Data Analysis of Indoor Air Quality and Thermal Comfort in Dwellings in Santiago, Chile

Constanza Molina*[1], Benjamin Jones[2], Ignacio Garrido[1], Giobertti Morantes[2]

1 Pontificia Universidad Católica de Chile
Vicuña Mackenna 4860, Macul, Santiago
Chile
**Corresponding author: cdmolina@uc.cl*

2 Department of Architecture and Built Environment,
University of Nottingham
Nottingham, NG7 2UH, UK.

ABSTRACT

Achieving better energy efficiency requires dwellings to face a delicate equilibrium, balancing thermal comfort and indoor air quality. This longitudinal study uses crowdsourced data collected over a year from 15 residences in Santiago, Chile, to examine the intricate relationship between these two parameters and the houses' typology.
Results highlight considerable variability in $PM_{2.5}$ and PM_{10} concentrations and thermal comfort across the sample. PM concentrations are below the worldwide representative value, but the maximum values are above the representative maximum. Chronic harm from exposure to these concentrations is 1271 and 683 (DALYs/10^5 person/year) for $PM_{2.5}$ and PM_{10}. Moreover, the annual WHO 2021 recommendations are not met during the measured time, and the daily mean is met by 25% and 72% of the measured days for $PM_{2.5}$ and PM_{10}, respectively. Determinants of these variations may include geographical location and construction materials, which will be included in future research. The indoor environment does not provide the hygrothermal conditions to achieve acceptable thermal comfort, which is only reached during 56% of the measured time.
This research advocates for a comprehensive regulatory approach, ensuring that interventions are needed to optimize energy efficiency and prioritize occupant well-being. Insights from this study contribute to a better understanding of competing objectives in residential architecture, offering informed perspectives for strategic decision-making and impactful interventions.

KEYWORDS

Indoor air quality; Thermal comfort; Particles; Longitudinal study; Residential

1 INTRODUCTION

Residential buildings play a crucial role in shaping human health and well-being by serving as the primary environments where individuals spend a significant proportion of their time. Comfortable and healthy homes contribute to physical health by providing optimal conditions that reduce the risk of respiratory illnesses and support mental well-being by creating spaces of shelter and relaxation, fostering emotional stability, and reducing stress. However, indoor environments and housing conditions vary from house to house due to a complex interplay of factors, such as household lifestyle, preferences and behavior, and socio-demographic and economic inequalities, all of which influence human health and well-being within residential settings. Furthermore, environmental factors such as climate, ambient pollution, and extreme weather events demand a more resilient housing design to mitigate these risks and safeguard occupants' safety and comfort.

Worldwide, diverse geographical, cultural, and socio-economic factors contribute to a wide range of housing conditions and lifestyles. From urban cities to rural villages along the country, housing types vary significantly, ranging from modern high-rise apartments to traditional single-family dwellings and informal settlements. Over the last two decades, Chile has begun to improve the efficiency and sustainability of its housing stock. The increasing demand for energy-efficient buildings has led to tighter envelopes with a corresponding reduction in ventilation and air infiltration. An unintended consequence of the drive for heating and cooling demand reduction is a reduction in IAQ (Shrubsole et al., 2014; Molina et al., 2021). Therefore, if the Chilean housing stock aims to develop sustainably, it must simultaneously consider energy and IAQ targets, and the impacts of energy-demand reduction interventions on IAQ must be understood. Accordingly, it must promote occupant health and become people-centered.

This paper reports the beginning of a study that investigates particle concentrations and hygrothermal conditions in a sample of Santiago houses to inform public health interventions, urban planning, and sustainability in Chile. The study is currently measuring indoor environmental parameters in a sample of houses in Santiago. Section 2 describes the study and the methods of analysis carried out for this paper, and Section 3 shows and discusses the results.

