Data in Brief 52 (2024) 109877

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Personal airborne particulate matter exposure and intake dose, indoor air quality, biometric, and activity dataset from the city of Ljubljana, Slovenia

Rok Novak[a,b,*], Johanna Amalia Robinson[a,b,c], Tjaša Kanduč[a], Dimosthenis Sarigiannis[d,e,f], David Kocman[a]

[a] Department of Environmental Sciences, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b] Ecotechnologies Programme, Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia
[c] Center for research and development, Slovenian Institute for Adult Education, Ulica Ambrožiča Novljana 5, 1000 Ljubljana, Slovenia
[d] Environmental Engineering Laboratory, Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[e] HERACLES Research Center on the Exposome and Health, Center for Interdisciplinary Research and Innovation, Balkan Center, Bldg. B, 10th km Thessaloniki-Thermi Road, 57001, Greece
[f] Environmental Health Engineering, Department of Science, Technology and Society, University School of Advanced Study IUSS, Palazzo del Broletto - Piazza Della Vittoria 15, 27100 Pavia, Italy

ARTICLE INFO

Article history:
Received 13 November 2023
Revised 20 November 2023
Accepted 23 November 2023
Available online 30 November 2023

Dataset link: Ljubljana Multi-Sensor Indoor and Outdoor PM Exposure, Environment, and Personal Health Dataset (Original data)

Keywords:
Personal monitors
Urban stressors
Wearable sensors

ABSTRACT

To gain a comprehensive understanding of the effects of air pollution on human health, it is imperative to assess exposure at an individual level. This necessitates the use of personal monitoring equipment that can record exposure in real-time while simultaneously capturing relevant biometric data and the activities being undertaken by the individual. The dataset presented herein encompasses data from five distinct sources, collected in Ljubljana, Slovenia:

(a) Questionnaire filled out by all participants for general information, i.e., age, height, weight.
(b) Time Activity Diaries maintained by participants.

* Corresponding author at: Department of Environmental Sciences, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia.
E-mail address: rok.novak@ijs.si (R. Novak).
Social media: @Novak_Rok (R. Novak)

https://doi.org/10.1016/j.dib.2023.109877

(a) A personal particulate matter sensing unit that recorded data on PM_1, $PM_{2.5}$, PM_{10}, ambient temperature, relative humidity, and geographical location coordinates.
(b) A wrist-worn Smart Activity Tracker that captures heart rate and movement data.
(c) An Indoor Air Quality sensing unit, specifically the uHoo device, which measures parameters such as Temperature, Relative Humidity, CO_2, TVOC, $PM_{2.5}$, CO, O_3, NO_2, and Air Pressure.
Participants were equipped with these sensors for two separate 1-week durations in 2019, specifically during a heating season and a non-heating season, as part of the ICARUS H2020 project. Throughout these periods, data was recorded at a 1-minute resolution. Additionally, based on the collected data, an inhalation rate was determined, and an inhalation adjusted exposure (or intake dose) was computed. The richness and granularity of this dataset not only offer insights into the direct health implications of air pollution exposure but also provide a valuable resource for urban planners, environmental scientists, and policymakers aiming to design healthier living environments and implement effective air quality regulations.

Specifications Table

Subject	Environmental Science: Pollution
Specific subject area	The dataset examines individual-level air pollution exposure, its potential health implications, and offers insights for urban planning and air quality regulations.
Data format	Data frame: .csv format, raw
Type of data	Table
Data collection	Participants contributed data through various means: 1) Providing age, height, gender, and weight information, 2) Maintaining daily time activity diaries (TADs), 3) Wearing a Plantower PMS5003 personal particulate matter (PPM) sensor on their person or nearby, 4) Wearing a Garmin Vivosmart 3 wristband as a smart activity tracker (SAT), 5) Hosting a uHoo indoor air quality sensor (IAQ) in their residence. The SAT, worn on their dominant wrist, recorded 1-minute resolution data, stored locally, and later transferred to a computer via USB. The PPM sensor recorded 1-minute data, which was simultaneously transferred to a remote server via GPRS, and stored on an SD card. The IAQ sensor, located in the most frequently used room during the day, recorded 1-minute data, which was transferred to uHoo servers, and later to a local device. TADs, completed daily with 1-h resolution, were collected in paper format. Personal information was gathered via researcher-conducted interviews.
Data source location	Institution: Jožef Stefan Institute City: Ljubljana Country: Slovenia
Data accessibility	Repository name: Zenodo Data identification number: https://doi.org/10.5281/zenodo.10142677 Direct URL to data: https://zenodo.org/records/10142677
Related research article	R. Novak, J.A. Robinson, T. Kanduč, D. Sarigiannis, D. Kocman, Simulating the impact of particulate matter exposure on health-related behaviour: A comparative study of stochastic modelling and personal monitoring data, Health & Place. 83 (2023) 103111. https://doi.org/10.1016/j.healthplace.2023.103111 [1]

