Towards a harmonized database of indoor air contaminant concentrations: methods and application to CO2

Rojas, G., Staffer, R., Casquero-Modrego, N., Loomans, M., Abadie, M., Alhindawi, I., Babich, F., Bekö, G., Coggins, M., Cremers, B., Damberger, B., Freundorfer, T., Garcia-Ortega, S., Hassan, H., Jones, B., Justo Alonso, M., Lara-Ibeas, I., McGill, G., McGrath, J., ... Toledo, L. (2026). Towards a harmonized database of indoor air contaminant concentrations: methods and application to CO2. *Building and Environment*, *292*, Article 114297. https://doi.org/10.1016/j.buildenv.2026.114297

Published in:
Building and Environment

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2026 the authors.

This is an open access article published under a Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/.), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Download date:09. Nov. 2026

Building and Environment 292 (2026) 114297

Contents lists available at ScienceDirect

Building and Environment

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

Towards a harmonized database of indoor air contaminant concentrations: Methods and application to CO_2

Gabriel Rojas [a,*], Reto Stauffer [b], Núria Casquero-Modrego [c], Marcel Loomans [d], Marc Abadie [e], Ibrahim Alhindawi [g], Francesco Babich [j], Gabriel Bekö [f], Marie Coggins [g], Bart Cremers [h], Bernhard Damberger [i], Timm Freundorfer [a], Sonia Garcia-Ortega [k], Hala Hassan [g], Benjamin Jones [l], Maria Justo Alonso [m], Irene Lara-Ibeas [j], Grainne McGill [n], James McGrath [o], Aurora Monge-Barrio [p], Constanza Molina [q], Alejandro Moreno-Rangel [n], Peter Tappler [i], Linda Toledo [n]

[a] Unit of Energy Efficient Building & Digital Science Center, University of Innsbruck, Austria
[b] Department of Statistics & Digital Science Center, University of Innsbruck, Austria
[c] Residential Building Systems Group, Lawrence Berkeley National Laboratory (LBNL), Berkeley, USA
[d] Eindhoven University of Technology, the Netherlands
[e] LaSIE (UMR CNRS 7356) - La Rochelle University, France
[f] International Centre for Indoor Environment and Energy, Department of Environmental and Resource Engineering, Technical University of Denmark
[g] School of Natural Sciences & Ryan Institute, University of Galway, Ireland
[h] Zehnder Group International, Zwolle, the Netherlands
[i] IBO Innenraumanalytik OG, Vienna, Austria
[j] Institute for Renewable Energy, Eurac Research, Bolzano, Italy
[k] Eduardo Torroja Institute for Construction Sciences (IETcc), Spanish National Research Council (CSIC), Spain
[l] Department of Architecture & Built Environment, University of Nottingham, UK
[m] Department of Architecture, Materials and Structures, SINTEF Community, Trondheim, Norway
[n] Department of Architecture, University of Strathclyde, Glasgow, UK
[o] Department of Physics, Maynooth University, Ireland
[p] School of Architecture, Universidad de Navarra, Spain
[q] Escuela de Construcción Civil, Pontificia Universidad Católica de Chile

ARTICLE INFO

Dataset link: https://zenodo.org/records/14917724

Keywords:
Indoor air quality
Database
Measurement data
Contaminants
CO_2 concentration
Residential buildings
Regression analysis

ABSTRACT

Reliable data on indoor air quality (IAQ) is essential for design and operation of energy-efficient and healthy residential buildings. This work introduces a standardized methodology to systematically compile, process, and analyze IAQ measurement data, enabling the creation of a harmonized IAQ database. Developed within the IEA EBC Annex 86 project, the open-source algorithm ensures compatibility, data protection, and scalability by aggregating time series measurements into monthly summaries while preserving distributional information and linking them to relevant meta-information. This methodology was applied to create a comprehensive, open-access dataset integrating indoor air contaminant measurements from numerous IAQ studies covering residential homes with varied building types, ventilation strategies, climates and occupancy patterns. The dataset enables consistent statistical analysis and interpretation of IAQ data collected through heterogeneous methods, supporting robust cross-study comparison and benchmarking. As the dataset grows, it will enable harm-based analyses, an emerging paradigm in indoor air quality regulation.

The paper exemplifies the utility of the harmonized database by analyzing CO_2 concentration data from 18 studies encompassing over 1000 homes, with a focus on bedrooms. The results quantify the concentration distribution encountered in bedrooms across diverse regulatory and building contexts. Results highlight significant variations in CO_2 levels influenced by ventilation type, building characteristics, and ambient temperature, emphasizing the importance of standardized data for advancing IAQ research, policy development, and occupant health assessments. Ultimately, the harmonized dataset and methodology can serve as a critical resource for

* Corresponding author.
E-mail address: gabriel.rojas@uibk.ac.at (G. Rojas).

https://doi.org/10.1016/j.buildenv.2026.114297
Received 19 November 2025; Received in revised form 12 January 2026; Accepted 25 January 2026
Available online 26 January 2026
0360-1323/Crown Copyright © 2026 Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

researchers, practitioners, and regulators aiming to optimize indoor environmental quality while advancing energy-efficient residential building design globally.

1. Introduction

Buildings account for approximately 30 % of global final energy consumption, positioning the built environment at the center of climate mitigation strategies [36]. While energy efficiency has traditionally dominated policy efforts, there is growing recognition that climate goals must be pursued alongside a healthy indoor environment, certainly also in residential buildings where individuals spend the majority of their time [39,75]. As buildings become more energy efficient with increased airtightness, reduced infiltration and the lack of adequate ventilation provisions can elevate indoor contaminant levels, such as radon or cooking fumes, and cause issues like dampness, mold, or overheating, all detrimental to occupant health [43,47,72].

In response, several countries have begun integrating indoor environmental quality (IEQ) metrics into building regulations, putting, amongst others, indoor air quality (IAQ) and adequate building ventilation stronger in the design focus. A policy example is the revised Energy Performance of Buildings Directive [22], which explicitly defines IEQ and establishes it as a co-benefit of energy efficiency. This directive requires European Union member states to set national IEQ standards and mandates IAQ monitoring and control in non-residential buildings, reflecting the shift toward holistic building performance frameworks that support both decarbonization and occupant health.

A key challenge in advancing these integrated standards lies in the availability and consistency of data. Historically, data-driven efforts have played a foundational role in shaping IEQ standards. For example, the ASHRAE Thermal Comfort Database compiled over 22,000 subjective-objective field observations across 52 studies to inform the adaptive thermal comfort model [17,26]. However, for IAQ in particular, data remain fragmented and difficult to access, limiting model development and policy alignment. Early calls for harmonized contaminant databases [34] emphasized the urgent need for well-designed, accessible databases of measured contaminant emission data to facilitate IAQ modeling. Responding to this challenge, initiatives such as PANDORA [1,2], have emerged, compiling emission rates of both gaseous and particulate contaminants from various indoor sources to provide up-to-date data for researchers and practitioners. More recently, international collaboration has been established to address another data gap. In 2020, the International Society of Indoor Air Quality and Climate (ISIAQ) launched Scientific and Technical Committee STC34 to compile and maintain global IEQ guidelines [74].

However, no database collecting contaminant measurements from occupied buildings is available up to date. There have been many post-occupancy IAQ evaluations in new or retrofitted buildings [63], assessing the real-world performance of energy-efficient buildings and/or identify unintended consequences, such as under-ventilation, elevated contaminant concentrations, or overheating [13,14,19,21,25,40,55,73]. But differences in methodologies and reporting, in particular in terms of assessment metrics and meta-information, make it difficult to utilize the wealth of existing measurement data to draw broad conclusions about how different ventilation concepts may impact IAQ performance and therefore occupant health. Based on the original work by Logue et al. [44], Morantes et al. [47] have provided a framework to assess health impacts from residential indoor air contaminants. Given concentration and exposure time, this approach estimates the chronic harm caused to occupants, expressed in disability adjusted life years (DALYs), an emerging paradigm in IAQ regulation. This harm-based assessment method underpins ASHRAE 62.2, is being adopted in ASHRAE 62.1, and is influencing new European standards. To support more consistent harm-based benchmarking and to derive conclusions on the effectiveness of IAQ management strategies, the access to databases containing empirical observed contaminant concentration distributions is critical. The creation of such a database that can pool IAQ measurement data from around the world requires a harmonized data structure, and a standardized processing and analysis methodology. This work presents the development of such a method, the preliminary creation of a harmonized database from 18 different IAQ studies and its application for the meta-analysis of CO_2 concentration measurements.

Humans emit carbon dioxide (CO_2) as a product of their metabolism. At concentration levels typically encountered indoors, CO_2 itself cannot be linked with impacts on perceived air quality, health or work performance with certainty as inconsistent results from different studies exist [18,24,46]. Nevertheless, as the emissions of other pollutants from human activity correlate with CO_2 emissions, indoor CO_2 concentrations have long been used as a proxy for bio-effluent exposure, perceived air quality and/or ventilation adequacy. But without knowledge or control of potential pollutants stemming from other sources, it should not be used as a direct indicator of overall IAQ. Even when used to assess ventilation, one needs to consider the type of space and its occupancy [56,57].

As a metric for ventilation adequacy, elevated CO_2 levels have been linked to reduced cognitive performance and poor sleep quality [23,38,78]. In residential settings, the dilution of bio-effluents and the removal of humidity are likely to be the driving factors for choosing ventilation rates, if adequate source-control measures are in place [60], e.g. the use of low-emitting building products, cooktop-ventilation, radon control. But relying on a fixed CO_2 threshold like 1000 ppm might be problematic, as concentrations are influenced by occupancy characteristics, activity, ventilation, and air distribution, highlighting the need for updated, ventilation guidelines for energy-efficient buildings [3,56]. As a result, there is a need for a better understanding how ventilation rates and contaminant concentrations, including CO_2, vary across residential homes, building types, countries, climates, etc. [53]. Together, these circumstances highlight the need for a more integrated and data-driven approach to IAQ assessment, also providing the necessary data for harm-based analysis.

This paper contributes to these efforts by introducing a standardized methodology to systematically compile, process, and analyze indoor contaminant concentrations in residential buildings from different sources. Developed within the context of an IEA project [35], this work presents an open-source algorithm for the generation of harmonized IAQ datasets. The benefits and potential of such an IAQ database are demonstrated exemplarily by analyzing the CO_2 concentration from 18 studies with over 1000 homes while focusing on bedrooms.

2. Methods

The methods section is divided in the following two subsections:

- "Algorithm" introduces the newly developed algorithm to generated harmonized datasets suitable for IAQ databases and,
- "Application" demonstrates the use of this algorithm to generate harmonized datasets and produce a meta-analysis from IAQ 18 studies covering more than 1000 residential homes.

The algorithm has been implemented as open-source R code [70] and is available at this GitHub URL: https://github.com/IEA-EBC-Annex86/annex.

The harmonized dataset has been published in a Zenodo repository [64] and can be downloaded under this URL: https://zenodo.org/records/14917724.

2.1. Algorithm

2.1.1. Overall concept

The creation of a harmonized IAQ database requires a standardized methodology and a harmonized data structure to systematically compile, compare, and analyze indoor measurements stemming from different sources. The developed methodology is able to accommodate different kinds of data sources, e.g., measurements from short-term (few days or weeks) vs. long-term (several months or years) campaigns, measurements with different sampling rates using different measurement devices, etc. and allow a unified assessment.

Such a unified analysis could be accomplished by centrally collecting the original time series data from the different studies and processing those data sets in a standardized way jointly. Alternatively, the datasets can be processed in a decentral manner, i.e. the datasets stay with the data owners. In this approach the standardized analysis method is applied by the research institution that collected the data and only the output of the statistical analysis, enriched with relevant meta-information, is collected centrally for the meta-analysis. In this work, the latter approach was chosen for two main reasons:

- Data protection: Since the time series data can contain a wealth of time-resolved information with respect to home occupancy, occupant activities, etc., potential data protection issues can be avoided. Since the introduction of the General Data Protection Regulation [30] in the European Union, some uncertainties among research institutions have arisen on how GDPR-compliant data handling of IAQ data should look like. Consequently, very restrictive consent declaration forms were drafted and used in IAQ field studies, limiting the processing of the original data to the research institution performing the measurements. Therefore, the chosen approach avoids the need to hand out the original time series data to a third-party data processing team, and only highly aggregated (monthly) data is collected centrally.
- Scalability and continuity: The aggregated datasets from the statistical analysis (and its respective meta-information) can be processed and stored easier by potential users and databases, if they have a standardized format and a notably reduced size compared to the original time series data sets. The decentral approach facilitates the continuous integration of additional data from new IAQ studies extending and updating the database and potential products resulting from it, while minimizing the workload of database maintainers.

2.1.2. Analytical procedure

The following analytical procedure is applied for measurement data obtained from IAQ studies. Apart from the collection of meta-information, statistics are determined on four distinct levels, namely for each study, home, room, and contaminant (relevant measured variable). To allow a distinct assessment of a certain time of the year or time of day, the statistics are calculated for each month of the year, and for daytime (07:00 – 23:00 h) as well as nighttime (23:00 – 07:00 h) hours. This results in a data structure as depicted in Fig. 1. More information on the collected meta-information can be found in Appendix A.