2 METHODS

The research study measures IEQ parameters in a sample of 32 houses in Santiago using PurpleAir sensors, including relative humidity, air temperature, and particulate matter concentrations ($PM_{2.5}$ and PM_{10}). Measurements are made every two minutes and recorded in the cloud. Sensors were installed in the living room or kitchen, away from direct solar radiation or sources of contamination, as these locations are where the concentration of the more harmful pollutants are expected to be. Data was collected over a year to represent the four seasons and then extracted from the platform at a ten-minute resolution (averaged). However, not all sensors recorded sufficient data to run the analyses, so a cleaning process was carried out, and only those sensors with adequate data (15) were retained for the study. The sample includes both houses and flats, all of which utilize natural ventilation exclusively. The climate in Santiago can be described as *Csb*, meaning *Temperate/DrySummer/Warm summer/ Rain in winter*, according to the Köppen classification (Peel *et al.*, 2007).

2.1 Data gathering

Web scraping techniques from publicly available sources included wildfires and treated them as *events*. These sources of information included CONAF, the Chilean National Forestry Corporation, and social media platforms. Monitored data was extracted using an automated code in Python and the "BeautifulSoup4" and "Requests" libraries, known for being effective at conducting clean Web-Scraping processes. Fifteen out of the 32 dwellings with complete data over a year were selected for this analysis to ensure coverage across different seasons. Parameters included the relative humidity, air temperature, atmospheric pressure, and particle concentrations.

2.2 Data processing and data analyses

Indoor thermal comfort and air quality analyses were carried out on two levels: i) inferential and descriptive statistics (Casella *et al.*, 2001) and ii) for thermal comfort, a hygrothermal analysis based on a psychrometric chart with standardized comfort zones and, for IAQ, the PM concentration exceeding the WHO recommendations for 24h mean averages (WHO, n.d.), and health impacts using Harm Intensities according to Morantes *et al.*, (2023). Morantes *et al.* developed an expression for quantifying chronic harm in DALYs based on harm intensities and contaminant concentrations. They combined toxicological and epidemiological data to calculate median harm intensities and uncertainties for the most common and reported indoor air contaminants.

A second analysis was carried out using the "k-means clustering" algorithm. This technique organizes the data into groups based on similarities. We used the "*scikit-learn*" library (Kramer *et al.,* 2016), which provides machine learning tools for data processing to create user-type profiles by identifying patterns in thermal comfort time-series data. This clustering considers the median relative humidity and air temperature in homes to provide a representative value for categorizing them.

After determining that the cleaned temperature data followed a normal distribution— minimizing errors and outliers— we used the *elbow method* to determine the number of clusters or subgroups (Shi *et al.*, 2021).

2.3 Data Visualization

A psychrometric chart visually represented thermal comfort, indicating whether the temperature and relative humidity fell within the acceptable comfort zones for the summer and winter seasons. A provisional dataset was also created with the dates of major extreme events (such as fires). This was generated using CONAF Chile data until July 2023 and the social network X (formerly known as Twitter) with data after July 2023. This information was used to find patterns in PM concentrations and increases in temperature in homes when these extreme events occurred.

3 RESULTS AND DISCUSSION

3.1 Data visualization

Data visualization and analysis tools are hosted online (GitHub repository link: https://eccuc.github.io), providing ongoing updates and interactive visualizations of the study's findings. Figure 1 shows the location of the 15 sensors in Santiago de Chile.

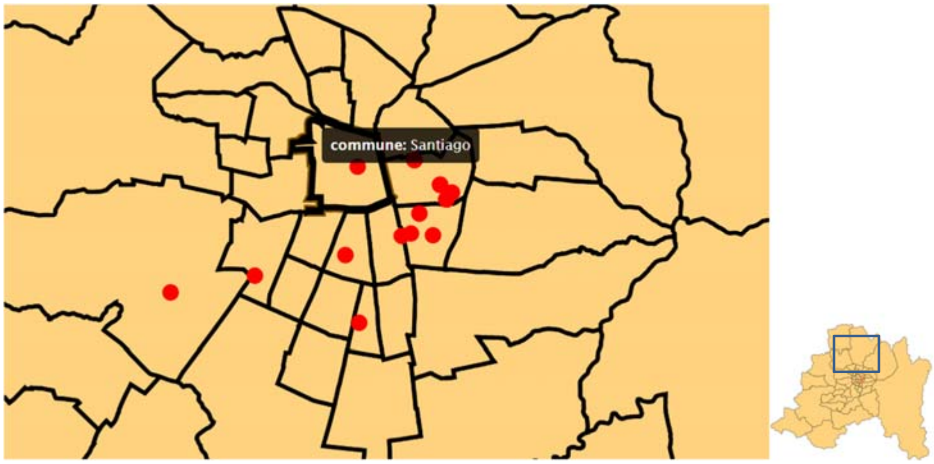

Figure 1: Location of the 15 sensors in the Metropolitan region. In the graph, only 13 sensors appear because two of them are very close to each other and thus appear overlapped.

