicated and preconditioned on setting up exclusive deals with companies. Even when the deal is set, the entire data retrieval process is reliant on the cooperation of the company. It would be better if the raw data streams were open access.

A key improvement for the TADs would be to allow more granular activity logging during the day, e.g., every 15 min. TADs could also be somewhat customized to different participants or days of the week, e.g., participants who perform only one activity, such as work, during morning hours could have a different TAD during workdays than during weekends. The hourly resolution of TADs caused some issues when presenting and visualizing data for the participants as the average values were skewed, due to the fact that most activities do not have a duration of one hour nor do they start and end at full hours. This meant, for example, that someone who went for a 40 min run followed by smoking a cigarette might only have recorded “running” for that hour. Even though the person could have checked both activities, there would still be no information as to which part of that full hour the “smoking” vs. “running” occurred. Recording activities minute by minute would be a heavy burden for participants, so future research should focus on other solutions, such as complex activity recognition using machine learning, smartphones, or other tools [45,56].

Visualizing the data proved challenging at times and required unique solutions. The main challenge in producing the plots in Figure 2 proved to be the horizontal ribbon with “optimal values”, which had to be referenced in a way that would allow this value to be presented for each individual hour, while also enabling faceting of the plots. Additional variables with minimum and maximum data for each season were introduced, which shortened the script for the final construction of the plots.

A rather easy, though important, improvement for the plots in Figure 6 would be to show specific dates and days of the week instead of the number of days since the participant joined the sampling campaign. Participants do not always remember which day they started the campaign and would have to go back to the IAQ figure to find out. As a significant amount of time can elapse between the campaign and the distribution of reports to participants, it could be good for future studies to always indicate in the figures the date and day of the week.

Figure 8 could be improved by indicating the number of instances for each activity by coloring the bars according to a color scale reflecting the frequency of activities or by changing the width of each bar accordingly. The activities without data should be clearly marked with a symbol or a text. A requirement for a minimal amount of data should be considered to remove activities with only a few instances. The color schemes should also be intuitive, such as coloring winter blue (cold color), summer orange (warm color), or smoking black, which instinctively guides the reader.

Manually collected data from TADs were double-checked by the researchers as there were some non-obvious errors, e.g., smoking selected for a person who designated that they do not smoke, which sometimes indicated a user error and other times that the person was in fact an infrequent smoker. As with any dataset, these inconsistencies and all permutations can be very time-consuming to implement into the report-generating code. A large number of reports (and the associated data) also necessitates that there is a careful process when deciding what functions to use, and how much time and processing power they will need.

#### 4. Conclusions

Data fusion and visualization of data obtained in personal exposure campaigns performed in seven European cities within the ICARUS project were conducted. By using a diverse set of devices (wearable and static, commercially available, and custom-made) with different temporal and spatial resolutions, a significant amount of data was obtained for each participant. Data fusion was performed in order to integrate the multi-sensor and multi-parameter datasets collected from the > 600 participants. Individualized reports were compiled for each participant with an automated process. Following these large-scale campaigns, several lessons were drawn and recommendations for future studies were provided.

Using low-cost sensors to assess air quality on an individual level presented some unique challenges, e.g., fusing data by rounding, duplicating, and removing certain parts of the timestamps, which allowed a uniform presentation on several plots. Mostly simple modifications were enough to provide some clarity and make data fusion more straightforward. Appropriate guidelines have to be considered carefully to avoid confusing the participant or giving false impressions on otherwise non-harmful concentrations of pollutants.

Participants should not be overwhelmed with the report, rather it should provide sufficient data for them to obtain as much meaningful information as possible. While the SAT provided a large amount of data, a decision was taken to only include the visualizations that were the most effective at communication. Apart from the number of visualizations, the appropriate type must also be carefully selected and curated. Our approach with relative values for NO<sub>2</sub>, CO<sub>2</sub>, and TVOCs provided enough data to clearly see some trends, without providing unreliable absolute data values. On the other hand, the higher reliability and accuracy of PM concentrations and meteorological parameters enabled us to provide absolute values. A clear option should be included to observe trends between days, seasons, and activities. These visualizations must also reflect the results of collocations and validations made prior to deploying these devices. Citizens must be made aware of the accuracy (and shortcomings) of the device they are using and to what extent can they rely on the results. A properly structured report will guide them through the report itself and give them enough support to extract the most useful interpretation of the data that they can.