The procedure is designed towards the processing of time series data gathered during IAQ measurement studies, e.g., CO_2 or $PM_{2.5}$ concentrations measured over several days, weeks or years. However, the data structure can also be used to accommodate single-value measurements, stemming from e.g. active or passive sampling techniques for volatile organic compounds (VOC) measurements. For time series data, the following statistical information and data quality indicators are derived for each branch of this data structure, i.e. for each possible combination of "User-Study-Home-Room-Contaminant-Time of Year-Time of Day" (see Fig. 1). It will be referred to as (data) tuple in the remainder of the text:

- arithmetic mean
- standard deviation
- percentiles from 0 to 100 % in 1 %-steps, plus 0.5 %, 2.5 %, 97.5 % and 99.5 % resulting in 105 percentile values (note that median and lower and upper quantile are also included)
- start- and end-date
- statistics on sampling interval: mean, median, min, max, lower and upper quantile
- total number of data points and number of invalid data points (non-numeric entries)
- estimated number of data points assuming an uninterrupted measurement from start- to end-date with a sampling interval equal the observed median sampling interval
- percentage of data points exceeding a lower or upper plausibility threshold (see below)

Thus, each tuple summarizes the monthly conditions for a given "User-Study-Home-Room-Contaminant-Time of Year-Time of Day" combination. Note that even if the measurement duration was short, e.g. a few days, this short period will be summarized in a monthly tuple. Subsequent processing will have to take this into account, e.g. by using the data quality indicators to calculate a weighting factor representing

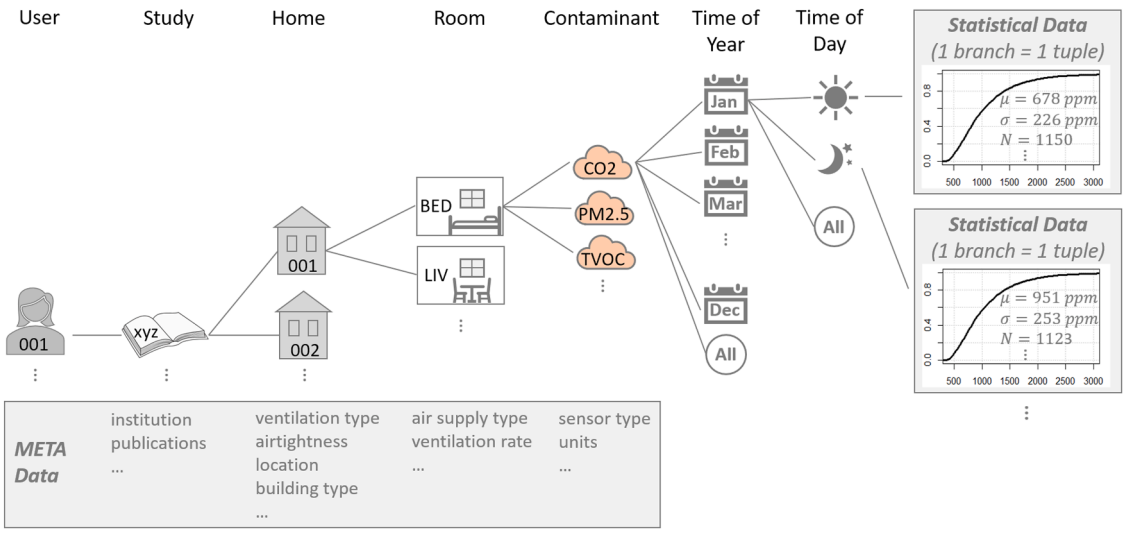


Fig. 1. Schematic overview of data structure of the presented analytical procedure.

the data coverage for that month as done in the meta-analysis (see Section 2.2.2.3). In any case, the following cleaning and selection steps are performed within the standardized analysis procedure to remove tuples with low statistical explanatory power. An important assumption is that systematic errors in the measurement data have been removed or corrected beforehand, e.g. CO_2 sensor drifts, cross-sensitivities, etc.

- Invalid non-numeric entries or data gaps are marked as NA and excluded from further analysis.
- The minimum sample size (valid number of measurement values) needed to calculate all the statistical measures as described above is 30.
- If the number of valid data points is lower than 30, but at least two, the data quality indicators and the percentiles are calculated. However, mean and standard deviation are set as NA as their explanatory power becomes low.
- If there is only one data point (measurement value), sampling interval statistics as well as the estimated number of data points will also be set to NA.

In addition to this data cleaning procedure, values are checked for plausibility. For each possible contaminant (measured variable) a plausible range is defined, see Table 1. Unrealistically low or high values are identified, the user (of the executing script) is warned, and the respective amount of exceedance (% of data points below or above threshold) is provided as a data quality metric in the analysis results (e.g. 1.5 % of measured CO_2 concentration $<$ 350 ppm. This lower CO_2 threshold was chosen to allow historic data and/or small, possibly short, sensor drifts.).

Each tuple, containing the statistics and the data quality indicators, is represented by a single row in the output table. E.g., if an IAQ study measured CO_2, T and RH in bedroom and living room for six consecutive months, the results are summarized in 126 rows, i.e. 2 rooms x 3 variables/contaminants x 7 (6 months + entire period) x 3 (daytime, nighttime and 24-hour-day). An example with three tuples is shown in Fig. 2.

Table 1
List of measured variables that can be processed and their respective attributes, including their lower and upper bounds for the plausibility check and their default and allowed units. A binary field indicating if additional information about the variable, is mandatory or not is also included.

Measured Variable	Additional info. req'd	Lower bound	Upper bound	Default (allowed) units
CO_2	No	350	1.0E+04	ppm (%)
CO	No	0	1.0E+02	µg/m^3 (mg/m^3, ppm, ppb)
HCHO	No	0	1.0E+03	µg/m^3(mg/m^3, ppm, ppb)
UFP	No	0	1.0E+09	#/cm^3
PM_1	No	0	1.0E+04	µg/m^3 (mg/m^3)
$PM_{2.5}$	No	0	1.0E+04	µg/m^3 (mg/m^3)
PM_{10}	No	0	1.0E+04	µg/m^3 (mg/m^3)
PM_{Other}	Yes	n.a.	n.a.	n.a.
NO_2	No	0	1.0E+04	µg/m^3 (mg/m^3)
NO_x	No	0	1.0E+04	µg/m^3 (mg/m^3)
TVOC	Yes	0	1.0E+04	µg/m^3 (mg/m^3)
VOC	Yes	0	1.0E+04	µg/m^3 (mg/m^3)
O_3	No	0	1.0E+02	µg/m^3 (mg/m^3, ppm, ppb)
RH	No	5	105	%, -
Radon	No	0	1.0E+04	Bq/m^3 (pCi/L)
Fungi	No	0	1.0E+09	CFU/m^3
Ions	No	0	1.0E+09	1/cm^3
Other	Yes	n.a.	n.a.	n.a.
T	No	−50	50	C, K, F
Pressure	No	500	1100	hPa, Pa, mbar, mmHg
SolRad	Yes	0	1500	W/m2
Flow	No	0	500	l/s, m3/h, cfm

2.1.3. Implementation in R

The analytical procedure described above was implemented in R, a programming language for statistical computing and data visualization [58]. For easier usability, distribution, and maintenance the algorithm has been made available as an R package [70]. Prior to launching the analysis functions, the user has to import the measurement data as proper R data frames. The user must also assign contaminant name, unit of measurement as well as a home and room identifier to each column of the data frame within an external configuration file or within a configuration variable.

After importing the time series datasets and configuration file, the following functions are available to perform the standardized analysis in a few steps as shown in Fig. 3. At this stage it is assumed, that systematic measurement errors have been addressed (e.g. set to "NA") and that the date/timestamp is interpreted correctly.

1. annex_check_config() can be used to check if the configuration file was defined correctly, and all the necessary data columns have been assigned correctly.
2. annex_prepare() prepares the format of the measurement data set for further processing steps (can be bypassed if data frame is provided as needed for the next step).
3. annex() is the creator function for annex objects. It defines which measured variables (contaminants) should be processed and checks for potential duplicate measurements.
4. annex_stats() takes up the annex object from the previous step to calculate the statistics (aggregated statistical measures) and returns an annex_stats object.
5. annex_write_stats() writes the object from the previous step into the output XLSX file. It contains the statistics as well as sheets for the required meta-information. This information has to be filled in manually or in a post-processing step by the user.
6. annex_validate() can be used to check if all meta-information has been filled in properly.

The analysis functions will take care of proper unit conversions and account for time zones and/or time changeovers when calculating the statistics and data quality indicators as described above.

Table 1 shows the contaminants (and other variables) that have been defined and can be processed currently with these R analysis functions. It also shows the allowable units (where unit conversion has been defined) and lower and upper bound values used for calculation of data quality indicators. The column "Additional info. req'd" defines if additional information must be provided by the user when filling out the meta-information regarding any of the measurement variables.

The documentation for the tool with detailed description of the functions and a tutorial for two different sample datasets has been developed and can be found at [70].

2.1.4. Providing meta-information

The analytical functions implemented in R and described above provide the results in a standardized spreadsheet file containing the statistical results in one sheet and four additional sheets to provide relevant meta-information. The meta-information that can be provided, is structured according to the four levels "study", "home", "room" and "variable" and contains fields such as:

- Study: contact, institution, publications, links, additional information
- Home: country, city, ventilation type, ventilation rate on home level, airtightness, type of building, energy standard, year of construction or renovation, size of home, type of occupants, further comments, …

user	study	home	room	year	month	tod	variable	quality_lower	quality_upper	quality_start	quality_end	interval_Min	interval_Q1	interval_Median	interval_Median	interval_Mean	interval_Q3	interval_Max	Nestim	N	NAs	Mean	Sd	p00	p00.5	p01
0010	NL_ML	H0100	BED2	2012	11	all	CO2	0	0	28-11-2012	30-11-2012	58	180	180	180	155.6	180	180	1440	1191	1	518.9146	92.66	368.4967	370.6336	372.9723
0010	NL_ML	H0100	BED2	2012	11	all	RH	0.08	0	28-11-2012	30-11-2012	58	180	180	180	155.6	180	180	1440	1191	0	39.7517	3.3566	0	34.73	34.825
0010	NL_ML	H0100	BED2	2012	11	all	T	0	0	28-11-2012	30-11-2012	58	180	180	180	155.6	180	180	1440	1191	0	19.8708	0.4882	4.69	19.39	19.39

Fig. 2. Example of three data tuples with the results of the statistical analysis of the time series for the month of November, for CO_2 in ppm, temperature (T) in °C and relative humidity (RH) in %. Time of day (tod) is "all", i.e. 0–24 h. Note that the remaining columns with all the percentiles are not shown here.

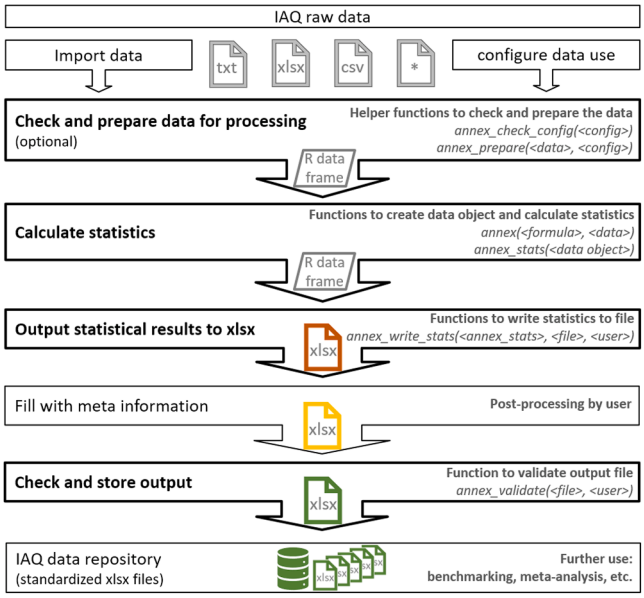


Fig. 3. Schematic representation of the developed workflow of the standardized data analysis process for IAQ measurement data. For the steps drawn with bold frames, easy to use functions have been programmed and made available in the R package "annex".

- Room: type of fresh air supply, ventilation rate, method of ventilation rate determination, type and number of occupants, further comments, …
- Variable: additional information on the measured variable, unit, measurement device, further comments.

Not all studies will have enough information available to fill all fields. Therefore, most fields are considered optional and only an informative note about missing fields will be printed when the validation function is run. The only fields which are considered essential and must be filled out for each home are country and ventilation type, otherwise a warning will be given by the validation function. If ambiguous measurement variables, such as "VOC, "TVOC", "PMOther" or "Other" have been processed, then the field "additional information" for these variables must be filled out too. For comparability and compatibility reasons, some of the meta-information fields allow only a defined set of answers (in form of a dropdown list). E.g. ventilation type must be selected from one of the following four categories:

- Window airing (not designed): homes which are only ventilated via occupant operated windows (besides natural in-/exfiltration), often also referred as purged ventilation,
- Natural ventilation (designed): homes which have been designed for natural ventilation with dedicated shafts, openings, automated window control, etc.,
- Hybrid/mixed mode ventilation: homes which apply natural with a supplemental mechanical ventilation including homes with intermittent, occupant driven fans,
- Mechanical ventilation: homes with continuous mechanical ventilation, e.g. mechanical extract ventilation (MEV), or balanced mechanical ventilation with heat recovery (MVHR).

Other predefined entries are summarized in Appendix A.

2.2. Application

2.2.1. Generation of harmonized IAQ datasets

To showcase the application of the proposed standardized analysis procedure, the measurement data from 18 IAQ studies have been processed as described above. The studies are summarized in Table 2. Fig. 4 provides an overview of the countries where the homes are located as well as their composition in terms of ventilation and building type. More background information and a short description are available in S1 (supplemental material). The CO_2 measurements (analyzed in Section 3) were performed with nondispersive infrared sensors (NDIR) in all studies. More details regarding measurement method can be found in the respective publications (see column Reference in Table 2). The resulting outputs from the herein presented analysis procedure (and the available meta-information) have been published in an open repository [64]. This repository can be considered a preliminary form of a harmonized IAQ database. It has been used for the meta-analysis described in the next subsection.