3.2 Particle concentrations and harm

Table 1 shows the annual descriptive statistics of PM$_{2.5}$ and PM$_{10}$ concentrations with annual medians of 21.19 and 22.76 µg/m^3, respectively. Harm Intensities for PM$_{2.5}$ and PM$_{10}$ are 30 and 60 DALY/µg/m^3/10^5 person/year, giving 1271 and 683 DALYs/10^5 person/year, respectively, and a total harm of 1954 DALYs/10^5 person/year.

Table 1: Descriptive statistics and harm of particle concentrations (annual data). The worldwide statistics are in brackets. Arrows indicate whether the sample is below or above the world representative statistics from Morantes et al..

Statistic		**PM$_{2.5}$ (µg/m^3)**		**PM$_{10}$ (µg/m^3)**
Max	↑	1688 (430)	↑	2082 (350)
P$_{97.5}$	-	91.40	-	96.47
Median	↓	21.19 (26)	↓	22.76 (62)
Mean	↓	28.52 (52)	↓	32.31 (82)
P$_{2.5}$	↓	3.85 (0.022†)	↓	4.35 (17†)
SD	↓	29.92 (67)	↓	33.69 (76)
Harm (DALYs/10^5 person/year)	↓	1271.22 (1560)	↓	682.71 (1860)

† : Minimum concentration reported by Morantes et al., 2024.

The $PM_{2.5}$ and PM_{10} annual mean in 2023 were 28.52 and 32.31 µg/m^3, surpassing the WHO guidelines of 5 µg/m^3 and 15 µg/m^3. The 99th percentile of the daily 24-hour means were 94.10 and 102.11 µg/m^3 for $PM_{2.5}$ and PM_{10}, respectively, above the recommended 15 and 45 µg/m^3.