1. Value of the Data

- This dataset represents a comprehensive assessment of air quality-oriented environmental, biometric, and personal health variables, collected from a large and diverse group of individuals. While air pollution exposure assessments are frequently made using data from sparse environmental monitoring stations [2,3], this dataset offers fine-grained temporal resolution of exposure with ample context provided via recorded activities, microlocations, and biometric data. Moreover, the geographic location of the Ljubljana municipality offers valuable insight into how exposure to poor air quality relates to specific weather conditions characteristic of a sub-alpine basin.
- A diverse set of stakeholders can benefit from this comprehensive dataset. Environmental researchers can explore further the direct health implications of various pollution levels and pollutants. Public health researchers and epidemiologists can correlate biometric data with exposure for comprehensive health studies. Policymakers can use this dataset as a preliminary indicator, highlighting areas where more detailed environmental or exposure studies are required. Technology developers can improve or innovate monitoring devices leveraging insights from existing data. Lastly, the general public can make informed decisions about activities during pollution peaks, contributing to individual and community well-being.
- By utilizing this dataset, researchers can improve exposure assessments and reduce the frequency of gaps and erroneous recordings. Biometric data, in combination with exposure assessment and the calculated intake dose, can provide a basis for an intake dose assessment study. The data can be reused for research related to policy development and interventions in urban environments.
- Local communities living in similar geographical locations, frequently afflicted by atmospheric thermal inversions, foggy days, and poor air circulation, can benefit from the insights of this dataset [4].

2. Background

This dataset, part of the ICARUS H2020 project focused on air quality and climate change governance in EU Member States, was curated to develop tailored strategies for assessing exposure to particulate matter on an individual level in Ljubljana, Slovenia.

Utilizing portable, cost-effective sensor systems, this dataset aimed to monitor personal exposure, providing a comprehensive understanding of individual air pollutant exposure concerning diverse microenvironments and daily activities.

The sensors employed in the ICARUS project captured personal movement, activity intensity, location, and key meteorological parameters. They included technologies for monitoring environmental factors and pollutants using adapted smartphone technology, off-the-shelf sensors, and academic technological advancements. Participant questionnaires supplemented this data, offering sociodemographic insights.

This specific dataset from Ljubljana, Slovenia, contributed to an agent-based model simulation evaluating particulate matter dose and exposure. This analysis formed a part of the research paper titled “Simulating the impact of particulate matter exposure on health-related behavior: A comparative study of stochastic modelling and personal monitoring data,” presenting findings on assessing air pollution exposure via digital twin simulations [1]. A subset of the dataset was used to compare particulate matter intake dose models [5] and assess exposure to particulate matter in periods of atmospheric thermal inversion [4]. Moreover, two research papers were published based on the harmonization and visualization of this dataset [6], and the user-centered design of the final results report for the participants [7].

3. Data Description

The dataset described in this work encompasses data collected and provided by 82 participants living in the municipality of Ljubljana, Slovenia, as part of a multinational European Horizon 2020 project ICARUS (Integrated Climate forcing and Air pollution Reduction in Urban Systems). The project was conducted between 2016 and 2020 [8], with data collected in seven European cities (Athens, Basel, Brno, Ljubljana, Madrid, Milan, and Thessaloniki). Participants in Ljubljana were distributed across the municipality, as shown in Fig. 1.