2.2.2. Meta-analysis of CO_2 concentration

The harmonized dataset has been used in a meta-analysis to investigate the CO_2 concentration in 1166 residential homes from 11 countries. CO_2 concentration values were available throughout all studies presented in Table 2. The workflow is illustrated in Fig. 5 and is described in the following subsections. The respective programming scripts (R, version 4.4.2) have been made available [62]

2.2.2.1. Read data. All output files generated with the algorithm described above and published in [64] are imported and processed in R. Therein, each row represents one (data) tuple, containing the statistical indicators (mean, standard deviation, percentiles) and the data quality indicators as described above. For each contaminant, there is one tuple for each room (of each home) for each month where data has been collected, which will be referred to as room-month tuples in the remaining text. As described above, room-month tuples were generated for different times of the day, i.e. daytime hours (07–23 h), nighttime hours (23–07 h) and for the entire day (0–24 h). Note that for each room-month tuple the empirical cumulative distribution function (eCDF) can be generated via the provided percentiles.

Depending on when the measurements in that particular month started or ended, this tuple might represent the statistics for only a few hours or up to the entire month. At this point, some basic data cleaning was performed to remove empty or incomplete tuples, and tuples that

Table 2
Summary table of IAQ studies included in this meta-analysis.

ID	Project Name (Study acronym)	Reference	Location	Country	New Construction / Existing Homes	Miscellaneous	Objective	Number of Homes	Standard / Code	Variables / Contaminants Measured	Measurement duration (CO_2)	Rooms (CO_2)
1	Lüftung 3.0 (Lue3)	[76]	Different Locations	Austria	New Construction (2010–2012)		Difference in IAQ equipped with and without mechanical balanced ventilation	123		CO_2, T, RH. VOCS, aldehydes, mould spores, dust mite allergens, radon Air volume flows, noise levels	2 × 1 week (3 and 15 months after move-in, no summer months)	Bedroom
2	Lodenareal and Kufstein (Lod_Kuf)	[65]	Innsbruck and Kufstein	Austria	New Construction (2008–2009)	Social housing	Difference in IAQ equipped with and without mechanical balanced ventilation	18+6	Passive House (with MVHR) + Low Energy (without MEV)	Energy use, CO_2, T, RH, VOCs	2 years	Living room, bedroom
3	Sinfonia	[9]	Innsbruck	Austria	Existing Buildings	Social housing, with and without MVHR	Measure building performance after refurbishments	185	deep retrofit aiming enerPHit standard	Energy use, CO_2, T	>12 months	Central corridor or dwelling extract
4	Büren	[15,16]	Büren	Switzerland	New Construction (2017)	Homes have either a Zehnder balanced ventilation unit with an enthalpy exchanger or only manually operated windows	Type of ventilation used (windows only, or mechanical balanced ventilation)	4		CO_2	12 months	Living room, bedrooms
5	IECH	[8]	Odense	Denmark		Less than 2 % of the homes had mechanical ventilation, 70 % ventilated via windows, the rest had natural ventilation or hybrid ventilation implemented.	Impact of the IEQ on asthma and allergy among children	500		CO_2, Ta, RH	>48 h (Mar through May)	Bedroom (childrens bedroom)
6	ClimateReady (CliReady)	To be published	Different Locations (Pamplona and Seville)	Spain	Existing Buildings and new construction (2018)	Dwellings were mainly naturally ventilated	Assessment of adaptability in residential buildings to Climate Change	23	CTE	CO_2, Ta, RH	18 months (2 summers)	Living room, bedroom
7	CSIC	[29,28]	Madrid	Spain	Existing Buildings		Increase knowledge of IAQ in existing dwellings with natural ventilation systems	12		CO_2, Ta, RH	16 days (avg) repeated in different seasons throughout > 2 years	Living room, bedroom, kitchen, bathroom
8	ALLO	[54]	La Rochelle	France	Existing Buildings	Social housing	Study IEQ of the homes	6	RT2012	CO_2, Ta, RH, PM1 PM2.5 PM10	18 months	Living room
9	VentStd	[67]	Scotland	UK	New Construction		Investigate the effect the changes to Scottish Building Standards in 2015	16	BSD 2015 Ventilation Standard 3.14	CO_2, RH, Ta, pressure, tVOCs, PM1, PM2.5, PM10	6 months	Living room, Bedroom

(*continued on next page*)

Table 2 (*continued*)

ID	Project Name (Study acronym)	Reference	Location	Country	New Construction / Existing Homes	Miscellaneous	Objective	Number of Homes	Standard / Code	Variables / Contaminants Measured	Measurement duration (CO_2)	Rooms (CO_2)
10	AMR	[66]	Scotland	UK	New and Existing	Managed social housing of older adults (60+)	Examine the impact of building ventilation on the type and nature of microorganisms in homes	21		Ta, RH, CO2, VOCs, PM2.5	7 months	Living room, Bedrrom, Kitchen
11	VALIDate	To be published	Different Locations	Ireland	Newly constructed and retrofitted energy efficient dwellings,	60 mechanically ventilated, 23 naturally ventilated, and 4 with demand-control ventilation	Monitor IEQ in energy-efficient Irish homes, addressing gaps in long-term air pollutant data	87	National Buildings regulations and retrofit guidelines	Ta, RH, CO_2, VOCs, PM2.5, Radon	>18 months	Living rooms, bedrooms, kitchens, bathrooms
12	ALIVE	[4]	Different Locations	Ireland	New Construction		Performance of natural ventilation in maintaining acceptable IEQ in energy-efficient Irish dwellings	9	National Buildings regulations	PM2.5, CO_2, TVOCs, NO_2, CO, Ta, RH	12 months	Bedrooms, living rooms, kitchens
13	ARDEN	[12,32]	Different Locations	Ireland	Existing homes	Measured pre and post retrofits	Comparison of IAQ pre- and post- and energy retrofit	26	National Building Regulations	CO, CO_2, PM2.5, NO_2, Radon, Formaldehyde and BTEX, Ta, RH	48 h (pre and post retrofit)	Living rooms, bedrooms
14	HomeOffice_COVID19 (HO_COV)	[5]	Different Locations: Trondheim	Norway	New Construction and Existing homes	Natural and mechanically ventilated	Assess IAQ	21		Formaldehyde, TVOC, CO_2, PM1, PM2.5, RH, Ta	1–2 weeks (in winter and partly in summer)	Rooms used as home offices
15	NL_ML	[45]	Roosendaal	Netherlands	Existing homes		Assess whether the passive house energy requirements were met while achieving acceptable indoor climate conditions	10	Passive House	Ta, RH, CO_2, energy flows	>24 months	Living room, 2 bedrooms
16	HENGH	[68]	California	USA	New Construction and Existing homes (between 2011 and 2017)		Impacts of new home mechanical ventilation requirements included in the 2008 Title 24 building standard	70	California's Title 24	PM2.5 indoors and outdoors, formaldehyde, NO_2, and CO_2 indoors.	6–8 days	Cooktop, range hood and other exhaust fans, and the heating and cooling system.
17	LIA	[79]	California	USA	Existing homes (renovated 2013–2017)	Low income housing	Assess the inadequacy of kitchen ventilation systems in meeting California's Residential Building Code standards for controlling indoor air pollutants	23+71	California's Title 24	PM2.5 indoors and outdoors, formaldehyde, NO_2, and CO_2 indoors.	1 week	
18	MX1, SC1 and SF1	[49–51]	Mexico City, Dunfermline	International (Mexico, UK, USA)			Explores the suitability of the Passivhaus to	7	Local building regs and Passive	PM2.5, tVOC, CO_2, Ta, RH	6 – 12 months	Living room, kitchen, bedroom

(*continued on next page*)

Table 2 (*continued*)

ID	Project Name (Study acronym)	Reference	Location	Country	New Construction / Existing Homes	Miscellaneous	Objective	Number of Homes	Standard / Code	Variables / Contaminants Measured	Measurement duration (CO_2)	Rooms (CO_2)
			and San Francisco				provide and maintain high levels of IAQ		House on each case			

were calculated from a short time period only, e.g., when the measurements ended only a few hours into a specific month. Specifically for this study, all room-month tuples that were generated with less than 30 individual measurements were removed. Note that the studies included in this meta-analysis measured continuously for a minimum of 48 h per room (e.g., see IECH study). Others measured for a few weeks in each room, some for several years. So, the total number of rooms represented in the dataset was not diminished in this step. Only a few room-month tuples with little or no statistical power were removed.

Additionally, all room-month tuples where more than 1 % of the available measurement values was outside the defined lower and upper bound of its quality indicator, were removed. For CO_2 the lower and upper bound to identify implausible values were defined as 350 ppm and 10,000 ppm.

2.2.2.2. Filter. Depending on the investigated aspect, a filter was applied to select all relevant room-month tuples. E.g. filtering for CO_2 measurements left 32,046 room-month tuples, while filtering for CO_2 in bedrooms during nighttime hours left 4311 tuples. When applying an additional filter to only include colder months (Oct-May) 3244 tuples remained. Alternatively, a filter selecting only tuples with mean outdoor temperatures below 10 °C was also applied, resulting in 2386 entries (see results).

2.2.2.3. Aggregate months (to one tuple per room). Since the measurement duration varied strongly in the different IAQ studies (from two days to several years) and to avoid a bias towards studies with many room-month tuples (per room), the room-month tuples (after filtering) were aggregated to obtain one tuple per room, referred to "room tuple". It is representative for the entire period of interest depending on the chosen filter, e.g. entire year, cooler time of the year, etc. This way, each room can be treated equally in subsequent steps, independent of the measurement duration. As a consequence, the distribution stemming from different spatial samples (rooms), rather than temporal samples of the same entity are analyzed.

For the aggregation, the overall mean is calculated as weighted mean applying the fraction of the month where measurement data was actually collected as weighting factor (w_i). This factor ranges between 0 and 1 and represents the relative data availability when generating each room-month tuple. It was calculated using formula (1).

$$w_i = \frac{(N - NAs)}{nr_{days} \cdot nr_{hours} \cdot nr_{samples\ per\ hour}} \tag{1}$$

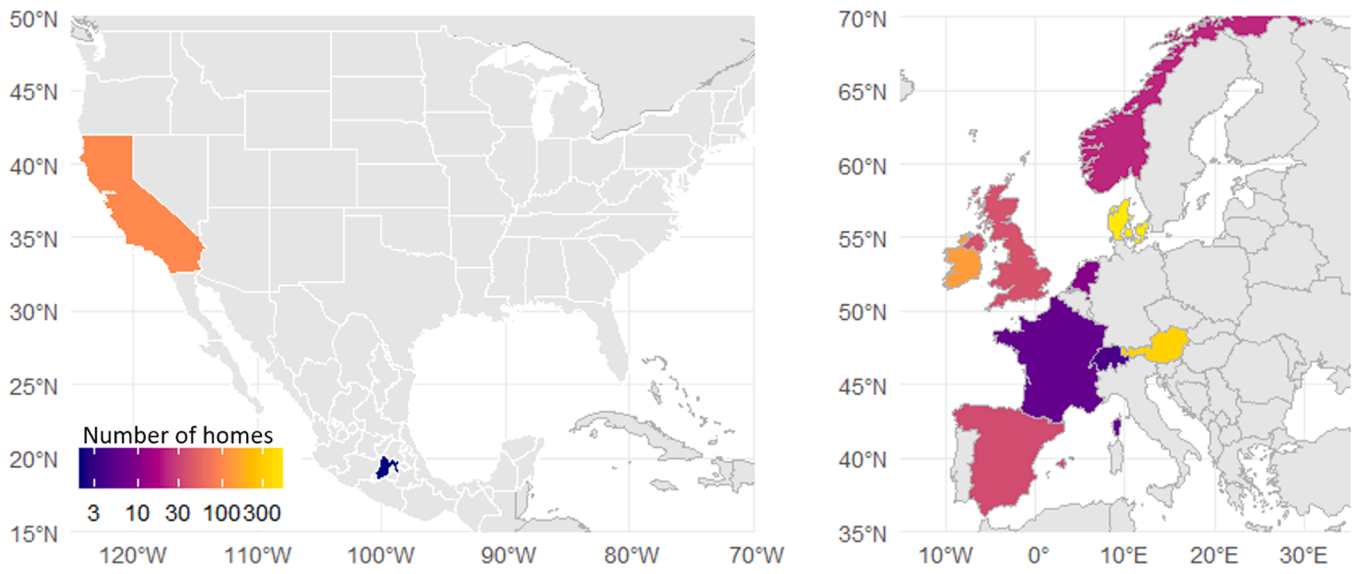
Here N stands for the total number of measurements, while NAs quantifies the number of measurements with missing or undefined values. Therefore, $(N - NAs)$ represents the number of valid measurement samples used to generate the room-month tuple, nr_{days} the number of days for given month, nr_{hours} the number of hours for the given time of day (i.e. 8, 16 or 24 h) and $nr_{samples\ per\ hour}$ the number of measurement samples per hour based on the median sampling interval.

The aggregated empirical cumulative distribution function for each room ($eCDF_{ag}$) is calculated based on the law of total probability as the finite mixture of the individual eCDFs of the room-month tuples ($eCDF_i$) that are being aggregated by determining the weighted arithmetic mean, according to formula (2).

$$eCDF_{ag} = \frac{\sum_{i=1}^{N} w_i \cdot eCDF_i}{\sum_{i=1}^{N} w_i} \tag{2}$$

The weighting ensures that room-month tuples that were generated from only a few measurements, e.g. only a few hours or days, count less than the room-month tuples (for the same room) with continuous measurements.