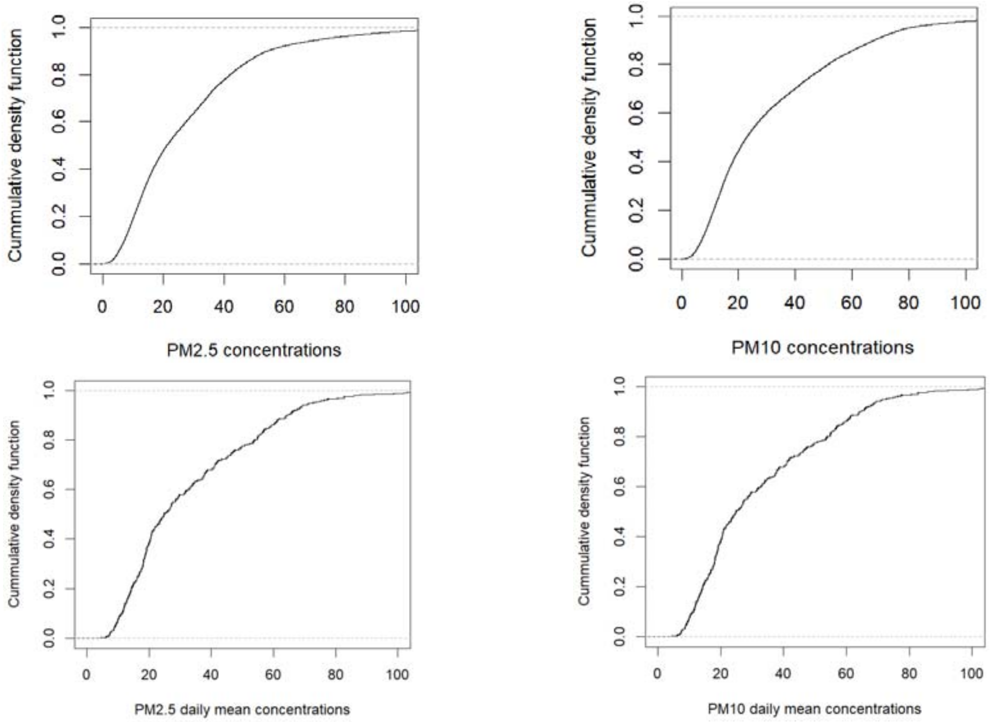

Figure 2: Cumulative density functions for PM concentrations during 2023 (above) and the daily means (below). Notice that the WHO 24-hour mean averages are met by 25% and 72% of the measured days for $PM_{2.5}$ and PM_{10}, respectively.

3.3 Hygrothermal performance

A psychrometric chart with winter and summer comfort zones visually represented each house's thermal comfort; see Figure 3. Data indicate that households fell within acceptable comfort zones during 56% of the measured time. The indoor air temperatures ranged between 9.4 and 41°C, with a median of 26°C, whereas the relative humidity ranged between 9% and 73%, with a median of 34%. The charts are available for each house in both HTML web and Python versions. The charts display a heat map, where lighter to yellow areas indicate higher concentrations of absolute humidity and indoor temperature tuples.

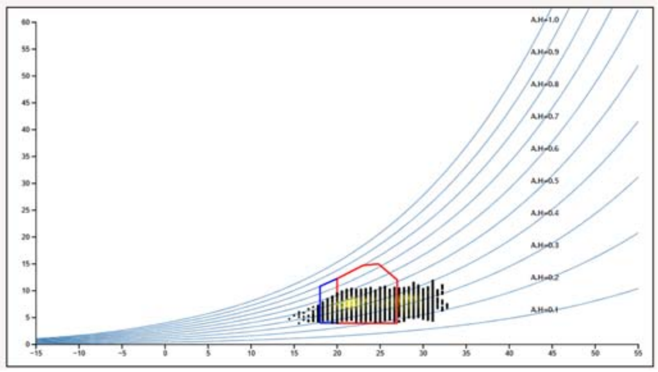

Figure 3: Psychrometric chart for one sensor, showing the data within the thermal comfort zone—red area for summer and blue and red area for winter. Gradients of yellow indicate data saturation.

For charts of other houses, visit the GitHub repository at https://eccuc.github.io.

According to the heat map shown in Figure 3, most dwellings tend to concentrate their hygrothermal conditions within the thermal comfort zone. These graphs can be viewed at https://eccuc.github.io/fondecyt/higrometric/hygrometric.html.

A rigorous future analysis might require incorporating construction materials (such as insulation materials and thermo panel windows) and conditioning elements (such as air conditioning and heating) into the equation, as they would allow for the development of a meta-statistical profile of thermal comfort in Chilean dwellings, creating more reliable profiles for developing public policies on hygrothermal conditioning.

3.4 Hygrothermal conditions during extreme or seasonal events

Overall, no statistical patterns regarding wildfires and hygrothermal conditions were found. This conclusion was reached through graphical visualization of the time-series data of temperature and relative humidity within the houses over the measured time while overlapping the extreme events and their dates. This first examination did not show a pattern worth exploring any deeper in Santiago, and the most visible pattern over time is the temperature and humidity variations, following a periodic function with high temperatures and low humidity in summer and the opposite in winter.

3.5 Comparative analysis and data clustering

The variation in indoor air quality and thermal comfort was influenced by factors such as the geographical location and the construction materials of the residences. *The k-means* clustering identifies distinct thermal comfort profiles, suggesting a need for tailored approaches in architectural design to enhance energy efficiency and thermal comfort. This is a subject of further investigation. Two primary clusters of dwellings were identified in the *Jupyter Notebook* based on the median relative humidity and temperature, with indoor relative humidity below and above 38% (Group 0 and Group 1, respectively), which showed different profiles of thermal comfort among the dwellings; see Figure 4a. A second iteration was applied using the elbow method to further sub-categorize the homes; see Figures 4a and 4b. Notice that two profiles, in red in Figure 4.a and green in 4.b, are overlapped.