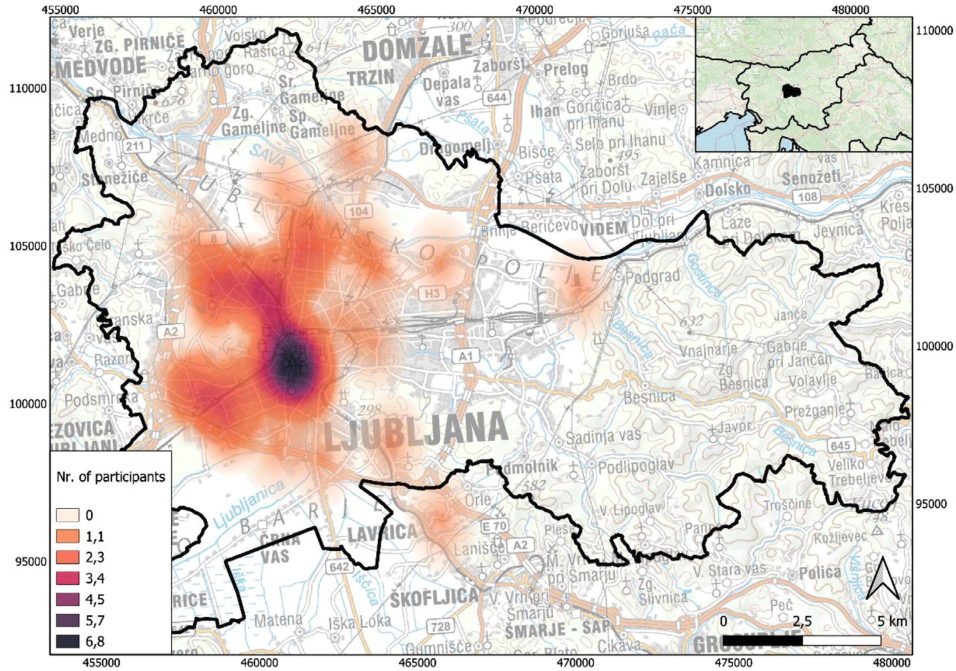

Fig. 1. Geographical heatmap illustrating participant household distribution within the ICARUS campaign in Ljubljana municipality (outlined by the black line). The larger map utilizes the National General Map of Slovenia at a scale of 1:250,000, while the smaller map employs the OSM Standard. Coordinates are depicted in meters within the Gauss-Krüger projection.

The data collection period took place between February 16, 2019, and May 25, 2019. This period encompassed different seasons, in terms of weather, air quality, temperature, and other ambient variables. Moreover, it included a period with lower temperatures when individuals in Ljubljana would be heating their homes, i.e., a “heating season”, and a period with higher temperatures, when heating was no longer required, i.e., “non-heating season”. Data collected before March 12 represents the “heating season,” while data after April 27 corresponds to the “non-heating season.” This paper provides an overview of the cleaned/filtered dataset with outliers caused by software or hardware error removed. Derived from the larger ICARUS project, which included data collection in multiple European cities, this dataset integrates five data sources collected in Ljubljana:

(a) Participant Questionnaires: Containing certain individual information, i.e., age, height, gender.
(b) Time Activity Diaries (TADs): Providing hourly activity records for each participant.
(c) Personal PM Monitors (PPM): Measuring indoor and outdoor PM concentrations.

(d) Smart Activity Trackers (SAT): Recording heart rate and movement data.
(e) Indoor Air Quality Station (IAQ): Capturing indoor air quality parameters.

The dataset includes the calculation of inhalation rate based on heart rate data, allowing for the determination of inhalation rate-adjusted exposure or intake dose, also included in the dataset.

3.1. Participant questionnaires

Questionnaires at the beginning of the first sampling period were filled out by answers provided by the participants. This was usually done by the researcher or in the presence of the researcher. The original questionnaire contained several questions regarding the living environments and habits of the involved participants. However, this particular dataset contains only data on the height, weight, age, and gender of each participant. The data on gender and age were collected from the questionnaire, and the height and weight data during the SAT set-up procedure by interviewing the participants. Table 1 shows the statistical analysis of the data collected from the questionnaire, calculated as one recording per individual. The analysis was conducted in R using base statistical functions, i.e., mean, median, sd, max, min, and quartile. Additionally, the data on gender showed that there were 41 and 47 individuals who identified as female and male, respectively.

Table 1
Statistical analysis of weight, height, and age of participants in 2019.

Variable	Median	Mean	SD	Max	Min	1stQ	3rdQ
Age	41	39.9	15.4	75	11	31.25	49
Weight	70	71.4	20.1	160	28	60	80
Height	171.5	172.0	11.5	196	130	166	180

3.2. Time activity diaries

Time Activity Diaries (TADs) were collected from each participant, for all days that they conducted measurements. Additionally, to the data collected from the TAD, in the dataset described in this paper, there are two extra columns, labelled as “location” and “details”. “Location” uses data from the other columns and provides feedback if the person was in a location, described as “indoor”, “outdoor”, “in transit” or NA. Similarly, “details” provides information if the person was commuting via “bus”, “car”, “motorbike”, “bicycle”, or “foot” (indicating walking), or located in “office”, “home”, or “other”. Sums of all instances of the activities and house conditions are listed in Table 2 in minute resolution, for both seasons.