A well-informed public can collaborate with and react to changes in their environment, be it by influencing policy decisions or making changes to their individual lifestyles and behaviors. Using LCS provides a conduit for citizens to be empowered by data that they can collect, observe, and interpret. We, as researchers, must provide the necessary tools and options for them and guide them through the process. Changes in policy can come from the top-down or from the ground-up. In both cases, the citizens that are



affected by these policy changes must be active participants in designing and implementing these solutions.

**Supplementary Materials:** The following are available online at [www.mdpi.com/article/10.3390/ijerph182111614/s1](http://www.mdpi.com/article/10.3390/ijerph182111614/s1) (SD-A: Example of timestamp gaps in PPM dataset; SD-B: Example of dataset used for visualizations and reports; and SD-C: Example of entire report for participants).

**Author Contributions:** Conceptualization, R.N., D.K., B.F., D.V. and C.D.; methodology, R.N., B.F. and C.D.; software, R.N. and I.P.; validation, R.N., D.K., J.A.R. and T.K.; formal analysis, R.N.; investigation, R.N., D.K., J.A.R., T.K., D.P., A.S., D.S. (Dikaia Saraga), M.G.P., O.M., O.S., C.D., J.V., S.G.D.S.-A., D.C. and S.K.; resources, R.N., D.K., J.A.R., T.K., T.M., D.C. and S.K.; data curation, R.N.; writing—original draft preparation, R.N.; writing—review and editing, R.N., D.K., J.A.R., T.K., D.S. (Dikaia Saraga), B.F., D.V., C.D. and A.G.; visualization, R.N. and J.A.R.; supervision, D.K.; project administration, D.S. (Dimosthenis Sarigiannis) and A.G.; funding acquisition, D.S. (Dimosthenis Sarigiannis) All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has received funding from the European Union’s Horizon 2020 Programme for Research, Technological Development, and Demonstration, under grant agreement No. 690105 (Integrated Climate forcing and Air Pollution Reduction in Urban Systems (ICARUS)). This work reflects only the authors’ views, and the European Commission is not responsible for any use that may be made of the information it contains. Funding was received from the Young Researchers Program and the P1-0143 program “Cycling of substances in the environment, mass balances, modelling of environmental processes and risk assessment”, both funded by the Slovenian Research Agency. The authors thank RECETOX Research Infrastructure (No. LM2018121) and ACTRIS-CZ Research Infrastructure (No. LM2018122) financed by the Ministry of Education, Youth and Sports and the Operational Programme Research, Development and Innovation (CZ.02.1.01/0.0/0.0/16\_013/0001315 and CZ.02.1.01/0.0/0.0/16\_013/0001761), as well as the CETO-COEN Excellence Project, supported by the Teaming Action of the EU Horizon 2020 programme (857560) and the Operational Programme Research, Development and Innovation (No. CZ.02.1.01/0.0/0.0/17\_043/0009632).

**Institutional Review Board Statement:** Ethical approval for the ICARUS project in Slovenia was obtained from the National Medical Ethics Committee of the Republic of Slovenia (approval nr. 0120-388/2018/6 on 22 August 2018). The data in this paper were selected only from participants in Slovenia.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data available on request due to privacy restrictions.