The effect of this aggregation step (compared to not performing an aggregation to room tuples) will depend on the composition of the



			Ventilation type fractions [%]				Building type fractions [%]			
Country	N	Study acronyms	Win	Nat	Hyb	Mech	SFH	TH	MFH	AB
AUT	332	Lod_Kuf, Lue3, Sinfonia	30	0	0	70	25	0	68	8
CHE	4	Büren	50	0	0	50	0	0	100	0
DNK	499	IECH	69	16	13	2	68	0	18	12
ESP	33	CISC, CliReady	36	30	12	21	0	0	73	27
FRA	6	ALLO	0	0	100	0	0	0	0	0
GBR	37	VentStd, AMR, SC1	0	22	0	78	0	24	0	76
IRL	129	VALIDate, ALIVE, ARDEN	10	21	5	64	74	26	0	1
MEX	2	MX1	0	50	50	0	0	0	0	100
NLD	10	NL_ML	0	0	0	100	0	100	0	0
NOR	21	HO_COV	0	52	24	24	33	24	10	33
USA	93	HENGH, LIA, SF1	0	0	0	100	75	0	0	25

N: Number of homes; Win: Window airing (not designed), Nat: Natural ventilation (designed), Hyb: Hybrid/mixed mode ventilation, Mech: Mechanical ventilation; SHF: single family house, TH: terraced house, MFH: multi family house, AB: apartment block

Fig. 4. Overview of location and composition of homes investigated in this meta-analysis.

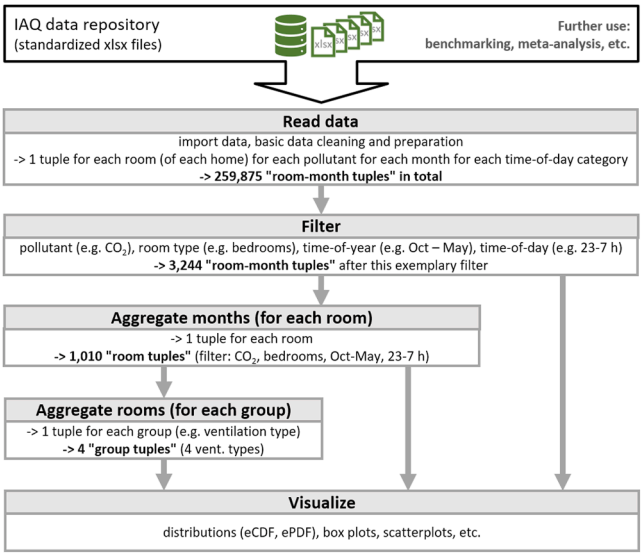


Fig. 5. Schematic illustration of workflow for meta-analysis.

investigated samples. When comparing the CO_2 concentration for different ventilation types, the relative difference between statistical values (mean, median, quantiles) of aggregated and unaggregated data was mostly below 5 %, except for the group "hybrid/mixed mode ventilation", where the deviations were around 10 %. Here, the dataset for this ventilation type included one study (IECH) with many bedrooms and three other studies (ClimateReady, MX1, VALIDate) with only a few bedrooms each, but measured over many months (see S2 in supplemental material). The aggregation had a larger effect when analyzing the differences between day- and nighttime hours, again due to large sample size of this study (IECH), see 3.1 and S2.

2.2.2.4. Aggregate rooms (to one tuple per ventilation type). For further analysis, e.g., for the comparison of different ventilation strategies, the obtained distributions for each room are aggregated over the groups of interest, e.g., the ventilation type. This aggregation step to obtain "group tuples" is performed in a similar manner as the aggregation to room tuples (described above), except that no weighting is applied to avoid the aforementioned bias towards studies with long measurement periods. Thus, every room tuple is treated equally representative for the investigated period, as the resulting eCDFs should represent the distributions encountered in investigated set of homes (with its characteristics) for the investigated time of year and time of day. Naturally, possible influencing factors need to be considered when doing a comparison to ensure that the resulting groups are comparable. Based on the results (see 3.2) and the available meta-information, two of the variables considered most relevant were accounted for during the generation of

the group tuples for ventilation type: climate via outdoor temperature and occupant density via building type.

Note that the distribution in group tuples represents the total measurement time from numerous investigated rooms, not a distribution for a room with a particular ventilation type. It should be interpreted as a measurement prevalence (i.e. amount of data points above or under a certain CO_2 concentration) in many bedrooms. In other words, room tuples represent the distribution of the contaminant in question for individual rooms, whereas group tuples represent the distribution over selected time and space. Therefore, while group tuples can be used for comparing the effectiveness of different ventilation concepts, the set of room tuples should be used for estimating contaminant exposure experienced by occupants.

2.2.2.5. Evaluation metric. The temporal mean might not be the most appropriate statistical metric to evaluate indoor CO_2 concentration. E.g. if CO_2 concentration is used to assess the ventilation rate, a metric representing the steady state concentration would be more favorable. Analyzing the time series data of four of the herein presented studies with and without mechanical ventilation (AMR, HENGH, Lodenareal and Lüftung 3.0) by applying peak-detection algorithms indicates that the best proxy for the nighttime peak, and therefore presumably also for the steady state concentration in bedrooms might be somewhere around the 88th and 94th percentile of the nighttime concentrations. More precisely, the root mean square error was minimized for the 92nd percentile when considering all nighttime measurements (23–7 h) from those four studies, see supplemental material S5. Nevertheless, as this result has not been validated and might not be transferable to other studies or other room types, the following analysis reports the 90th percentile unless otherwise noted. It is interpreted as a measure of the ventilation rate and therefore of the dilution of other indoor generated contaminants. However, since the mean (over occupied time periods) resembles the most direct measurement of the exposure to bio-effluents, selected figures showing the mean have also been generated and are available as supplemental material S6.

3. Results

3.1. Impact of measurement location, time and context

Fig. 6 shows the measured CO_2 concentration for the different room categories. It was generated after aggregation to one tuple per room. A total of around 16,000 room-month tuples were aggregated to form 1882 room tuples, i.e. 1882 distinct measurement locations (rooms + ambient locations). So, the box-plots depict the distribution of the 90th percentile of all distinct rooms of a given room category. The mean of the 90th percentiles of all room types are significantly different compared to the mean 90th percentile in ambient air, measured within the HENGH study (USA) (all $p < 0.001$). Looking at the bedrooms, the median and the mean (of the 90th percentiles) are around 1180 ppm and 1320 ppm while the upper quartile reaches almost 1600 ppm, notably more than in the other room categories shown in Fig. 6.

Bedrooms and living rooms make up more than 75 % of all available distinct measurement locations and they can be considered most

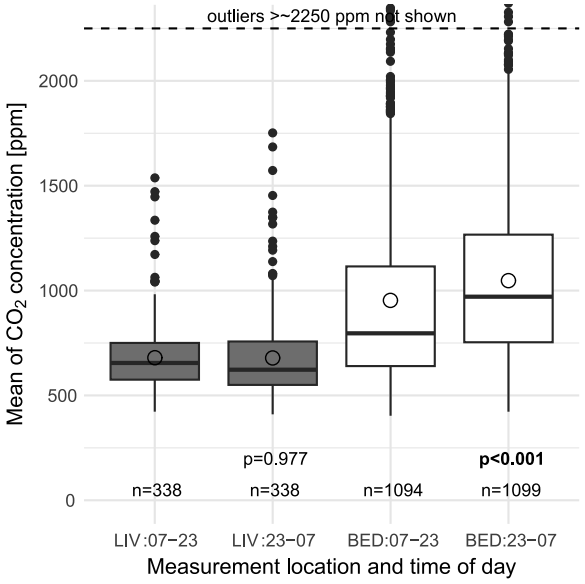


Fig. 7. Left: Boxplot of mean CO_2 concentration in bedrooms (BED) and living rooms (LIV) during day (07–23 h) and nighttime (23–07 h) hours. The mean is shown as a circle, n indicates the number of distinct rooms. The p-value of the regression analysis informs if the groups have a significantly different mean compared to the reference group (first).

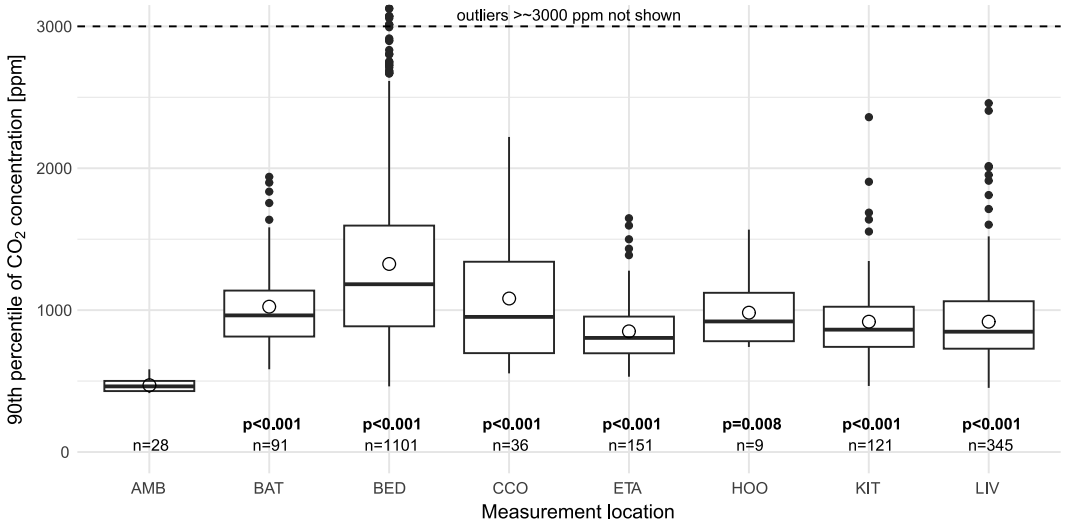


Fig. 6. Boxplots of the 90th percentile of CO_2 concentration in different measurement locations (AMB: ambient/outdoor, BAT: bathroom, BED: bedroom, CCO: central corridor, ETA: extract air, HOO: home-office room, KIT: kitchen, LIV: living room including undefined living area). The mean is shown as a circle, n indicates the number of distinct measurement locations, e.g. the number of measured bedrooms (1101) or living rooms (338). The p-value of the regression analysis informs if the mean is significantly different (if e.g. <0.05) from the reference location (ambient).

relevant in terms of occupant usage time. Fig. 7 evaluates the mean CO_2 concentration for those two room types during day and nighttime. Here the mean was chosen because the 90th percentiles depend much stronger on the chosen time periods. The mean concentration over all measured bedroom datapoints is around 100 ppm higher during nighttime compared to daytime, which is statistically significant ($p < 0.001$). Not so for the living room. Note that this difference between day and night in the bedrooms, doubles to 207 ppm when removing one particular study (IECH) which investigated 500 child bedrooms, comprising about 45 % of the total sample. In child bedrooms a different occupancy and door opening profile during the day can be assumed. A more detailed comparison between day and nighttime hours is provided in the supplemental materials S3.

Fig. 8 and Fig. 9 visualize the overall differences between the studies for bedrooms and living rooms. One can see a great variation between studies, even within the same ventilation type category, in particular for bedrooms. This could be due to differences in climatic conditions, occupant behavior or building characteristics stemming from general geographical or cultural differences and/or different study foci. For most cases, the homes selected cannot be considered representative for their entire country. As shown below (see 3.2), the study is a significant descriptor of the CO_2 concentration, this should be kept in mind when interpreting aggregated results.

To investigate the seasonal variation of indoor CO_2 concentrations, the difference between the monthly mean for each bedroom (i.e. from room-month tuples) and the overall mean in January was analyzed. It showed significantly lower CO_2 concentrations during the months June through October (see supplemental material S3). All homes of all studies were located in the northern hemisphere. It is most likely that the seasonal variation was due to changing airing behavior by occupants as the ambient (outdoor) temperature varied. To investigate the potential influence of ambient temperature on the CO_2 concentration in bedrooms, a linear regression was performed. The ambient temperature data was either measured locally by the individual IAQ studies (LIA, Lodenareal, HENGH, MX1, VentStd), or gathered from the respective national weather agencies or a global weather service provider [77] for the respective locations. Differences between locally measured temperature data and data from weather agencies / service providers were investigated for seven cities. For most cases / homes, a good agreement (<0.5 °C) was observed. For a few homes (from two studies) the differences were around 1–2 °C, in rare occasions up to 3.5 °C (see S4 in supplemental material).

Fig. 10 shows the relationship between indoor CO_2 concentration and ambient temperature, differentiating between dwellings with mechanical and natural ventilation (including dedicated natural, window and hybrid ventilation). Each point represents the monthly 90th percentile measured in bedrooms during nighttime hours (23–7) and the monthly mean ambient temperature for the respective location. Although the data scatter broadly, driven by other influencing factors, a statistically significant trend towards lower CO_2 concentrations with increasing temperature, especially above 10 °C, can be observed. The Generalized Additive Model (GAM), a model commonly used for scatterplot smoothing [33], indicated a linear decline in CO_2 concentration for ambient temperatures above roughly 5 to 10 °C and more or less constant values for lower temperatures (very similar when applying a LOESS smoother, not shown in the figure).

Since the LOESS and GAM smoothers exhibited a linear decline above a certain temperature, a segmented linear regression was performed [52] (Table 3. It identified a breakpoint at 5 °C for mechanically and 12 °C for naturally ventilated homes. The small in-/decline in CO_2 concentration below this breakpoint was not significant ($p = 0.8$ and $p = 0.2$) and the adjusted R^2-value indicates that these linear models can only explain around 8 % and 5 % of the variation for mechanically and naturally ventilated bedrooms, respectively. Nevertheless, despite a big variability due to other influencing factors, the change in slope is statistically significant (see p-value in Table 3) for temperatures above the breakpoint for both ventilation types, resulting in a decline in the 90th percentile of CO_2 concentration of 22 ppm and 57 ppm per degree Celsius for mechanically and naturally ventilated bedrooms, respectively. During the cooler time of the year the naturally ventilated bedrooms had on average a roughly 400 ppm higher 90th percentile than mechanically ventilated bedrooms. For temperatures above roughly 12 °C this difference is gradually reduced. To reduce a potential bias due to different outdoor temperatures, the following analysis was performed for tuples where the ambient temperature was ≤ 10 °C. This filter was chosen for all ventilation types, despite the fact that the segmented linear regression gave different breakpoints, to simplify the filtering procedure while avoiding a large reduction in sample size.