Table 2
Number of instances for all activities and house conditions included in the TADs.

Activity	home_in	office_in	other_in	home_out	office_out	other_out	bus
Count	858886	163746	129784	29536	29670	109611	13021
Activity	**car**	**motorbike**	**bicycle**	**foot**	**other**	**resting_in**	**sleep_in**
Count	74316	540	19321	63991	2580	300116	431245
Activity	**playing_in**	**sports_in**	**cooking_in**	**smoking_in**	**cleaning_in**	**running_out**	**sports_out**
Count	25780	19335	41742	16831	29681	5700	29985

3.3. Personal particulate matter monitors

Each participant wore the personal particulate matter (PPM) monitor the entire sampling period. They were instructed to wear the sensor on their clothes, e.g., clipped to belt or pocket, or on a backpack or handbag, when they were moving around. In cases when the individual was sedentary or laying down, they were instructed to place the device as near as possible. At all times the area with the vent for the sensors had to be open and unobstructed. The battery life of the device was up to 8 hours, which necessitated that it was charged as often as possible. All variables collected from the PPM were numeric. As per the dataset, read from left to right, the columns in order represent:

- Columns 1–3: recorded values based on an internal "environmental" algorithm for transforming particle counts to mass or concentration, for three size classes of PM, i.e., aerodynamic diameters of 1, 2.5 and 10 µm.
- Columns 4–5: same as previous three, but calculated using the "standardized" algorithms, simulating controlled laboratory conditions.
- Columns 6–7: ambient temperature and humidity, which were somewhat influenced by the heat coming from the lithium-ion battery, when the device was charging.
- Column 8: battery capacity at the specific minute in percent, from 1 to 100.
- Columns 9–10: Altitude at location, and speed of movement in km/h.
- Column 11: Counting up by 60 every minute from the moment the "reset button" was pressed on the device.

Already published works describe the design of the PPM device [1] and our in-house evaluation of the monitor [5].

Table 3 provides an overview of the statistical analysis of the dataset, showing median, mean, standard deviation, maximum, minimum, 1st quartile, 3rd quartile, and percent of NA values. Importantly, the dataset was cleaned by removing the outliers caused by software of hardware errors. In the part of the dataset provided by the PPM, the outliers removed were: 1) all PM env values above 3333 and PM std values above 5000, as the data suggests this to be the limit of the sensor, 2) temperatures >40 °C, as it is not feasible, with only two recordings, both without an accompanying trend and significantly above the limit (51 °C and 486 °C), 4) relative humidity >100 %, as it is not possible, 5) altitude <0 as it is not possible, and >2864 as the highest point in Slovenia, 6) speed >180 km/h, as the next highest value shown was >350 km/h, which is not likely real and does not have an accompanying trend.

Table 3
Statistical analysis of all variables collected with the PPM, and the share of not available values (NAs).

Variable	Median	Mean	SD	Max	Min	1stQ	3rdQ	NAs (%)
PM_1 (env)	9.0	15.0	31.2	3333.0	0.0	5.0	17.0	68.4
$PM_{2.5}$ (env)	12.0	22.3	49.3	3333.0	0.0	6.0	24.0	68.4
PM_{10} (env)	13.0	25.5	56.3	3333.0	0.0	7.0	26.0	68.4
PM_1 (std)	9.0	18.2	46.9	5000.0	0.0	5.0	17.0	68.4
$PM_{2.5}$ (std)	12.0	27.4	74.2	5000.0	0.0	6.0	24.0	68.4
PM_{10} (std)	13.0	30.4	84.3	5000.0	0.0	7.0	26.0	68.4
Temperature	24.1	24.1	2.8	37.6	1.6	22.7	25.4	68.4
Rel. Humidity	31.6	31.8	7.6	84.3	1.7	26.5	36.9	68.4
Battery	94.0	82.5	21.1	100.0	0.0	69.0	98.0	68.4
Altitude	272.1	170.2	158.6	1258.0	0.0	0.0	307.2	69.1
Speed	0.4	2.2	8.6	179.6	0.0	0.0	1.7	69.0

3.4. Smart activity tracker

Smart Activity Trackers (SAT) provided data based on measurements made by the Garmin Vivosmart 3 wristwatches worn by all participants at all times. Average heart rate per minute, number of steps, distance travelled (in meters), kcal, Metabolic Equivalent of Task (M.E.T.), motion intensity, and stress were recorded as numeric variables. Other variables provided by the SAT were strings/factors (sleep level, activity type, intensity). No outliers were removed from the SAT part of the dataset.