**Acknowledgments:** The authors would like to thank the participants from all the cities, that generously provided their time and effort to this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Wellenius, G.A.; Schwartz, J.; Mittleman, M.A. Health and the environment: Addressing the health impact of air pollution: Draft resolution proposed by the delegations of Albania, Chile, Colombia, France, Germany, Monaco, Norway, Panama, Sweden, Switzerland, Ukraine, United States of America, Uruguay and Zambia. *Sixty-Eighth World Health Assembly* **2015**, *14*, 68.
- Payne-Sturges, D.C.; Marty, M.A.; Perera, F.; Miller, M.D.; Swanson, M.; Ellickson, K.; Cory-Slechta, D.A.; Ritz, B.; Balmes, J.; Anderko, L.; et al. Healthy Air, Healthy Brains: Advancing Air Pollution Policy to Protect Children’s Health. *Am. J. Public Health* **2019**, *109*, 550–554, <https://doi.org/10.2105/AJPH.2018.304902>.
- Sicard, P.; Agathokleous, E.; De Marco, A.; Paoletti, E.; Calatayud, V. Urban population exposure to air pollution in Europe over the last decades. *Environ. Sci. Eur.* **2021**, *33*, 28, <https://doi.org/10.1186/s12302-020-00450-2>.
- Jerrett, M.; Donaire-Gonzalez, D.; Popoola, O.; Jones, R.; Cohen, R.C.; Almanza, E.; de Nazelle, A.; Mead, I.; Carrasco-Turigas, G.; Cole-Hunter, T.; et al. Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. *Environ. Res.* **2017**, *158*, 286–294, <https://doi.org/10.1016/j.envres.2017.04.023>.
- Hubbell, B.J.; Kaufman, A.; Rivers, L.; Schulte, K.; Hagler, G.; Clougherty, J.; Cascio, W.; Costa, D. Understanding Social and Behavioral Drivers and Impacts of Air Quality Sensor Use. *Sci. Total Environ.* **2018**, *621*, 886–894, <https://doi.org/10.1016/j.scitotenv.2017.11.275>.
- Miskell, G.; Salmond, J.; Williams, D.E. Low-cost sensors and crowd-sourced data: Observations of siting impacts on a network of air-quality instruments. *Sci. Total Environ.* **2017**, *575*, 1119–1129, <https://doi.org/10.1016/j.scitotenv.2016.09.177>.