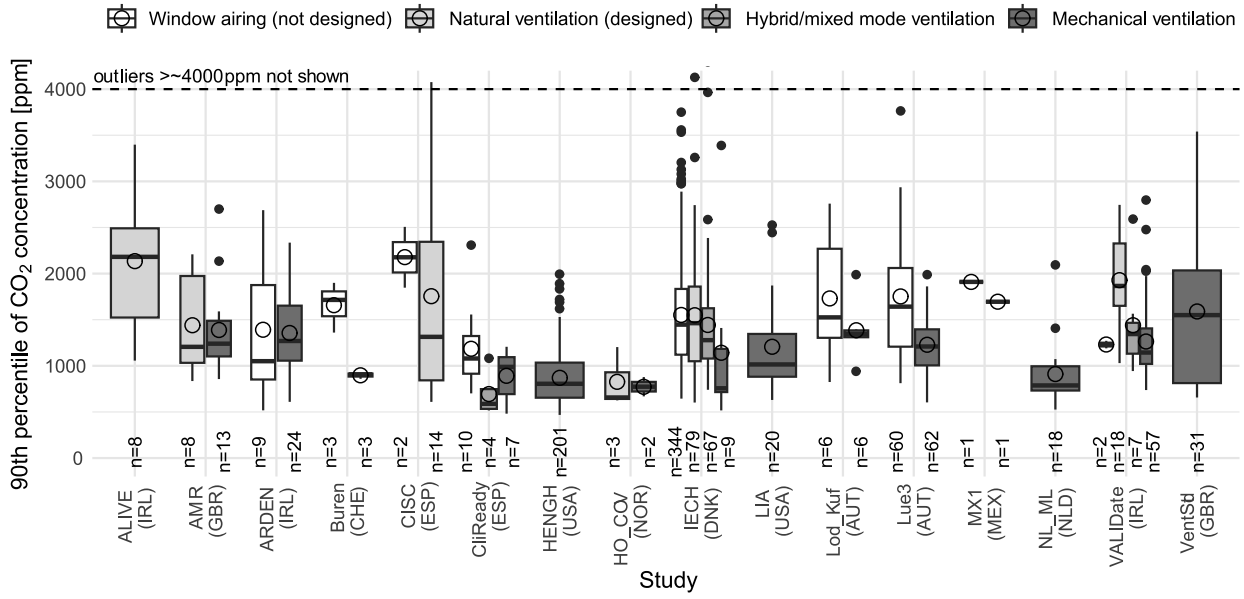


Fig. 8. Boxplots of the 90th percentile of the CO_2 concentration in bedrooms measured in the different studies (see Table 2) during nighttime hours (23–07). The mean is shown as a circle, n indicates the number of distinct rooms.

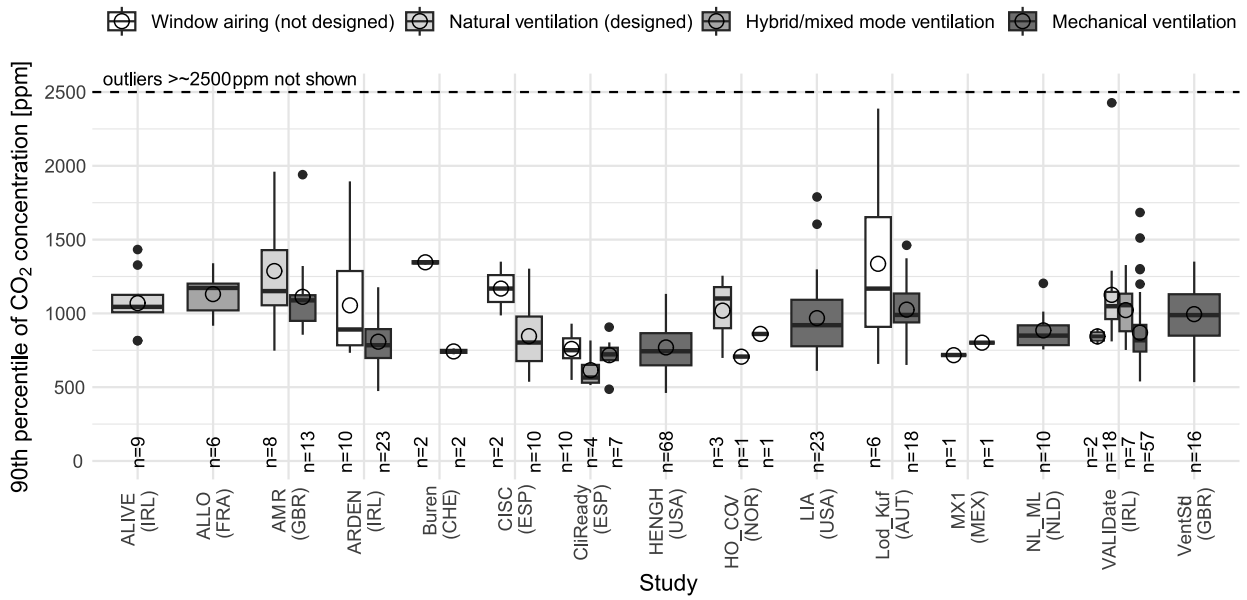


Fig. 9. Boxplots of the 90th percentile of the CO_2 concentration in living rooms measured in the different studies (Table 2) during daytime hours (07–23). The mean is shown as a circle, n indicates the number of distinct rooms.

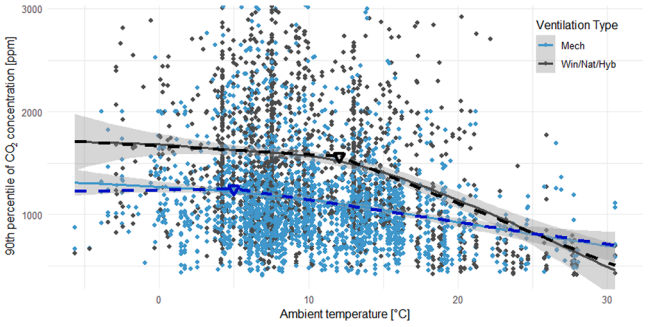


Fig. 10. Scatterplot of 90th percentile of CO_2 concentration in bedrooms during nighttime hours as a function the monthly mean ambient temperature for mechanically and naturally ventilated homes (the latter combining window airing, dedicated natural ventilation and hybrid solutions). The fitted curves from a Generalized Additive Model (GAM) are also shown (solid lines plus grey areas for standard error). Additionally, the results of segmented linear regressions with breakpoints at around 5 °C (Mech) and 12 °C (Win/Nat/Hyb) are also shown (dashed lines). Note that points above 3000 ppm are not shown but included in the regression.

Table 3
Results of segmented linear regression investigating relation between 90th percentile of nighttime CO_2 concentration in bedrooms and mean outdoor temperature.

Ventilation type	Breakpoint (BP) [°C]	Intercept @ 0 °C [ppm]	Slope before BP [ppm/ °C]	Change in Slope after BP [ppm/ °C]
Mechanical	5.0 ± 1.7	1236 ± 35 ($p < 0.001$)	2.5 ± 10.5 ($p = 0.814$)	−24.0 ± 10.7 ($p = 0.026$)
Window airing, natural or hybrid	12.0 ± 1.2	1665 ± 49 ($p < 0.001$)	−8.3 ± 6.8 ($p = 0.228$)	−49.1 ± 9.5 ($p < 0.001$)

3.2. Multiple linear regression

To investigate possible relationships between relevant building and study characteristics, a stepwise model selection procedure to fit a multiple linear regression model was performed [71]. Since only categorical variables were included it can be considered an "Analysis of Variance" (ANOVA). Starting from the most trivial model (intercept only) the algorithm tested if adding (and in subsequent steps, removing) one of the parameters listed in Table 4 reduced the Akaike Information Criteria (AIC). This is repeated to the point where no further reduction of the AIC is possible. Interaction terms (between the chosen parameters) were also allowed and tested, e.g. to account for differences in concentration for certain ventilation types given a certain building type. Unfortunately, other obviously relevant parameters like year of construction, occupant number, room size, etc. were not provided in a sufficiently consistent manner to investigate a further expansion of the linear regression model.

For the 90th percentile of the nighttime CO_2 concentration in bedrooms for months with ambient temperature below 10 °C the model selection using AIC chose the following model (Eq`. (3)):

$$c_{p90} = \beta_0 + \beta_1 X_{vent\ type\ red} + \beta_2 X_{study} + \beta_3 X_{bld\ type\ red} \quad (3)$$

The coefficients and p-values are given in supplemental material S7. The intercept value (for reference case "Mech"-"ALIVE"-"MFH/AB") is 1846 ± 230 ppm. After accounting for study site and building type, buildings with natural (Win/Nat/Hyb) ventilation tend to have higher c_{p90} values (by 463 ± 72 ppm with $p < 0.001$) compared to a mechanical ventilation. Most study datasets (AMR, Buren, CliReady, HENGH, HO_COV, IECH, LIA, Lue3, NL_ML) have significantly lower c_{p90} values

Table 4
List of parameters and their possible values tested within stepwise model selection procedure.

Parameter	Description	Possible parameter values (number = degree of freedom+1)
$X_{vent\ type\ red}$	Ventilation type -reduced	Mech, Win/Nat/Hyb (2)
$X_{vent\ type}$	Ventilation type - all	Window airing (not designed), Natural ventilation (designed), Hybrid/mixed mode ventilation, Mechanical ventilation (4)
$X_{country}$	Country of home	AUT, CHE, DNK, ESP, GBR, IRL, NLD, NOR, USA (9)
X_{study}	Name of study	ALIVE, AMR, … (15)
$X_{bld\ type\ red}$	Building type - reduced	SFH/TH, MFH/AB (2)
X_{bld_type}	Building type - all	SFH, TH, MFH, AB (4)

compared the reference dataset (with $p < 0.05$). Single-family and terraced houses (SFH/TH) show 135 ± 55 ppm lower c_{p90} values than the reference building type (MFH/AB) with $p = 0.014$. This model fits the data significantly overall ($p < 0.001$), but it explains only about 15 % of the variation (R^2_{adj}=0.146), meaning other important factors likely exist.

Performing the stepwise parameter selection for the living room data, identified only ventilation type ($X_{vent\ type}$) and country ($X_{country}$) as variables that improve the AIC. However, from the three alternative ventilation options (with window airing being the reference case), only mechanical ventilation has a coefficient that is significantly different from zero (-238 ± 70 ppm). Again, the model is significant but can only explain about 20 % of the variation (R^2_{adj}=0.22).

3.3. Comparing ventilation concepts

In the following section the impact of the ventilation concept on the CO_2 concentration, in particular in the bedrooms, is further highlighted. Given the results presented above, the comparison is made for measurements during cooler periods where the monthly mean ambient temperature did not surpass 10 °C. By applying this filter, the influence of ambient temperature on occupant ventilation behavior can be considered small. Since the building type was chosen in the model selection procedure as relevant parameter, results are distinguished between two categories, one combining single family houses (SFH) and terraced houses (TH), and the other combining multi-family houses (MFH) and apartment blocks (AB).

The regression analysis of the 90th percentile in bedrooms shows that, with window airing considered the reference situation, only mechanically ventilated bedrooms have a significantly lower value ($p < 0.001$) for both building type categories (see Fig. 11). With mean c_{p90} values around 1200 to 1300 ppm, it is about 300 to 400 ppm lower than ventilation options relying on natural forces. Natural and hybrid ventilation show no significant difference compared to window airing (p-values between 0.21 and 0.63). The same tendencies can be observed when comparing temporal mean values of the CO_2 concentration in bedrooms, see supplemental material S7.

The empirical cumulative distribution function (eCDF) for CO_2 is available for each room and month. Fig. 12 and Fig. 13 show the eCDFs for CO_2 in single rooms. The shaded areas show the eCDF for the 5th, 25th, 75th and 95th percentile of the bedrooms. Note that the shading boundaries (and the dashed line) do not necessarily depict a concrete room, but a fictive room where the respective percentage of other rooms have a "lower" eCDF curve. The data distribution for all bedrooms is depicted with solid line. It represents the distribution of all observations in all bedrooms. Note that e.g. only a few rooms could be responsible for defining the highest (or the lowest) percentile points.

The eCDF for SFH and TH during nighttime hours in months with mean ambient temperature below 10 °C (Fig. 12) shows, that in 67 % and 20 % of the time the median bedroom had CO_2 concentrations above 1000 ppm in naturally and mechanically ventilated homes, respectively. One can also see that the worst 5 % of bedrooms experienced a CO_2 concentration above 2000 ppm for around 54 % and 18 % of the night hours, respectively. In MFHs and ABs the concentrations tend to be slightly higher (see Fig. 13), with concentrations at median level surpassing 1000 ppm in 77 % and 33 % of the nighttime hours in naturally and mechanically ventilated homes, respectively. The eCDFs of mechanically ventilated homes vary less compared to naturally ventilated homes, i.e. the width of the shaded areas (Fig. 13) is notably smaller. One can also see that the "worst" 25 % of mechanically ventilated homes seem to be performing unsatisfactory presenting a substantial fraction of nighttime hours above 1500 ppm (see light grey shaded area to the right of median in Fig. 13).