Calculated statistics for some of the numeric variables are shown in Table 4, showing median, mean, standard deviation, maximum, minimum, 1st quartile, 3rd quartile, and percent of NA values.

Table 4
Statistical analysis of eight numeric variables collected with the SAT, and the share of not available values (NAs).

Variable	Median	Mean	SD	Max	Min	1stQ	3rdQ	NAs (%)
Average heart rate	73.0	76.3	17.8	205.0	33.0	64.0	86.0	21.7
Steps	0.0	6.6	21.7	317.0	0.0	0.0	0.0	10.8
Distance	0.0	5.3	18.2	355.0	0.0	0.0	0.0	10.8
kcal	0.0	0.3	1.1	31.0	0.0	0.0	0.0	10.8
M.E.T.	0.1	0.2	0.4	45.6	0.0	0.1	0.1	10.8
Mean motion intensity	2.4	2.5	2.3	7.0	0.0	0.2	4.6	10.8
Max motion intensity	4.0	4.4	2.7	7.0	0.0	3.0	7.0	10.8
Stress	22.0	29.2	23.3	99.0	1.0	12.0	44.0	52.0

3.5. Indoor air quality station

Each participants' household had an Indoor Air Quality station (IAQ), in the form of a uHoo device, installed in the room that they used most frequently during the day. Ten variables were recorded using the IAQ: a timestamp, three variables recording atmospheric conditions (ambient temperature, relative humidity, and air pressure), and six air quality parameters ($PM_{2.5}$, Total Volatile Organic Compounds (TVOC), CO_2, CO, ozone, and NO_2). The statistical analysis of all the non-timestamp variables is shown in Table 5, showing median, mean, standard deviation, maximum, minimum, 1st quartile, 3rd quartile, and percent of NA values. Some outliers were removed for the IAQ section of the dataset, namely: 1) all TVOC values >1156 ppm, as this was the detection limit of the sensor, 2) CO_2 values >10,000 ppm, as it is not feasible, with only a single recording above this limit, showing a value of 1,18,000 ppm, 3) ozone values >50 ppm, deemed as not feasible, with two recordings showing higher values, both >900 ppm.

Table 5
Statistical analysis of variables collected with the IAQ, and the share of not available values (NAs).

Variable	Median	Mean	SD	Max	Min	1stQ	3rdQ	NAs (%)
Temperature	21.8	21.6	1.6	28.7	6.4	20.9	22.6	10.9
Relative humidity	45.6	45.7	9.4	77.4	11.3	39.1	52.4	10.9
$PM_{2.5}$	15.4	26.8	32.1	209.1	2.4	6.8	33.6	10.9
TVOC	92.0	190.6	247.8	1156.0	0.0	34.0	239.0	10.9
CO_2	827.0	923.4	420.5	4358.0	400.0	618.0	1116.0	10.9
CO	0.0	0.0	0.1	36.7	0.0	0.0	0.0	10.9
Air pressure	980.4	981.8	7.3	1007.3	965.4	976.9	985.5	10.9
Ozone	9.4	9.2	2.6	30.9	1.2	7.7	10.6	10.9
NO_2	8.5	15.5	20.8	317.8	0.0	0.6	24.8	10.9

3.6. Inhalation and intake dose

Based on published models an inhalation rate was calculated by using the recorded average heart rate data [5,9]. Additionally, the model required information on age, gender, and height. By using this calculated inhalation rate [9] and the recorded $PM_{2.5}$ concentration, an inhalation adjusted exposure rate or intake dose was calculated. The information on the weight of the participant was needed to calculate the dose per kg of body mass. Fig. 2 shows violin and box plots of the calculated minute resolution intake doses per age, gender, height, and weight.