7. Morawska, L.; Thai, P.K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **2018**, *116*, 286–299, <https://doi.org/10.1016/j.envint.2018.04.018>.
8. Rai, A.C.; Kumar, P.; Pilla, F.; Skouloudis, A.N.; Di Sabatino, S.; Ratti, C.; Yasar, A.; Rickerby, D. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci. Total Environ.* **2017**, *607–608*, 691–705, <https://doi.org/10.1016/j.scitotenv.2017.06.266>.
9. Robinson, J.A.; Kocman, D.; Horvat, M.; Bartonova, A. End-User Feedback on a Low-Cost Portable Air Quality Sensor System—Are We There Yet? *Sensors* **2018**, *18*, 3768, <https://doi.org/10.3390/s18113768>.
10. Goal 11. Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable—Indicators and a Monitoring Framework Available online: <https://indicators.report/goals/goal-11/> (accessed on 2 March 2021).
11. Mean Urban Air Pollution of Particulate Matter (PM10 and PM2.5)—Indicators and a Monitoring Framework Available online: <https://indicators.report/indicators/i-69/> (accessed on 2 March 2021).
12. Jarvis, D.J.; Adamkiewicz, G.; Heroux, M.-E.; Rapp, R.; Kelly, F.J. *Nitrogen Dioxide*; World Health Organization: Geneva, Switzerland, 2010.
13. Nuvolone, D.; Petri, D.; Voller, F. The effects of ozone on human health. *Environ. Sci. Pollut. Res.* **2018**, *25*, 8074–8088, <https://doi.org/10.1007/s11356-017-9239-3>.
14. Shuai, J.; Kim, S.; Ryu, H.; Park, J.; Lee, C.K.; Kim, G.-B.; Ultra, V.U.; Yang, W. Health risk assessment of volatile organic compounds exposure near Daegu dyeing industrial complex in South Korea. *BMC Public Health* **2018**, *18*, 528, <https://doi.org/10.1186/s12889-018-5454-1>.
15. Casset, A.; de Blay, F. Health effects of domestic volatile organic compounds. *Rev. Mal. Respir.* **2008**, *25*, 475–485, [https://doi.org/10.1016/s0761-8425\(08\)71587-0](https://doi.org/10.1016/s0761-8425(08)71587-0).
16. Jacobson, T.A.; Kler, J.S.; Hernke, M.T.; Braun, R.K.; Meyer, K.C.; Funk, W.E. Direct human health risks of increased atmospheric carbon dioxide. *Nat. Sustain.* **2019**, *2*, 691–701, <https://doi.org/10.1038/s41893-019-0323-1>.
17. Castanedo, F. A Review of Data Fusion Techniques. *Sci. World J.* **2013**, *2013*, e704504, <https://doi.org/10.1155/2013/704504>.
18. Okafor, N.U.; Alghorani, Y.; Delaney, D.T. Improving Data Quality of Low-cost IoT Sensors in Environmental Monitoring Networks Using Data Fusion and Machine Learning Approach. *ICT Express* **2020**, *6*, 220–228, <https://doi.org/10.1016/j.icte.2020.06.004>.
19. Schneider, P.; Castell, N.; Vogt, M.; Dauge, F.R.; Lahoz, W.A.; Bartonova, A. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environ. Int.* **2017**, *106*, 234–247, <https://doi.org/10.1016/j.envint.2017.05.005>.
20. Gressent, A.; Malherbe, L.; Colette, A.; Rollin, H.; Scimia, R. Data fusion for air quality mapping using low-cost sensor observations: Feasibility and added-value. *Environ. Int.* **2020**, *143*, 105965, <https://doi.org/10.1016/j.envint.2020.105965>.
