

Unveiling Cleaning Workers' Chemical Exposure through Data-Driven Field Monitoring and Activity-Based Analysis

Claudia Falzone^{1,2*}, Nathalie Redon¹, Hatim Moujahid¹, Frédéric Thevenet¹, Marie Verrièle¹

¹ IMT Nord Europe (Centre for Energy and Environment), Lille, France

² Liege University (Sensing of Atmospheres and Monitoring), Arlon, Belgium

*Corresponding email: marie.verrièle@imt-nord-europe.

SUMMARY

The EXAMEN project investigates how cleaning practices in schools affect workers' exposure to indoor air pollutants using personal low-cost sensors and activity diaries. Sensor data (CO₂, PM_{2.5}, NO₂, O₃, TVOC, microclimate) and tablet-based logs are pre-processed in R, addressing missing and aberrant values, time alignment, and encoding errors. Sankey diagrams, time series, and boxplots reveal practice-specific peaks and room effects, while emphasizing sensor limitations and the need for informed IAQ data science.

KEYWORDS

Occupational exposure; Cleaning practices, Personal monitoring; Sensor-based assessment; Data integration

1 INTRODUCTION

The EXAMEN project investigates the exposure of cleaning staff and school occupants to indoor air pollutants generated by cleaning activities. Among other aims, it seeks to link professional practices to pollutant concentrations in order to identify realistic levers for reducing exposure. To this end, the project combines data from portable air quality monitors with activity logs completed by cleaning staff. To capture both environmental conditions and work organization, the study integrates personal monitoring with self-reported activity data. Portable sensing systems (OEEIL), developed at IMT Nord Europe, were worn at torso level by 20 female cleaning workers employed in nursery and primary schools in the city of Nancy (France). These devices were used to track exposure to pollutants of interest across the different microenvironments frequented during working hours, including CO₂, VOCs, particulate matter, O₃, and NO₂. In parallel, participants used tablets to record each activity (location, task, product, and practice) over four consecutive working days. This campaign highlighted major challenges related to data processing and synchronization between sensor measurements and activity logs. The absence of readily available analytical packages places researchers in front of a significant barrier to fine-grained interpretation of exposure data. This work therefore proposes a dedicated framework for graphical data representation, enabling integrated analysis and supporting robust conclusions.

2 METHODS

R was selected as the core environment to structure, clean, and analyze these datasets, allowing continuous time series to be combined with detailed activity logs (R Core Team 2025). Processing included data import, harmonization of date–time formats, detection and treatment of missing and aberrant values, and correction of encoding errors in the activity logs. Instrumental and tablet data were then merged on a temporal basis, despite different time resolutions (≈ 30 -s logging for sensors versus start/end times and durations for activities). To support interpretation, graphical exploration was systematically used, with plots tailored to specific questions (e.g. activities vs. pollutant levels, time spent on each activity–practice–product combination).

3 RESULTS AND DISCUSSION

For each employee, several types of plots were produced depending on the analysis objective. Sankey diagrams illustrate the links and durations between activities, practices, and product use (Figure 1a), highlighting that activity alone is insufficient to characterize exposure and must be considered alongside practices and product use. Indeed, similar activities may involve different operational modes, leading to contrasting exposure conditions, which emphasizes the need for detailed activity recording. Time series plots identify activities associated with the highest emission peaks, while boxplots highlight rooms where practices have a greater impact, likely due to contextual factors such as ventilation or room volume (Figure 1b). In particular, toilet cleaning is associated with higher exposure levels, likely due to lower ventilation, specific product use, and potentially higher breathing rates linked to the physical nature of the task. The use of R made it possible to process a large number of datasets and standardize processing by pollutant. This made it possible to integrate or adjust the code for uncertainties related to human factors as well as sensor performance in order to avoid premature or overinterpreted conclusions.

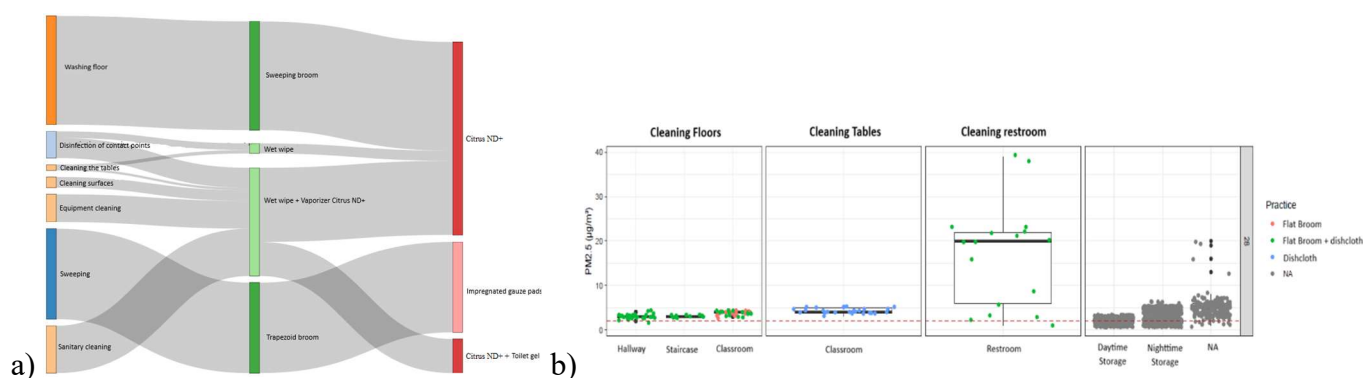


Figure 1. a) (from left to right) Links between activity, practice, and products used. The thickness of the links is proportional to the amount of time spent on the combinations, b) Distribution of PM_{2.5} for a working day broken down by activity/location and differentiated by practice.

4 CONCLUSIONS

Indoor air quality data scientists must be able to perform data processing with a thorough understanding of the instrumentation used, the design and execution of the monitoring campaign, and the level of engagement of participating workers. They also need strong command of analytical tools and a critical sense of when these tools are appropriate, given the study context and the objectives defined prior to data collection.

ACKNOWLEDGEMENT

This work was supported by ANSES through the EXAMEN project. The authors gratefully acknowledge the City of Nancy for providing access to its staff and the University of Liège–Arlon, specifically Claudia Falzone and Anne-Claude Romain for the collaborative work on data processing using R.

5 REFERENCES

- Karagulian, Federico, Maurizio Barbieri, Alexander Kotsev, Laurent Spinelle, Michel Gerboles, Friedrich Lagler, Nathalie Redon, Sabine Crunaire, et Annette Borowiak. 2019. « Review of the Performance of Low-Cost Sensors for Air Quality Monitoring ». *Atmosphere* 10(9):506. doi:10.3390/atmos10090506.
- R Core Team. 2025. « R: A Language and Environment for Statistical Computing ».