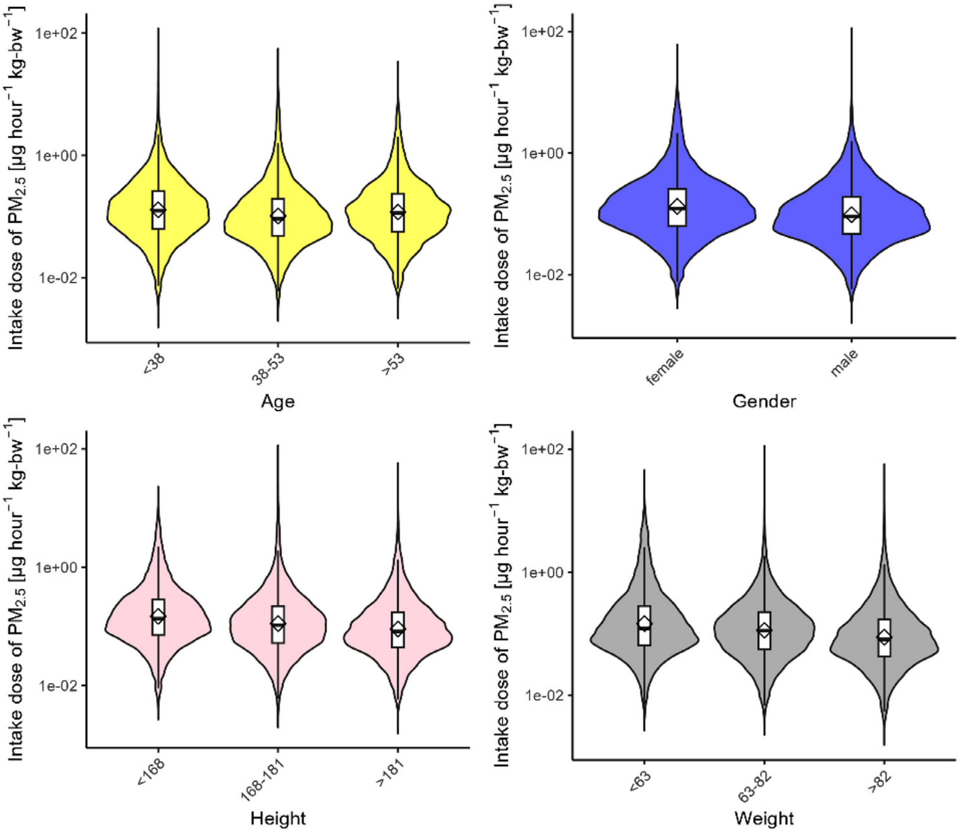

Fig. 2. Violin and box plots for the calculated intake dose per age, gender, height, and weight. Scale is logarithmically transformed to allow a more informative visualization.

4. Experimental Design, Materials and Methods

The dataset described in this paper has been collected in the scope of the ICARUS H2020 project (icarus2020.eu). It represents the data collected in the municipality of Ljubljana, Slovenia. Each individual involved in the data collection provided data by 1) providing data on their age, height, gender, and weight, 2) filling out a detailed time activity diary (TAD) each day, 3) a personal particulate matter monitor (PPM), with the Plantower PMS5003 sensing unit, worn on their clothes, backpack or set nearby, 4) wearing a smart activity tracker (SAT) in a form of a Garmin Vivosmart 3 wristband, 5) hosting an indoor air quality sensor (IAQ) station, the uHoo, set in their residence.

4.1. The TAD

One time activity diary (TAD) per day was provided to each individual involved in the sampling campaign. Per the campaign protocol, the goal was for each participant to collect data for seven days, consequently receiving seven TADs. However, in multiple cases this period was somewhat shorter or longer, depending on the availability of the participant and their willingness to collaborate. Ideally, the same participant collaborated in the heating (Feb–Mar) and non-heating (Apr–May) campaign, proving two sets of seven TADs altogether. Participants were required to fill out the TADs even on starting and ending days, when the measuring period did not encompass the entire day. One TAD was printed per A4 page, physically delivered to the participants, and later collected and digitized. It contained 24 lines, one for each hour of the day, and 27 columns, listing the time and day, 4 house conditions and 21 activities, as listed in Table 2.

4.2. The PPM

In the context of the ICARUS study, each participant was equipped with a Personal Particulate Matter (PPM) monitor incorporating a pms5003 sensor from Plantower, Nanchang Panteng Technology Co., Beijing, China. Accompanying the device, a micro-USB charger was distributed to ensure continuous operation, with the PPM requiring 3–4 h for a full charge and functioning for 6–8 h thereafter. Participants were advised to maintain the device charged, necessitating proximity to a power source.

The PPM was attached to clothing or a personal item, such as a backpack or handbag, to facilitate constant monitoring during various activities, including transitioning between different environments. Data acquisition was dual-faceted: storage on an internal SD card and concurrent real-time transmission to a server via GSM, allowing for immediate device status verification. Upon completion of the data collection phase, the devices were retrieved, and the stored data were transferred to a local system. Similar to the SAT protocol, participants did not have real-time access to the data but were provided with a detailed individual report post-campaign.