4. Discussion

4.1. The need for harmonized IAQ data

The presented work addresses the lack of databases collecting measured indoor contaminant concentrations in a harmonized manner. The developed methodology for gathering and structuring empirical data on IAQ and its meta-information allows comparisons, benchmarking and meta-analysis across studies, countries, building types and other parameters. It can provide the research community with a wide and diverse dataset, far beyond the individual research projects to continue to support research in fields where in-situ data is crucial, e.g. the assessment of IAQ management solutions [31], the evaluation of health impacts on population scale [47,48], or the evaluation of the effectiveness of certain policy interventions [10]. As the dataset grows it will support harm-based analysis, reflecting such new compliance

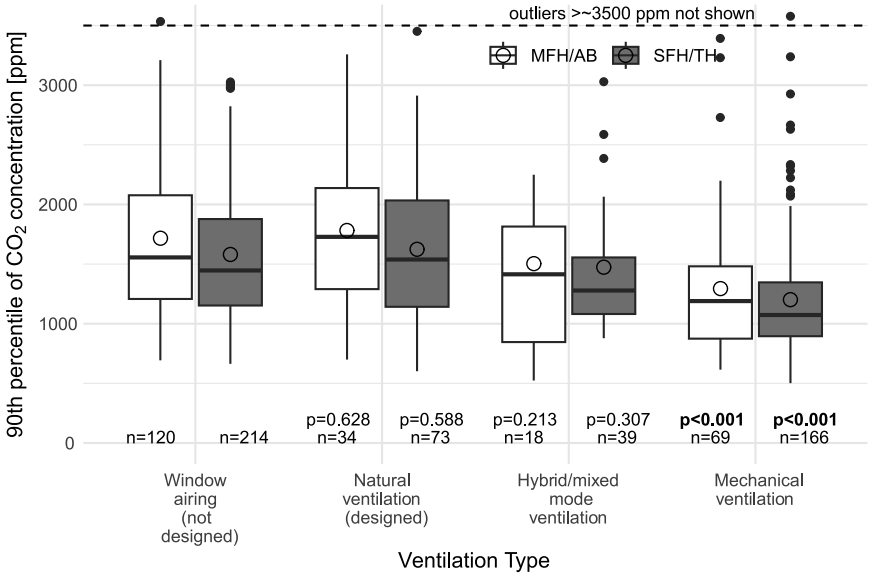


Fig. 11. Boxplots of 90th percentile of CO_2 concentration in bedrooms at nighttime for different ventilation and building types. The mean is shown as a circle, n indicates the total number of distinct rooms. The p-value of the regression analysis informs if the ventilation types have a significantly different mean compared to the reference type (Window airing).

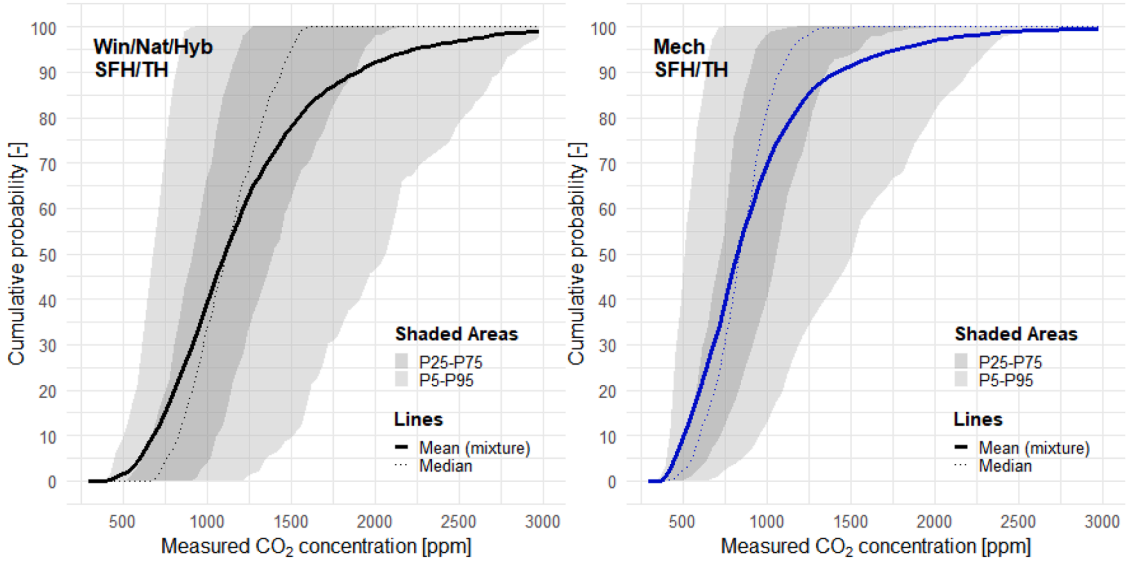


Fig. 12. eCDF plots for CO_2 in bedrooms in single-family and terraced houses (SFH/TH), during nighttime hours and months with mean ambient temperature ≤10 °C) for (left) window aired, naturally and hybrid ventilation homes (Win/Nat/Hyb, $n = 326$) and (right) mechanically ventilated homes (Mech, $n = 166$). The eCDFs representing the 5th, 25th, 75th and 95th percentile are shown as shaded areas, the median as dotted line. The eCDF representing the mixture distribution (representing the distribution of all measurements) is shown as solid line.

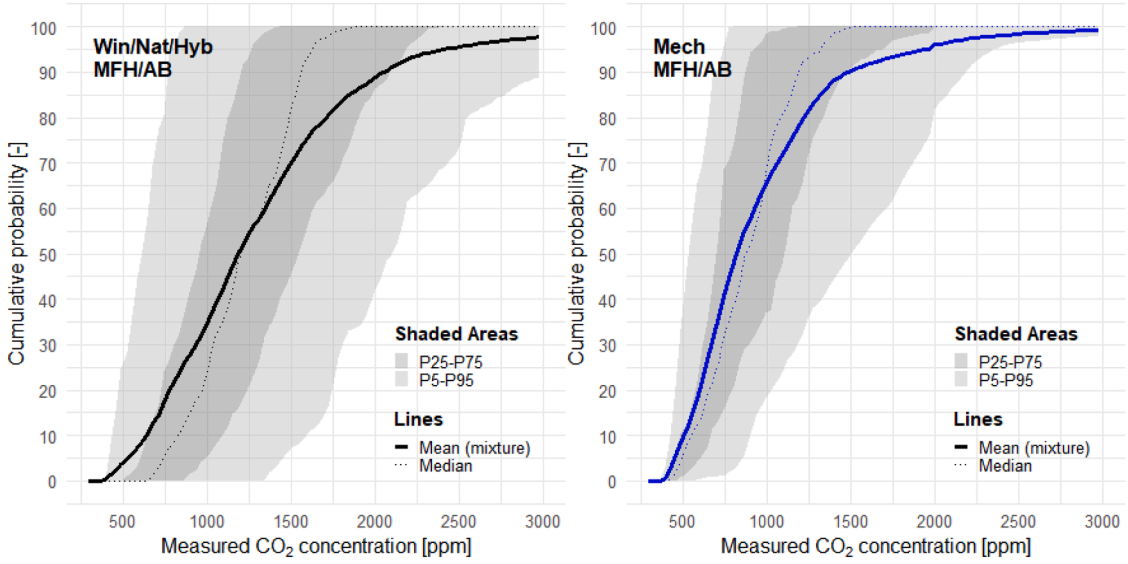


Fig. 13. eCDF plots for CO_2 in bedrooms in multi-family houses and apartment blocks (MFH/AB), during nighttime hours and months with mean ambient temperature ≤10 °C) for (left) window aired, naturally and hybrid ventilation homes (Win/Nat/Hyb, $n = 172$) and (right) mechanically ventilated homes (Mech, $n = 69$). The eCDFs representing the 5th, 25th, 75th and 95th percentile are shown as shaded areas, the median as dotted line. The eCDF representing the mixture distribution (representing the distribution of all measurements) is shown as solid line.

pathways as already included in ASHRAE 62.2.

4.2. Proposing a methodology and presenting an open-source implementation

The proposed harmonization ensures comparability and compatibility when combining different datasets. IAQ measurement studies often apply different statistical indicators making the subsequent quantitative meta-analysis difficult, an issue frequently reported in traditional review studies, e.g. [27,37,61]. The proposed approach also avoids potential data protection issues when analyzing time series data from IAQ measurements, in particular with respect to the EU's GDPR, [6,42,69]. The original time series data does not need to be collected centrally, only monthly aggregated data is generated by the algorithm. Nevertheless, the aggregated output contains empirical cumulative distribution functions (eCDF) enabling detailed statistical analysis that goes beyond the reporting of scalar aggregates (e.g. mean or median) by providing information about the temporal and the spatial (i.e. between rooms) distribution. In other words, it can answer questions on how the contaminant concentration varies over time within individual rooms as well as on how this temporal distribution varies between homes (e.g. see Fig. 12 and Fig. 13).

The herein developed open-source algorithm, implemented as an easy to deploy R package [70], was used to produce a harmonized datasets from IAQ measurements from over 1000 dwellings from 18 IAQ studies. This preliminary database has been published in an open data

repository [64]. Although the data is mainly confined to Europe and California, it provides a wealth of information and can be expanded to include more datasets including other countries, climates, contaminants and/or building types. This development will be linked to the global Indoor Air Quality Database (IAQ-DB) initiative [20], which is developing a database with an interactive web-based front end. The proposed procedure will allow the generation of harmonized datasets from IAQ measurements for the inclusion in data repositories and databases, such as the IAQ-DB.

4.3. Showcasing the potential of harmonized IAQ datasets for CO_2

The herein presented meta-analysis showcases the potential use of such a database. It does so by analyzing the measured indoor CO_2 concentration while focusing on bedrooms. While CO_2 is a widely used IAQ metric, one should be aware of its limitations [56,57]. In an ongoing work and as the dataset grows, the harmonized database will be used to analyze other, more health-related IAQ contaminants, such as $PM_{2.5}$ or certain VOCs. Nevertheless, CO_2 can be used for estimation of bio-effluent exposure and is considered a good metric for ventilation rate assessment. For the latter the steady state concentration would be most indicative, but it cannot be measured exactly in residential settings. Therefore, the 90th percentile of the measured CO_2 concentration is used as a proxy in the presented regression analysis.

The CO_2 concentration measurements show a great variability, which cannot be predicted applying regression models with satisfactory accuracy with the available (fully populated) set of variables (room type, ventilation type, building type, country, ambient temperature, study). Other relevant features such as occupant number, ventilation rate, room size, building age, etc. would have been only available for part of the dataset and were excluded in the presented regression results. Nevertheless, significant differences in concentration could be identified in various regressions. E.g. Fig. 6 and Fig. 7 confirm that nighttime hours in bedrooms are the most critical in terms of underventilation and exposure to bio-effluents. No significant difference in mean CO_2 concentration between daytime and nighttime hours could be observed in living rooms. A stepwise model selection procedure identified ventilation type, building type and study as relevant model parameters and found CO_2 concentration in bedrooms with mechanical ventilation to be significant lower compared to the other three ventilation types (window, natural, hybrid). The regression against building type reflects possible differences in occupant density, with single family and terraced houses having lower mean concentrations compared to multi-family houses and apartment blocks.

Analyzing the 90th percentile of the CO_2 concentration as a function of ambient temperature (Fig. 10) confirmed that mechanically ventilated bedrooms had, on average, a significantly lower CO_2 concentration, in particular during the colder months. This advantage disappears slowly for monthly mean outdoor temperatures between roughly 12 °C and 25 °C. It is assumed that this decline in CO_2 concentration is due to changes in occupant behavior with respect to opening windows and doors as the outdoor temperature rises. It seems plausible that this effect is more pronounced in homes where ventilation relies on occupant intervention. The results of the segmented linear regression (Table 3) may be used in future work to empirically model outdoor-temperature-driven airing behavior in building energy simulations. Overall, the regression results suggest that, when comparing IAQ across different ventilation strategies and studies, it is advisable to control for room type, building type (or occupant density, if available), and outdoor temperature (or season).

The large variation in CO_2 concentration suggests that additional explanatory variables can probably improve the statistical model to predict CO_2 concentrations in residential homes, e.g. occupant behavior and building characteristics. Unfortunately, they were not available for a sufficiently large sample size within this study. Further analysis of the data and the meta-information could provide insight into the extent to which ventilation, building and usage characteristics affect indoor air quality [8]. These analyses could also include the use of artificial intelligence and machine learning tools, the credibility of which increases with the availability of more datasets on which to perform analyses, ultimately providing input to models at the individual level [7]. In connection with the latter, with the wider distribution of low-cost sensors, we see that data on IAQ is becoming available more easily. The quality of these sensors often does not compare to research grade alternatives, but developments suggest that they are improving continuously [81].

The great advantage of the proposed data harmonization algorithm is, that the distribution of contaminant concentration may also be investigated. As an example, the empirical cumulative distribution functions (eCDFs) of the measured CO_2 concentration in bedrooms are shown (Fig. 12 and Fig. 13). If the median eCDF is considered as representative of the central tendency across rooms, the mechanically ventilated bedrooms in SFH and TH surpassed 1000 ppm during 20 % of the nighttime hours, i.e. for around 1.5 h if equidistant sampling interval is assumed. In contrast, the representative naturally ventilated bedroom (in SFH/TH) experienced concentrations >1000 ppm for 67 % of the nighttime, i.e. for more than 5 h. Note that a recent study recommends even lower CO_2 concentrations for increased sleep quality [38]. These figures also visualize the spread between different homes/bedrooms of the same category. It becomes clear that while mechanically ventilated bedrooms generally outperform naturally ventilated bedrooms in terms of CO_2 concentration (and ventilation rate), there is a non-negligible fraction of mechanically ventilated bedrooms that seem to be under-ventilated, see eCDFs representing the 95th percentile in Fig. 12 and Fig. 13. These outcomes generally align with those of other field or review studies, e.g. [11,41,59,61,80].

4.4. Limitations

One motivation for this work is to enable harm-based IAQ assessment by characterizing population-level distributions of concentrations across a heterogeneous housing stock. This relies on the assumption that sufficiently large and diverse samples can provide meaningful estimates of central tendency and variability for defined populations, rather than deterministic predictions for individual homes. In this respect, the following limitations should be acknowledged when interpreting the presented results.

- CO_2 as an IAQ metric: In the concentration range typically encountered in buildings, CO_2 itself is not a health-harming pollutant but an air contaminant used as proxy for ventilation adequacy and other contaminants related to occupants' emissions. Nevertheless, given sufficient data and meta-information such a database can also be applied to analyze other more health relevant pollutants such as $PM_{2.5}$ or certain VOC's to characterize their respective distributions in a very similar manner. Naturally, the meta-analysis would need to be adapted to control for other or additional relevant parameters, e.g. outdoor concentrations or construction materials.
- Lack of contextual / meta-information: Also, for the analysis of CO_2 concentration it would have been interesting to control for other influencing parameters describing the housing stock, e.g. building age or building energy standard. Unfortunately, such meta-information was not (yet) available for a sufficient large sample size.
- Limited geographical / climatic coverage: Great efforts have been made within the international collaborative project [35] to collect as many geographically diverse datasets as possible. The resulting database includes many homes from various European countries as well as California, covering a good range of different climatic conditions and building practices. Nevertheless, the presented data still lacks sufficient samples from many parts of the world, e.g. from developing countries, hot and humid climates, etc.

- Sample bias: Data from existing IAQ studies from within the project consortium was collected analyzed based on availability. By no means does the dataset form a perfectly representative sample of residential homes in certain countries, climates, buildings, etc.

5. Conclusions

This work demonstrates a practical and novel pathway to harmonize IAQ measurement data across studies, supporting robust comparison, benchmarking, and meta-analysis. The aggregation methodology preserves statistical richness via empirical distribution functions while avoiding centralized raw time-series collection. It has been implemented as an open-source algorithm in R and it was applied to create an open-access harmonized dataset integrating measurement data from 18 IAQ studies, 11 countries and 1166 residential homes — primarily in Europe and California.

The herein presented meta-analysis exemplifies the application of a harmonized IAQ dataset by analyzing CO_2 concentrations with a focus on bedrooms. It confirms that mechanically ventilated bedrooms exhibit significantly lower concentrations—particularly in colder months—while large variability across homes highlights the influence of occupant behavior, building characteristics, and season. Distributional analyses reveal both, typical benefits of mechanical systems, but also a non-negligible fraction of under-ventilated bedrooms, underscoring the need for improved controls and commissioning. The regressions indicate that, besides ventilation type and outdoor temperature, contextual information on building type and country can serve as predictors. While the regression results are constrained by the available meta variables and geographical coverage, the framework is well-suited to expand regression models as more meta-data becomes available and/or to include more health-relevant contaminants (e.g., $PM_{2.5}$, VOCs) in the analysis.

Together with initiatives such as IAQ-DB, it is intended to support the progressive expansion of harmonized IAQ datasets enabling the characterization of population-level distributions of indoor air contaminant concentrations for harm-based IAQ assessment.

CRediT authorship contribution statement

Gabriel Rojas: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Reto Stauffer:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Núria Casquero-Modrego:** Writing – review & editing, Writing – original draft, Data curation. **Marcel Loomans:** Writing – review & editing, Writing – original draft, Data curation. **Marc Abadie:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Ibrahim Alhindawi:** Writing – review & editing, Data curation. **Francesco Babich:** Writing – review & editing, Methodology, Conceptualization. **Gabriel Bekö:** Writing – review & editing, Data curation. **Marie Coggins:** Writing – review & editing, Writing – original draft, Data curation. **Bart Cremers:** Writing – review & editing, Data curation. **Bernhard Damberger:** Writing – review & editing, Data curation. **Timm Freundorfer:** Writing – review & editing, Visualization, Software, Investigation. **Sonia Garcia-Ortega:** Writing – review & editing, Data curation. **Hala Hassan:** Writing – review & editing, Data curation. **Benjamin Jones:** Writing – review & editing, Methodology, Conceptualization. **Maria Justo Alonso:** Writing – review & editing, Data curation. **Irene Lara-Ibeas:** Writing – review & editing, Methodology, Conceptualization. **Grainne McGill:** Writing – review & editing, Data curation. **James McGrath:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Aurora Monge-Barrio:** Writing – review & editing, Writing – original draft, Data curation. **Constanza Molina:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Alejandro Moreno-Rangel:** Writing – review & editing, Data curation. **Peter Tappler:** Writing – review & editing, Data curation. **Linda Toledo:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Gabriel Rojas reports financial support was provided by Austrian Research Promotion Agency. Marcel Loomans reports financial support was provided by Rijksdienst voor Ondernemend Nederland. Bart Cremers reports a relationship with Zehnder Group International that includes: employment. One co-author employed by manufacturer of mechanical ventilation systems If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was carried out to a great extend as part of the IEA research cooperation on behalf of the Austrian Federal Ministry for Climate Protection, Environment, Energy, Mobility, Innovation, and Technology. Additional support was provided by Rijksdienst voor Ondernemend Nederland (EGOI123012).

The authors would also like to acknowledge all the original research projects and their financial support providing the wealth on data for this work, in particular:

- The financial support provided by the Scottish Building Standards Division, Directorate for Local Government and Housing (UK) - research to identify if changes to guidance in standard 3.14 (Ventilation) in 2015 have been effective in improving ventilation and indoor air quality.

- CLIMAREADY “Adaptation Assessment of Spanish residential buildings” (2020–2023), funded by Convocatoria $I + D + I$ “Retos de Colaboración" del Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad. Plan Estatal de Investigación Científica y de Innovación”. Reference: PID2019-109008RB-C21.

- The ARDEN study was funded by the Sustainable Energy Authority of Ireland (SEAI) - 2018 National Energy Research, Development & Demonstration Funding Programme through research Grant No. RDD204.

- The study in Denmark was funded by the Villum Kann Rasmussen Foundation (now Villum Foundation).

- The Research Council of Norway (project no. 257660) and several partners through the Research Centre on Zero Emission Neighborhoods in Smart Cities (FME ZEN).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.buildenv.2026.114297.

Appendix A

Table A.1

Table A.1
Meta-information fields with predefined entries.

Ventilation type
Window airing (not designed)
Natural ventilation (designed)
Hybrid/mixed mode ventilation
Mechanical ventilation
Method of vent. rate determination
Active tracer gas measurement (in context of study)
Pressure-drop compensated Flow-hood measurement (in context of study)
Flow-hood / anemometer measurement (in context of study)
Measurement not in context of study (e.g. from ventilation system commissioning)
Design values (from ventilation system design)
Calculated (e.g. from CO2, passive tracer gas)
Other (comment reqd)
Airtightness normalization value
Building volume specific [1/h]
Envelope area specific / [m^3/hm^2]
Floor area specific / [m^3/hm^2]
Type of building
Single family house (SFH)
Terraced house (TH)
Multi-family house (MFH)
Apartment block (AB)
Undefined/Not known
Type of room / measurement location
Ambient
Bedroom
Living room
Bathroom / Toilet
Kitchen
Living area undefined
Central corridor
Home office
Exhaust air (vent. Sys.)
Outdoor air (vent. Sys.)
Extract air (vent. Sys.)
Supply air (vent. Sys.)
Other
Type of occupancy (in room)
adult
child
elderly
mixed
unknown
Fresh air supply (in room)
Through occupant operated window only
Through automatic operated window
Through permanent opening to the exterior (grille, …)
Through mechanically driven supply air (no recirculated air)
Through mechanically driven overflow (or recirculated air) from adjacent zones
Other

Data availability

https://zenodo.org/records/14917724 (The data has been published in Zenodo repository:)

References

[1] M. Abadie, E. Geffre, C. Picard, et al., PANDORA: an open-access database of indoor pollutant emission rates for IAQ modeling, J. Build. Eng. (2025) 114216, https://doi.org/10.1016/j.jobe.2025.114216. September 29.

[2] M.O. Abadie, P. Blondeau, PANDORA database: a compilation of indoor air pollutant emissions, HVACR. Res. 17 (4) (2011) 602–613, https://doi.org/10.1080/10789669.2011.579877.

[3] M. Akimoto, X. Fan, L. Lan, et al., New research on bedroom ventilation and sleep quality suggests that building standards should Be revisited (ASHRAE 1837-RP), Sci. Technol. Built. Environ. (2025) 1–12, https://doi.org/10.1080/23744731.2025.2531317. July 21.

[4] I. Alhindawi, J.A. McGrath, D. Sood, J. O'Donnell, M.A. Byrne, A seasonal assessment of indoor air quality and thermal performance in naturally ventilated airtight energy-efficient dwellings, Build. Environ. 276 (May) (2025) 112862, https://doi.org/10.1016/j.buildenv.2025.112862.

[5] M. Alonso, T.N. Moazami, P. Liu, R.B. Jørgensen, H.M. Mathisen, Assessing the indoor air quality and their predictor variable in 21 home offices during the Covid-19 pandemic in Norway, Build. Environ. 225 (November) (2022) 109580, https://doi.org/10.1016/j.buildenv.2022.109580.

[6] F. Armknecht, Z. Benenson, P. Morgner, C. Müller, C. Riess, Privacy implications of room climate data, J. Comput. Secur. 27 (1) (2019) 113–136, https://doi.org/10.3233/JCS-181133.

[7] F.T. Bahadur, S.R. Shah, R.R.R. Nidamanuri, A brief outline of indoor air quality: monitoring, modeling, and impacts, J. Environ. Eng. 151 (5) (2025) 03125001, https://doi.org/10.1061/JOEEDU.EEENG-7929.

[8] G. Bekö, T. Lund, F. Nors, J. Toftum, G. Clausen, Ventilation rates in the bedrooms of 500 Danish children, Build. Environ. 45 (10) (2010) 2289–2295, https://doi.org/10.1016/j.buildenv.2010.04.014.

[9] J. Bliem, Masterthesis, University of Innsbruck, Austria, 2021.

[10] BUILD UP, Acceptable Indoor Environmental Quality and Energy Efficiency Standards: Can Buildings Meet Both (and for Everyone)? BUILD UP, 2025. https://build-up.ec.europa.eu/en/resources-and-tools/articles/acceptable-indoor-environmental-quality-and-energy-efficiency.

[11] N. Canha, C. Teixeira, M. Figueira, C. Correia, How is indoor air quality during sleep? a review of field studies, Atmosphere 12 (1) (2021) 110, https://doi.org/10.3390/atmos12010110.

[12] A.M. Coggins, N. Wemken, A.K. Mishra, et al., Indoor air quality, thermal comfort and ventilation in deep energy retrofitted Irish dwellings, Build. Environ. 219 (July) (2022) 109236, https://doi.org/10.1016/j.buildenv.2022.109236.

[13] S. Colclough, R.O. Hegarty, M. Murray, et al., Post occupancy evaluation of 12 retrofit nZEB dwellings: the impact of occupants and high in-use interior temperatures on the predictive accuracy of the nZEB energy standard, Energy Build. 254 (January) (2022) 111563, https://doi.org/10.1016/j.enbuild.2021.111563.

[14] M. Collins, S. Dempsey, Residential energy efficiency retrofits: potential unintended consequences, J. Environ. Plan Manag. 62 (12) (2019) 2010–2025, https://doi.org/10.1080/09640568.2018.1509788.

[15] B. Cremers, The resulting CO2 levels and the heating/cooling consumption of apartments with balanced ventilation versus window ventilation, in: Proceedings of Healthy Buildings Conference, Oslo, 2021, pp. 162–171.

[16] B. Cremers, Comparison of balanced ventilation and window ventilation in terms of CO2 levels and space heating and cooling use, in: *41st AIVC - ASHRAE IAQ Joint Conference*, Athens, 2022.

[17] R.J D. Dear, G. Brager, Developing an adaptive model of thermal comfort and preference, ASHRAe Trans. 104 (1) (1998).

[18] B. Du, M.C. Tandoc, M.L. Mack, J.A. Siegel, Indoor CO2 concentrations and cognitive function: a critical review, Indoor. Air. 30 (6) (2020) 1067–1082, https://doi.org/10.1111/ina.12706.

[19] L. Du, V. Leivo, T. Prasauskas, M. Täubel, D. Martuzevicius, U. Haverinen-Shaughnessy, Effects of energy retrofits on indoor air quality in multifamily buildings, Indoor. Air. 29 (4) (2019) 686–697, https://doi.org/10.1111/ina.12555.

[20] L. Dusan and B. Du. 2025. "iAQ DB - the indoor air quality database." https://iaqdb-dev.epfl.ch/.

[21] M. Elsayed, P. Romagnoni, S. Pelsmakers, R. Castaño-Rosa, U. Klammsteiner, The actual performance of retrofitted residential apartments: post-occupancy evaluation study in Italy, Build. Res. Inf. 51 (4) (2023) 411–429, https://doi.org/10.1080/09613218.2022.2121908.

[22] (EU) 2024/1275, *Directive (EU) 2024/1275 of the European parliament and of the council of 24 April 2024 on the energy performance of buildings (recast) (text with EEA relevance)*. 2024/1275, Off. J. Eur. Union (2024). http://data.europa.eu/eli/dir/2024/1275/oj.

[23] X. Fan, C. Liao, K. Matsuo, et al., A single-blind field intervention study of whether increased bedroom ventilation improves sleep quality, Sci. Total. Environ. 884 (August) (2023) 163805, https://doi.org/10.1016/j.scitotenv.2023.163805.

[24] W.J. Fisk, P. Wargocki, X. Zhang, Do indoor CO2 levels directly affect perceived air quality, health, or work performance? ASHRAe J. (September) (2019). https://www.nxtbook.com/nxtbooks/ashrae/ashraejournal_201909/index.php#/p/70.