21. Senthilkumar, N.; Gilfether, M.; Metcalf, F.; Russell, A.G.; Mulholland, J.A.; Chang, H.H. Application of a Fusion Method for Gas and Particle Air Pollutants between Observational Data and Chemical Transport Model Simulations Over the Contiguous United States for 2005–2014. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3314, <https://doi.org/10.3390/ijerph16183314>.
22. Clements, A.L.; Griswold, W.G.; Rs, A.; Johnston, J.E.; Herting, M.M.; Thorson, J.; Collier-Oxandale, A.; Hannigan, M. Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). *Sensors* **2017**, *17*, 2478, <https://doi.org/10.3390/s17112478>.
23. Lewis, A.C.; von Schneidmeyer, E.; Peltier, R.E. *Low-Cost Sensors for the Measurement of Atmospheric Composition: Overview of Topic and Future Applications*; World Meteorological Organization (WMO): Geneva, Switzerland, 2018.
24. Paul, J.D.; Buytaert, W. Chapter One—Citizen Science and Low-Cost Sensors for Integrated Water Resources Management. In *Advances in Chemical Pollution, Environmental Management and Protection*; Friesen, J., Rodríguez-Sinobas, L., Eds.; Advanced Tools for Integrated Water Resources Management; Elsevier: Amsterdam, The Netherlands, 2018; Volume 3, pp. 1–33.
25. Wang, Y.; Han, F.; Zhu, L.; Deussen, O.; Chen, B. Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series. *IEEE Trans. Vis. Comput. Gr.* **2018**, *24*, 1141–1154, <https://doi.org/10.1109/TVCG.2017.2653106>.
26. Saket, B.; Endert, A.; Demiralp, Ç. Task-Based Effectiveness of Basic Visualizations. *IEEE Trans. Vis. Comput. Gr.* **2019**, *25*, 2505–2512, <https://doi.org/10.1109/TVCG.2018.2829750>.
27. Garcia-Retamero, R.; Galesic, M. Who profits from visual aids: Overcoming challenges in people’s understanding of risks. *Soc. Sci. Med.* **2010**, *70*, 1019–1025, <https://doi.org/10.1016/j.socscimed.2009.11.031>.
28. Saket, B.; Srinivasan, A.; Ragan, E.D.; Endert, A. Evaluating Interactive Graphical Encodings for Data Visualization. *IEEE Trans. Vis. Comput. Gr.* **2018**, *24*, 1316–1330, <https://doi.org/10.1109/TVCG.2017.2680452>.
29. ICARUS2020.eu Available online: <https://icarus2020.eu/> (accessed on 12 October 2018).
30. Kocman, D.; Kanduć, T.; Novak, R.; Robinson, J.A.; Mikeš, O.; Degrendele, C.; Sáňka, O.; Vinkler, J.; Prokeš, R.; Vienneau, D.; et al. Multi-Sensor Data Collection for Personal Exposure Monitoring: ICARUS Experience. *Fresenius Environ. Bull.* **2021**, *6* (accepted for publication).
31. Sarigiannis, D.; Chapizanis, D.; Arvanitis, A. D4.1 Report on the Methodology for Estimating Individual Exposure. ICARUS2020 Consortium Publication. Available online: [https://icarus2020.eu/wp-content/uploads/2018/03/ICARUS-Deliverable-D4.1\\_FINAL.pdf](https://icarus2020.eu/wp-content/uploads/2018/03/ICARUS-Deliverable-D4.1_FINAL.pdf) (accessed on 10 September 2021).