The PPM's primary function was to record geographic and particulate matter (PM) concentration data, alongside ambient temperature and humidity readings. The device, constructed on an Arduino base, featured the pms5003 sensor as its core component. This sensor, a nephelometer, operates on the light scattering principle to deduce particle size and mass concentration.

The pms5003 sensor, an optical particle sensor, employs a fan to draw air through a light beam, with particles scattering light onto a photodetector. The resulting electrical signal is then analysed to count particles by size class, specifically those with aerodynamic diameters ranging from 2.5 to 10 µm, and to estimate mass concentration per size class. However, inherent assumptions regarding particle size and composition can affect the accuracy of these sensors, necessitating regular calibration to maintain data integrity.

4.3. The SAT

Every individual taking part in the ICARUS study was given a SAT, chosen from two different sizes. They were advised to wear the SAT on their dominant wrist, ensuring it was secure yet comfortable. Before their initial meeting to receive the SAT, participants had their personal details (including gender, age, height, and weight) collected and entered into the device. Each participant was also provided with a SAT charger. The SAT was designed to last approximately five days on a single charge. The data collected by the participants was saved on the SAT and, after the collection phase, a field worker or researcher would upload it to the Garmin server. This upload could be done using a USB charging cable or wirelessly through a Bluetooth connection to a smartphone with the Garmin Connect app installed. During the data collection phase, participants were not able to access their data and did not have the Garmin Connect app on their smartphones.

The Garmin Vivosmart 3 is outfitted with four sensors: a Garmin Elevate wrist heart rate monitor, a barometric altimeter, an accelerometer, and an ambient light sensor. It gauges heart rate using the photoplethysmography (PPG) technique, which involves a light source and a sensor on the skin detecting blood flow changes [10]. The device opts for a green LED for its deeper skin penetration. The wrist, offering cost-effectiveness, portability, and ease of use, is the preferred site for the PPG sensor [10]. The heart rate sensor is positioned on the underside of the SAT, in contact with the skin. Garmin's software leverages heart rate data, including variability, to calculate physical activity levels, stress, sleep patterns, breathing rate, calorie expenditure, and more. While the heart rate data is collected in per-second intervals, the reported data in the ICARUS campaign is the average heart rate per minute.

4.4. The IAQ

During the ICARUS campaign, each participating home was equipped with an Indoor Air Quality (IAQ) station, a multi-sensor device from uHoo Limited, Hong Kong, to monitor various parameters indicative of indoor air quality. This device measured ambient temperature, relative humidity, air pressure, $PM_{2.5}$ levels, and a range of gases (CO_2, NO_2, CO, TVOC, and O_3).

A single IAQ station was allocated to each household, with installation conducted by a field worker/researcher on their initial visit. The installation process required verification from a household member that a Wi-Fi network was available, with the option for the participant to either share the Wi-Fi password with the field worker or enter it themselves for privacy. The optimal placement of the IAQ station was collaboratively determined by the participant and the field worker, typically in a common area such as the living room or a combined space that included the living room, dining room, and kitchen.

Once installed, the IAQ station began transmitting data to the uHoo server, a process that was verified by the field worker. Participants were advised to avoid any interaction with the device throughout the monitoring period unless directed by the research team. At the end of the sampling period, the IAQ station was disconnected and retrieved from the household. Data from the uHoo server was then downloaded to a local device and integrated with the broader dataset collected during the campaign.

Limitations

The dataset focuses solely on Ljubljana's municipality, posing a geographical constraint. The transition between non-heating and heating seasons was narrow. Some challenges arose with the SAT during data collection. Despite instructions, not all participants wore the device snugly, causing gaps in heart rate data collection. Data was stored locally, preventing real-time monitoring for malfunctions. PPM devices often malfunctioned, resulting in data gaps. Researchers had to prompt participants for adjustments, adding inconvenience. These devices also had short battery lives, and when charged, affected the integrated temperature and humidity sensor data. Vigorous movements occasionally halted the device, and its size posed wearability issues during sports. Time stamp inconsistencies meant data rounding. The IAQ device needed a constant Wi-Fi connection, as per the manufacturers design. Disruptions led to data gaps, and power outages also halted data recording. As the TADs had only one hour activity resolution, it was difficult to determine the exact amount of the participant performed an activity.

Ethics 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 August 22, 2018).