[25] V. Foldvary, G. Beko, S. Langer, K. Arrhenius, D. Petras, Effect of energy renovation on indoor air quality in multifamily residential buildings in Slovakia, Build. Environ. 122 (September) (2017) 363–372, https://doi.org/10.1016/j.buildenv.2017.06.009.

[26] V. Földváry, T. Cheung, H. Zhang, et al., Development of the ASHRAE global thermal comfort database II, Build. Environ. 142 (September) (2018) 502–512, https://doi.org/10.1016/j.buildenv.2018.06.022.

[27] A. Fonseca, I. Abreu, M.J. Guerreiro, N. Barros, Indoor air quality in healthcare units—A systematic literature review focusing recent research, Sustainability 14 (2) (2022) 967, https://doi.org/10.3390/su14020967.

[28] G. Ortega, S. María. 2024. "Caracterización experimental y modelo predictivo de la calidad del aire interior en viviendas con sistemas de ventilación natural." Phd, E. T.S. Arquitectura (UPM). East=-3.6981904; north=40.4253562; name=Madrid, Madrid, Comunidad de Madrid, España. https://oa.upm.es/80533/.

[29] S. Garcia-Ortega, P. Linares-Alemparte, Indoor air quality in naturally ventilated dwellings in Spain, Informes de La Construcción 75 (572) (2023), https://doi.org/10.3989/ic.6447 e519–e519.

[30] GDPR. 2016. "General Data Protection Regulation." https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

[31] G. Guyot, M.H. Sherman, I.S. Walker, Smart ventilation energy and indoor air quality performance in residential buildings: a review, Energy Build. 165 (April) (2018) 416–430, https://doi.org/10.1016/j.enbuild.2017.12.051.

[32] H. Hassan, A. Kumar Mishra, N. Wemken, et al., Deep energy renovations' Impact on indoor air quality and thermal comfort of residential dwellings in Ireland – ARDEN Project, Build. Environ. 259 (July) (2024) 111637, https://doi.org/10.1016/j.buildenv.2024.111637.

[33] T.J. Hastie, R.J. Tibshirani, Generalized additive models, in: Monographs On Statistics and Applied Probability, 43, Chapman and Hall, 1991.

[34] C. Howard-Reed, B. Polidoro, Database Tools for Modeling Emissions and Control of Air Pollutants from Consumer Products, Cooking, and Combustion, National Institute of Standards and Technology, 2006, https://doi.org/10.6028/nist.ir.7364.

[35] IEA EBC Annex 86. 2025. "IEA EBC Annex 86 - energy efficient indoor air quality management in residential buildings." https://annex86.iea-ebc.org/.

[36] International Energy Agency (IEA), Buildings, International Energy Agency, 2022. https://www.iea.org/energy-system/buildings.

[37] C. Kakoulli, A. Kyriacou, M.P. Michaelides, A review of field measurement studies on thermal comfort, indoor air quality and virus risk, Atmosphere 13 (2) (2022) 191, https://doi.org/10.3390/atmos13020191.

[38] M. Kang, Y. Yan, C. Guo, et al., Ventilation causing an average CO2 concentration of 1000 ppm negatively affects sleep: a field-lab study on healthy young people, Build. Environ. 249 (February) (2024) 111118, https://doi.org/10.1016/j.buildenv.2023.111118.

[39] N.E. Klepeis, W.C. Nelson, W.R. Ott, et al., The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants, J. Expo. Anal. Environ. Epidemiol. 11 (3) (2001) 3, https://doi.org/10.1038/sj.jea.7500165.

[40] S. Langer, G. Bekö, E. Bloom, A. Widheden, L. Ekberg, Indoor air quality in passive and conventional new houses in Sweden, in: *Building and Environment*, Special Issue: Indoor pollutants, chemistry and health- Selected papers presented at Indoor Air 2014 conference in Hong Kong 93, 2015, pp. 92–100, https://doi.org/10.1016/j.buildenv.2015.02.004.

[41] J. Laverge, M. Delghust, A. Janssens, Carbon dioxide concentrations and humidity levels measured in Belgian standard and low energy dwellings with common ventilation strategies, Int. J. Vent. 14 (2) (2015) 165–180, https://doi.org/10.1080/14733315.2015.11684078.

[42] N. Leesakul, C. Morisset, Position paper: the role of law in achieving privacy and security measures in smart buildings from the GDPR context, in: 2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), 2023, pp. 619–626, https://doi.org/10.1109/EuroSPW59978.2023.00073. July.

[43] J.M. Logue, T.E. McKone, M.H. Sherman, B.C. Singer, Hazard assessment of chemical air contaminants measured in residences, Indoor. Air. 21 (2) (2011) 92–109, https://doi.org/10.1111/j.1600-0668.2010.00683.x.

[44] J.M. Logue, P.N. Price, M.H. Sherman, B.C. Singer, A method to estimate the chronic health impact of air pollutants in U.S. Residences, Environ. Health Perspect. 120 (2012) 216–222, https://doi.org/10.1289/ehp.1104035.

[45] M. Loomans, G. Boxem, *Rapportage Monitoring Kroeven 2013-2014*. Nos. BWK2015-1580648, Eindhoven University of Technology, The Netherlands, 2015.

[46] S.D. Lowther, S. Dimitroulopoulou, K. Foxall, et al., Low level carbon dioxide indoors—A pollution indicator or a pollutant? A health-based perspective, Environments. 8 (11) (2021) 125, https://doi.org/10.3390/environments8110125.

[47] G. Morantes, B. Jones, C. Molina, M.H. Sherman, Harm from residential indoor air contaminants, Environ. Sci. Technol. 58 (1) (2024) 242–257, https://doi.org/10.1021/acs.est.3c07374.

[48] G. Morantes, B. Jones, M. Sherman, C. Molina, A preliminary assessment of the health impacts of indoor air contaminants determined using the DALY metric, Int. J. Vent. 22 (4) (2023) 307–316, https://doi.org/10.1080/14733315.2023.2198800.

[49] A. Moreno-Rangel, Glasgow School of Art Dissertations, School of Art, 2019.

[50] A. Moreno-Rangel, F. Musau, T. Sharpe, G. McGill, Indoor air quality assessment of Latin America's first passivhaus home, Atmosphere 12 (11) (2021) 1477, https://doi.org/10.3390/atmos12111477.

[51] A. Moreno-Rangel, T. Sharpe, G. McGill, F. Musau, Indoor air quality and thermal environment assessment of Scottish homes with different building fabrics, Buildings 13 (6) (2023) 1518, https://doi.org/10.3390/buildings13061518.

[52] V.M.R. Muggeo, Segmented: an R package to fit regression models with broken-line relationships, R. Jounal. 8 (1) (2008) 20–25.

[53] W.W. Nazaroff, Residential air-change rates: a critical review, Indoor. Air. 31 (2) (2021) 282–313, https://doi.org/10.1111/ina.12785.

[54] J. Nicolle, M. Abadie, F. Castet, *Données Individuelles de Qualité d'air et Changement de Comportement, ALLO - Accompagner Les Habitants Pour L'intégration De La qualité De L'air Dans Leur Logement*, ADEME, France, 2021.

[55] L. Pampuri, P. Caputo, C. Valsangiacomo, Effects of buildings' Refurbishment on indoor air quality. Results of a wide survey on radon concentrations before and after energy retrofit interventions, Sustain. Cities. Soc. 42 (October) (2018) 100–106, https://doi.org/10.1016/j.scs.2018.07.007.

[56] A. Persily, Development and application of an indoor carbon dioxide metric, Indoor. Air. 32 (7) (2022) e13059, https://doi.org/10.1111/ina.13059.

[57] A. Persily, W. Bahnfleth, H. Kipen, et al., ASHRAE Position Document On Indoor Carbon Dioxide, ASHRAE, 2022. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934476.

[58] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2024 released, https://www.r-project.org/.

[59] F.L. Rashid, M.A. Al-Obaidi, N.M.L. Al Maimuri, et al., Mechanical ventilation strategies in buildings: a comprehensive review of climate management, indoor air quality, and energy efficiency, Buildings 15 (14) (2025) 2579, https://doi.org/10.3390/buildings15142579.

[60] G. Rojas, Dissertation, University of Innsbruck, 2015.

[61] G. Rojas, M. Fletcher, D. Johnston, M. Siddall, A review of the indoor air quality in residential passive house dwellings, Energy Build. 306 (March) (2024) 113883, https://doi.org/10.1016/j.enbuild.2023.113883.

[62] G., Rojas, and T. Freundorfer. 2025. *IEA-EBC-Annex86/ST2analysis*. R. IEA-EBC-Annex86, released March 4. https://github.com/IEA-EBC-Annex86/ST2analysis.

[63] G., Rojas, and L. Mayr. 2025. "Registry of indoor air quality (IAQ) measurement studies in residential homes - summary tables from IEA EBC Annex 86 (ST2)." IEA EBC Annex 86, May 15. https://doi.org/10.5281/zenodo.15399188.

[64] G., Rojas, and R. Stauffer. 2025. "Indoor air quality in residential homes - monthly aggregated data from IEA EBC Annex 86 (ST2) project participants." Version 1.0.0. IEA EBC Annex 86, February 24. https://doi.org/10.5281/zenodo.14917724.

[65] G. Rojas, W. Wagner, J. Suschek-Berger, R. Pfluger, W. Feist, Applying the passive house concept to a social housing project in Austria – Evaluation of the indoor environment based on long-term measurements and user surveys, Adv. Build. Energy Res. 10 (1) (2016) 125–148, https://doi.org/10.1080/17512549.2015.1040072.

[66] T. Sharpe, G. McGill, S.J. Dancer, M.F. King, L. Fletcher, C.J. Noakes, Influence of ventilation use and occupant behaviour on surface microorganisms in contemporary social housing, Sci. Rep. 10 (1) (2020) 11841, https://doi.org/10.1038/s41598-020-68809-2.

[67] T. Sharpe, G. McGill, S. Howieson, et al., Research to Identify If Changes to Guidance in Standard 3.14 Ventilation in 2015 Have Been Effective in Improving

Ventilation and Indoor Air Quality, Building Standards Division Directorate for Local Government and Housing, Scotland, 2023. https://www.gov.scot/publications/research-identify-changes-guidance-standard-3-14-ventilation-2015-effective-improving-ventilation-indoor-air-quality/.
[68] B.C. Singer, W.R. Chan, Y. Kim, F.J. Offermann, I.S. Walker, Indoor air quality in California homes with code-required mechanical ventilation, Indoor. Air. 30 (5) (2020) 885–899, https://doi.org/10.1111/ina.12676.
[69] K. Singh, S. Gujar, S. Chaudhari, P. Kumaraguru, Exposing privacy risks in indoor air pollution monitoring systems, E3S Web. Conf. 620 (2025) 01001, https://doi.org/10.1051/e3sconf/202562001001.
[70] R., Stauffer, 2025. *Annex: IEA EBC Annex86 data analysis package*. R. V. 0.2-14. IEA-EBC-Annex86, released May 15. https://github.com/IEA-EBC-Annex86/annex.
[71] *Step Function - Stats Package*. n.d. R. V. 3.6.2. RDocumentation. Accessed November 16, 2025. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/step.
[72] J. Taylor, R. McLeod, G. Petrou, et al., Ten questions concerning residential overheating in central and Northern Europe, Build. Environ. 234 (April) (2023) 110154, https://doi.org/10.1016/j.buildenv.2023.110154.
[73] L. Toledo, A. Wright, P. Cropper, Case study investigation of overheating in low-energy homes: insights from a post-occupancy evaluation in England, Build. Res. Inf. 53 (1–2) (2025) 139–164, https://doi.org/10.1080/09613218.2024.2417274.
[74] O. Toyinbo, L. Hägerhed, S. Dimitroulopoulou, et al., Open database for international and national Indoor Environmental Quality guidelines, Indoor. Air. 32 (4) (2022), https://doi.org/10.1111/ina.13028.
[75] S. Vardoulakis, E. Giagloglou, S. Steinle, et al., Indoor exposure to selected air pollutants in the home environment: a systematic review, Int. J. Environ. Res. Public Health 17 (23) (2020) 23, https://doi.org/10.3390/ijerph17238972.
[76] P. Wallner, U. Munoz, P. Tappler, et al., Indoor environmental quality in mechanically ventilated, energy-efficient buildings vs. Conventional buildings, Int. J. Env. Res. Public Health 12 (11) (2015) 14132–14147, https://doi.org/10.3390/ijerph121114132.
[77] "Weather Underground." n.d. Accessed August 29, 2025. https://www.wunderground.com/.
[78] Y. Yan, L. Lan, M. Kang, et al., Emission rate of carbon dioxide by older adults while sleeping, Build. Environ. 236 (May) (2023) 110299, https://doi.org/10.1016/j.buildenv.2023.110299.
[79] H. Zhao, W.R. Chan, S. Cohn, W.W. Delp, I.S. Walker, B.C. Singer, Indoor air quality in new and renovated low-income apartments with mechanical ventilation and natural gas cooking in California, Indoor. Air. 31 (3) (2021) 717–729, https://doi.org/10.1111/ina.12764.
[80] Y. Zhao, H. Sun, D. Tu, Effect of mechanical ventilation and natural ventilation on indoor climates in Urumqi residential buildings, Build. Environ. 144 (October) (2018) 108–118, https://doi.org/10.1016/j.buildenv.2018.08.021.
[81] H. Zheng, V. Krishnan, S. Walker, M. Loomans, W. Zeiler, Laboratory evaluation of low-cost air quality monitors and single sensors for monitoring typical indoor emission events in Dutch daycare centers, Environ. Int. 166 (August) (2022) 107372, https://doi.org/10.1016/j.envint.2022.107372.