32. Sarigiannis, D.; Karakitsios, S.; Chapizanis, D.; Hiscock, R. D4.2\_ICARUS\_Methodology for Properly Accounting for SES in Exposure Assessment.pdf. ICARUS2020 consortium publication. Available online: [https://icarus2020.eu/wp-content/uploads/2019/02/ICARUS\\_D4.2.pdf](https://icarus2020.eu/wp-content/uploads/2019/02/ICARUS_D4.2.pdf) (accessed 10 September 2021).
33. Robinson, J.A.; Novak, R.; Kanduč, T.; Sarigiannis, D.; Kocman, D. *Articulating User Experience of a Multi-Sensor Personal Air Quality Exposure Study*; Department of Environmental Sciences, Jožef Stefan Institute: Ljubljana, Slovenia, 2021. manuscript in preparation.
34. R: The R Project for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 5 December 2019).
35. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
36. Wickham, H.; François, R.; Henry, L.; Müller, K. Dplyr: A Grammar of Data Manipulation. CRAN. 2018. Available online: <https://dplyr.tidyverse.org> (accessed 10 September 2021).
37. Xie, Y. Knitr: A General-Purpose Package for Dynamic Report Generation in R. CRAN. 2021. Available online: <https://yihui.org/knitr/> (accessed 10 September 2021).
38. Allaire, J.J.; Xie, Y.; McPherson, J.; Luraschi, J.; Ushey, K.; Atkins, A.; Wickham, H.; Cheng, J.; Chang, W.; Iannone, R. Rmarkdown: Dynamic Documents for R. CRAN. 2021. Available online: <https://pkgs.rstudio.com/rmarkdown/> (accessed 10 September 2021).
39. Novak, R.; Kocman, D.; Robinson, J.A.; Kanduč, T.; Sarigiannis, D.; Horvat, M. Comparing Airborne Particulate Matter Intake Dose Assessment Models Using Low-Cost Portable Sensor Data. *Sensors* **2020**, *20*, 1406, <https://doi.org/10.3390/s20051406>.
40. Industries, A. Adafruit PCF8523 Real Time Clock Assembled Breakout Board. Available online: <https://www.adafruit.com/product/3295> (accessed on 30 September 2020).
41. Garmin; subsidiaries, G.L. or its Garmin vivosmart® 3 | Fitness Activity Tracker. Available online: <https://buy.garmin.com/en-US/US/p/567813> (accessed on 3 September 2019).
42. uHoo | Product. Available online: <https://uhooair.com/product/> (accessed on 16 November 2018).
43. Mahajan, S.; Kumar, P.; Pinto, J.A.; Riccetti, A.; Schaaf, K.; Camprodon, G.; Smári, V.; Passani, A.; Forino, G. A citizen science approach for enhancing public understanding of air pollution. *Sustain. Cities Soc.* **2020**, *52*, 101800, <https://doi.org/10.1016/j.scs.2019.101800>.
44. Nikolakopoulos, T.; Gotti, A.; Tsiros, E.; Siora, E. D7.2: Report on the Design of Technical Framework and System Architecture of the ICARUS DSS. ICARUS2020 Consortium Publication. Available online: [https://icarus2020.eu/wp-content/uploads/2017/08/D.7.2\\_ICARUS\\_Design\\_of\\_%20technical\\_framework\\_and\\_system\\_architecture\\_of\\_the\\_ICARUS\\_DS\\_S\\_FINAL.pdf](https://icarus2020.eu/wp-content/uploads/2017/08/D.7.2_ICARUS_Design_of_%20technical_framework_and_system_architecture_of_the_ICARUS_DS_S_FINAL.pdf) (accessed on 10 September 2021).
45. Novak, R.; Kocman, D.; Robinson, J.A.; Kanduč, T.; Sarigiannis, D.; Džeroski, S.; Horvat, M. Low-Cost Environmental and Motion Sensor Data for Complex Activity Recognition: Proof of Concept. *Eng. Proc.* **2020**, *2*, 54, <https://doi.org/10.3390/ecs-7-08194>.
46. Robinson, J.A.; Novak, R.; Kanduč, T.; Maggos, T.; Pardali, D.; Stamatelopoulou, A.; Saraga, D.; Vienneau, D.; Flückiger, B.; Mikeš, O.; et al. User-Centred Design of a Final Results Report for Participants in Multi-Sensor Personal Air Pollution Exposure Monitoring Campaigns. *Preprints* **2021**, doi:10.20944/preprints202110.0031.v1.
47. Air Quality Now—Indices Definition. Available online: [http://airqualitynow.eu/about\\_indices\\_definition.php](http://airqualitynow.eu/about_indices_definition.php) (accessed on 20 January 2021).
48. Zhang, X.; Zhao, Z.; Nordquist, T.; Norback, D. The prevalence and incidence of sick building syndrome in Chinese pupils in relation to the school environment: A two-year follow-up study. *Indoor Air* **2011**, *21*, 462–471, <https://doi.org/10.1111/j.1600-0668.2011.00726.x>.
49. AQ-SPEC. *Field Evaluation—uHoo PM2.5, Ozone, and CO Sensor*; AQ-SPEC: Diamond Bar, CA, USA, 2019. Available online: <http://www.aqmd.gov/docs/default-source/aq-spec/field-evaluations/uhoo---field-evaluation.pdf?sfvrsn=12> (accessed on 10 September 2021).
50. Baldelli, A. Evaluation of a low-cost multi-channel monitor for indoor air quality through a novel, low-cost, and reproducible platform. *Meas. Sens.* **2021**, *17*, 100059, <https://doi.org/10.1016/j.measen.2021.100059>.
51. Tran, V.V.; Park, D.; Lee, Y.-C. Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2927, <https://doi.org/10.3390/ijerph17082927>.
52. Wei, Y.; Wang, Y.; Di, Q.; Choirat, C.; Wang, Y.; Koutrakis, P.; Zanobetti, A.; Dominici, F.; Schwartz, J.D. Short term exposure to fine particulate matter and hospital admission risks and costs in the Medicare population: Time stratified, case crossover study. *BMJ* **2019**, *367*, 16258, <https://doi.org/10.1136/bmj.l6258>.
53. Orellano, P.; Reynoso, J.; Quaranta, N.; Bardach, A.; Ciapponi, A. Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO<sub>2</sub>), and ozone (O<sub>3</sub>) and all-cause and cause-specific mortality: Systematic review and meta-analysis. *Environ. Int.* **2020**, *142*, 105876, <https://doi.org/10.1016/j.envint.2020.105876>.
54. WHO Guidelines: Ambient (Outdoor) Air Pollution (Prior to 2021). Available online: [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) (accessed on 3 March 2021).
55. New WHO Global Air Quality Guidelines Aim to Save Millions of Lives from Air Pollution. Available online: <https://www.euro.who.int/en/media-centre/sections/press-releases/2021/new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution> (accessed on 27 September 2021).
56. Park, J.; Kim, S. Machine Learning-Based Activity Pattern Classification Using Personal PM2.5 Exposure Information. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6573, <https://doi.org/10.3390/ijerph17186573>.