The data in this paper were selected only from participants in Slovenia. Informed consent was obtained from all subjects involved in the study.

CRediT Author Statement

Rok Novak: Validation, Formal analysis, Resources, Data curation, Writing – original draft preparation, Writing – review and editing, Visualization. **Johanna A. Robinson:** Resources, Data curation, Writing – review and editing. **Tjaša Kanduč:** Resources, Data curation, Writing – review and editing. **Dimosthenis Sarigiannis:** Project administration, Funding acquisition. **David Kocman:** Validation, Formal analysis, Resources, Data curation, Writing – review and editing, Supervision, Project administration.

Data Availability

Ljubljana Multi-Sensor Indoor and Outdoor PM Exposure, Environment, and Personal Health Dataset (Original data) (Zenodo)

Acknowledgments

The authors acknowledge and thank all the participants that volunteered to contribute to the sampling campaign in Ljubljana, Slovenia. Moreover, the authors express their gratitude to other ICARUS H2020 partners, including the teams responsible for designing some of the devices and protocols, and maintaining the servers for project-specific data storage.

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, P1-0143 program "Cycling of substances in the environment, mass balances, modelling of environmental processes and risk assessment«, all funded by the Slovenian Research and Innovation Agency.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the author(s) used Chat GPT v4 in order to provide some minor language editing. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Novak, J.A. Robinson, T. Kanduč, D. Sarigiannis, D. Kocman, Simulating the impact of particulate matter exposure on health-related behaviour: a comparative study of stochastic modelling and personal monitoring data, Health Place 83 (2023) 103111, doi:10.1016/j.healthplace.2023.103111.

[2] S. Lim, E. Bassey, B. Bos, L. Makacha, D. Varaden, R.E. Arku, J. Baumgartner, M. Brauer, M. Ezzati, F.J. Kelly, et al., Comparing human exposure to fine particulate matter in low and high-income countries: a systematic review of studies measuring personal PM2.5 exposure, Sci. Total Environ. 833 (2022) 155207, doi:10.1016/j.scitotenv.2022.155207.
[3] R. Novak, J.A. Robinson, C. Frantzidis, I. Sejdullahu, M.G. Persico, D. Kontić, D. Sarigiannis, D. Kocman, Integrated assessment of personal monitor applications for evaluating exposure to urban stressors: a scoping review, Environ. Res. 226 (2023) 115685, doi:10.1016/j.envres.2023.115685.
[4] R. Novak, J.A. Robinson, T. Kanduč, D. Sarigiannis, D. Kocman, Assessment of individual-level exposure to airborne particulate matter during periods of atmospheric thermal inversion, Sensors 22 (2022) 7116, doi:10.3390/s22197116.
[5] R. Novak, D. Kocman, J.A. Robinson, T. Kanduč, D. Sarigiannis, M. Horvat, Comparing airborne particulate matter intake dose assessment models using low-cost portable sensor data, Sensors (2020) 20, doi:10.3390/s20051406.
[6] R. Novak, I. Petridis, D. Kocman, J.A. Robinson, T. Kanduč, D. Chapizanis, S. Karakitsios, B. Flückiger, D. Vienneau, O. Mikeš, et al., Harmonization and visualization of data from a transnational multi-sensor personal exposure campaign, Int. J. Environ. Res. Public Health 18 (2021) 11614, doi:10.3390/ijerph182111614.
[7] J.A. Robinson, R. Novak, T. Kanduč, T. Maggos, D. Pardali, A. Stamatelopoulou, D. Saraga, D. Vienneau, B. Flückiger, O. Mikeš, et al., User-centred design of a final results report for participants in multi-sensor personal air pollution exposure monitoring campaigns, Int. J. Environ. Res. Public Health 18 (2021) 12544, doi:10.3390/ijerph182312544.
[8] ICARUS2020 ICARUS2020 Available online: https://icarus2020.eu/(accessed on 28 January 2022).
[9] R. Greenwald, M.J. Hayat, E. Dons, L. Giles, R. Villar, D.G. Jakovljevic, N. Good, Estimating minute ventilation and air pollution inhaled dose using heart rate, breath frequency, age, sex and forced vital capacity: a pooled-data analysis, PLOS One 14 (2019) e0218673, doi:10.1371/journal.pone.0218673.
[10] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, H. Nazeran, A review on wearable photoplethysmography sensors and their potential future applications in health care, Int. J. Biosens. Bioelectron. 4 (2018) 195–202, doi:10.15406/ijbsbe.2018.04.00125.