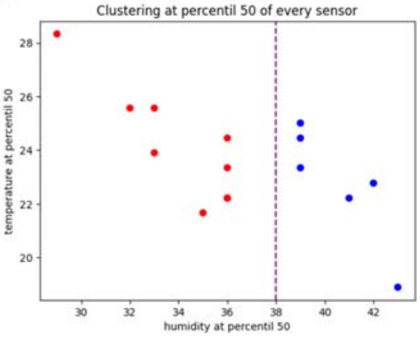

Figure 4.a: Sample clustering, groups 0 and 1.

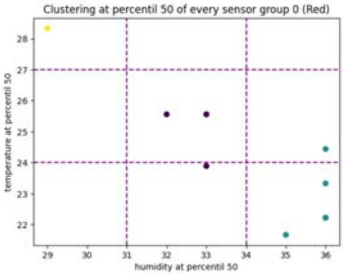

Figure 4.b: Clustering of Group 0, giving three user profiles.

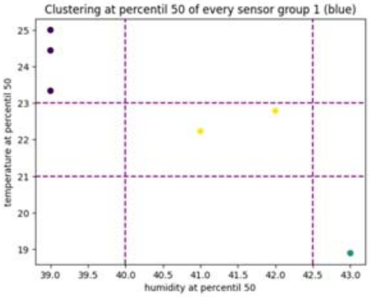

Figure 4.b: Clustering of Group 1, giving three user profiles.

In both cases, three subgroups were created within the same quadrants. However, this subclassification is temporal because, given the restricted number of homes and the lack of information about the presence of elements that might improve thermal comfort, such as HVAC equipment, it is impossible to state that these profiles represent the diverse Chilean realities.

4 OTHER LIMITATIONS

Limitations of this study include the lack of verification for potential 'deviation' in the two sensors' response once they were deployed, as well as the absence of individual uncertainty analyses and consideration of seasonal variations.

5 CONCLUSIONS

Data processing involved statistical and machine learning tools, both numerical and visual, including Python, R HTML, CSS, AND JavaScript. This study found great variability in $PM_{2.5}$ and PM_{10} concentrations in the sampled houses, with average yearly levels of 21.19 $\mu g/m^3$ for $PM_{2.5}$ and 22.76 $\mu g/m^3$ for PM_{10}. These concentrations are linked to health risks, contributing to harm quantified as 1271 DALYs/10^5 person/year for $PM_{2.5}$ and 683 DALYs/10^5 person/year for PM_{10}. The analysis has also found that homes only achieved acceptable thermal comfort levels 56% of the measured time. Future research will focus on the variability within each house and its relation with the house and household characteristics.

These findings emphasize the need for an integrated approach in residential building design, incorporating both architectural and technological solutions to achieve optimal thermal comfort and air quality. The study advocates for robust regulatory frameworks to ensure these aspects are prioritized in future housing policies. Further research is recommended to include more diverse environmental conditions and housing typologies to generalize the findings more broadly across different climates and construction styles. Additionally, incorporating more advanced metrics for air quality and thermal comfort could refine the understanding of these interactions.

For further details on quantifying the harm provided by exposure limit values (ELV) and their implications for public health, refer to the paper of this conference titled "The Protection from Harm to Populations of People Provided by Exposure Limit Values" by Jones et al..

6 ACKNOWLEDGEMENTS

Molina received funding from ANID Fondecyt Iniciación 11220965 and Garrido from IPRE-VRI 2023 from Pontificia Universidad Católica de Chile. Part of the data was gathered by André Vargas, whose work as a research assistant contributed to the quality of this research.

7 REFERENCES

Casella, George, and Berger, Roger L. (2001). Statistical Inference.

Givoni B, A. (1969). Man, Climate, and Architecture.

Kramer, O., & Kramer, O. (2016). Scikit-learn. *Machine learning for evolution strategies*, 45-53.

Morantes, G., Jones, B., Molina, C., & Sherman, M. H. (2023). Harm from Residential Indoor Air Contaminants. *Environmental Science & Technology*, *58*(1), 242-257.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences*, 11(5), 1633-1644.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP journal on wireless communications and networking*, *2021*, 1-16.

WHO. (n.d.). *What are the WHO air quality guidelines?*. World Health Organization. https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines