CHAARM: A model to predict uncertainties in indoor pollutant concentrations, ventilation and infiltration rates, and associated energy demand in Chilean houses

Constanza Molina[a,b,*], Benjamin Jones[a], Ian P. Hall[c], Max H. Sherman[d,a]

[a] *Department of Architecture and Built Environment, University of Nottingham, Nottingham, UK*
[b] *Escuela de Construcción Civil, Pontificia Universidad Católica de Chile, Santiago, Chile*
[c] *Division of Respiratory Medicine, School of Medicine, University of Nottingham, Nottingham, UK*
[d] *Lawrence Berkeley National Laboratory, Energy and Performance of Buildings Group, EETD, Berkeley, CA, USA*

Abstract

The housing stock of Chile is responsible for 15% of its total final energy consumption and so its Government is regulating the construction of dwellings. However, there is a need to establish models to help governments determine sensible guidance. This paper presents the Chilean Housing Archetypes AiR quality Model (CHAARM) and a stochastic framework for predicting uncertainties in indoor pollutant concentrations, ventilation and infiltration rates, and associated energy demand during the heating season. Pollutant sources are $PM_{2.5}$ emitted by cooking and unflued heaters present in 80% of houses.

CHAARM predicts that 66% of dwellings have a daily mean $PM_{2.5}$ concentration below the WHO 24–hour guideline value of 25 $\mu g/m^3$, even if their

*Corresponding author
Email address: cdmolina@uc.cl (Constanza Molina)

windows are always closed. Houses are not found to be airtight and 60% have $Q_{50} > 10\,\mathrm{m^3\,h^{-1}\,m^{-2}}$. Dwelling ventilation and infiltration heat loss is estimated to be 0.25–42.3 MWh with 90% confidence, and to account for at least 15% of the estimated total energy demand of the stock. Therefore, many houses require remediation measures to improve their airtightness and to reduce their annual space heating demand. However, to avoid negative health effects from exposure to $PM_{2.5}$, kitchen ventilation, such as a cooker hood, should be installed and unflued heaters should be replaced.

Keywords:
indoor air quality, uncertainty, housing stock, policy, exposure, open–source, stochastic, Monte Carlo

---

Highlights

- A probabilistic model and modelling framework to evaluate a housing stock is presented
- Uncertainties are predicted for exposures to $PM_{2.5}$, ventilation rates, and energy losses
- $PM_{2.5}$ emission rates and envelope leakage affect exposures the most

1. Introduction

A person's total exposure to airborne pollutants is a function of their time spent, and the concentrations found, in different micro-environments [1]. On average, people in countries like the UK and USA spend most of their time in their own houses [2, 3] and so pollutant concentrations in dwellings may have a greater influence on total personal exposures than outdoor concentrations.

The need to reduce the energy demand and equivalent carbon emissions (CO_2e) of dwellings can lead to unintended consequences [4]; for example, decreasing ventilation or infiltration rates can be detrimental to indoor air quality. Therefore, this phenomenon must be accounted for when interventions are applied to stocks of houses to avoid negative health effects at a population scale.

Methods for studying indoor air quality and the energy demand can be classified into two groups. Firstly, *direct* methods include monitoring or field measurements using personal or stationary equipment, and biomonitoring using biomarkers. Secondly, *indirect* methods comprise modelling and simulation techniques. Indirect approaches are often preferred to direct methods because they are less invasive and cost and time prohibitive than monitoring [5].

Although the indoor air quality and energy demand of a stock of houses can be modelled with a high level of accuracy [6], there are inevitable uncertainties in predictions. These are a function of the model itself, the heterogeneity of studied scenarios, and the random variation in inputs value, either due to lack of knowledge or natural randomness. Distributions of outputs can be generated that account for the uncertainty in model inputs when models are used with an appropriate statistical framework [7, 8], and the use of representative buildings can be used to reflect the variability between different groups [9, 10]. Additionally, sensitivity analyses can be used to determine the nature of the relationships between each of the inputs and the outputs, and to identify those that are the most important [7, 8]. Finally, the predictions can then be used to direct future research when sensitive inputs are of low quality, or to investigate the effectiveness and consequences of interventions [10, 11]. These techniques can be applied to any housing stock where there is sufficient data to generate archetypes.

Chile is a South American country whose territory traverses at least nine different climates [12], and whose >6 million houses have characteristics that vary according to the local weather conditions and the availability and affordability of building materials and energy resources. The variability of the Chilean housing stock has been studied previously and a series of representative *archetypes* have been developed by dividing the stock into groups where houses share similar properties, and each group weighted so that it represents a proportion of the whole stock[1] [13].

Logue *et al.* [14] evaluated the health impacts of a range of indoor pollutants in dwellings and identified fine particulate matter with a diameter of $\leq 2.5\,\mu\text{g/m}^3$ ($PM_{2.5}$) as the most harmful by an order of magnitude. Cooking has been identified as an important source of $PM_{2.5}$ in houses all over the world [15, 11], although its effects have not yet been investigated widely in Chilean houses. Additionally, 80% of Chilean houses use stoves for space heating, where 42% are fuelled by wood, 24% by bottled LPG, propane, and butane gases, and 10% by paraffin. The high use of wood is attributable to cultural and historical reasons, and to its low cost when compared to gas or electricity[16]. The remaining 20% of houses do not have a specific heating

[1]The code is available under a creative commons license from DOI: 10.13140/RG.2.2.16242.15041

system, using either charcoal, electricity, or solar energy [17]. Many stoves are unflued and so their contribution to total exposure is thought to be significant during the heating season [18]. It is clearly important to quantify how cooking and heating stoves affect the air quality in Chilean houses, and to understand how potential remediation measures, such as trickle vents, or opening windows, might help to dilute $PM_{2.5}$.

To do this, a wide ranging field survey is required. However, in the short term, a model of the Chilean housing stock could be used to predict uncertainty in pollutant concentrations that can be compared against international benchmarks [19] and used to identify potential impacts on population health.

This paper continues the work of Molina *et al.* [13] by developing the Chilean Housing Archetypes AiR quality Model (CHAARM). CHAARM is used to predict uncertainties in indoor pollutant concentrations, ventilation and infiltration rates and their associated energy demand, and the sensitivity of the model to its inputs. Its outputs can then be used to guide future interventions and field surveys.

The CHAARM model, its inputs, and statistical framework are described in Section 2 and a flowchart is shown in Figure 1. Section 3 presents and discusses its predictions, and Section 4 evaluates the sensitivity of CHAARM's predictions to its inputs.

2. Developing CHAARM

Stock models create a framework that can be used to evaluate and develop regulations and interventions for buildings, and the estimated outcomes can be compared against suitable indicators. Given the level of detail and quality of available data, this investigation is limited to ventilation, infiltration, and pollutant transport; see [13]. Indoor air quality is generally worse during the heating season because occupants choose to preserve their thermal comfort and minimize energy costs by closing their windows to minimize ventilation. Therefore, the models are simulated for the astronomical winter period between June 21st and September 21st. There are no reported measurements of window opening behaviour in Chilean houses at the stock scale. Therefore, two extreme scenarios are considered: (i) an all *windows open* scenario and (ii) an all *windows closed* scenario.

2.1. Model of ventilation, Infiltration, and pollutant transport

`CONTAM` [20] is a freely available multi–zone indoor ventilation and pollutant transport tool that models airflows between a building and its external environment, and between its zones. It has been validated by comparing its performance against other modelling tools [21], against measurements in controlled environments [22], and against field studies [7]. It has been used to model different types of building [23] and for evaluating input parameters [24] and pollutant concentrations [25, 26, 27]. `CONTAM` is selected over other tools because it can include multiple pollutants, multiple sources and sinks, and has 12 different emission and removal models. Therefore, many different species can be modelled simultaneously giving a more detailed representation of the indoor air.

`CONTAM` is not a dynamic thermal model and so indoor air temperatures must either be specified (see Section 2.3) or `CONTAM` must be coupled with a dynamic thermal model, such as `EnergyPlus` [21]. A dynamic thermal model can provide some predictive advantages [28], but the material properties required to model heat transfer in Chilean houses are unknown. Therefore, CONTAM is used in isolation.

Molina *et al.* [13] identified 496 common Chilean archetypes and the parameters required to describe them, showing that there is a law of depreciating returns when increasing the number of archetypes to increase the representativeness of the stock. Increasing the number of archetypes also increases the modelling and computational time and so a trade–off is required. We modelled eight archetypes in CONTAM to represent 35% of the national stock, which is predicted to be a medium representation of the stock by an *effect size* statistic [29]. The CONTAM models were manipulated using bespoke R code[2]. An example model of an archetype is shown in Figure 2a, which gives the layout of the first of the eight archetype, and in Figure 2b, which shows its implementation in CONTAM graphical user interface. For brevity, the details of each archetype are not given here, but an in–depth discussion of their parameters can be found in [13]. Rooms are represented as well–mixed zones indicated by squares, and airflow paths are indicated by diamonds. A well–mixed zone is a zone with discrete temperature, pressure, and contaminant concentration—where emissions are from a point source and instantaneously and homogeneously mixed [20]. Model inputs are described in Section 2.2. The procedures used to explore uncertainty in both the archetypes and their inputs, and to extrapolate predictions to the entire Chilean housing stock, are described in Sections 2.7–2.9.

[2] The R code is available under a creative commons license from DOI: 10.13140/RG.2.2.12641.04963.

2.2. Dwelling parameters

Dwelling properties, such as window opening area and opening schedules, are considered to be variables and were manipulated by the code. For simplicity, each archetype has fixed room volumes and floor areas, but the properties of each airflow path are variables. The models were simulated probabilistically to estimate uncertainty in predictions.

2.2.1. Airflow paths

Windows are modelled as sash types with a rectangular cross–section and a fixed cross–sectional area using the one–way orifice equation following [7, 20]. Open internal doors are modelled using the two–way flow two–opening model following [20], with a discharge coefficient of 0.78 [30] and its relative elevation is at the bottom of the door [7, 31]. Closed doors are modelled by a one–way flow power law as rectangular sections with a discharge coefficient of 0.68 and a flow exponent, n, of 0.5 [32].

In Chile, wall mounted extractors fans commonly have a minimum airflow rate of $48\,\mathrm{l\,s^{-1}}$ in kitchens and $14\,\mathrm{l\,s^{-1}}$ in bathrooms and so these rates are applied here uniformly. All bathrooms have an extractor fan and a window. Kitchens have an extractor fan and an air vent with an area of $100\,\mathrm{cm^2}$ to help dilute combustion gases and comply with Chilean Standard DS No. 66 [33]. Fans operate for both window opening scenarios according to fixed schedules; see Section 2.6.1.

All façades are assumed to be uniformly porous and the pressure distribution over vertical surfaces is assumed to be linear [8, 34]. Airleakage paths (ALPs) are modelled using a power law model and flow exponents are sampled from a normal distribution truncated between 0.5 and 1, $N(0.651,$

0.077), following [35]. Distributions of air permeability, $\dot{Q}_{50}$ ($\mathrm{m}^3\,\mathrm{h}^{-1}\,\mathrm{m}^{-2}$), have been derived for each geographic region and grouped by climatic zone; see Section 2.2.2 and Figures 5a and 5b. A value of air permeability is applied to each ALP as a function of surface area and the number of ALPs located in each wall. Party walls are assumed to be impermeable.

A single ALP is located in the floor and ceiling of each room [8], although the lower floor of each archetype is considered to be impermeable because they are assumed to be solid; see Section 2.2.2. However, the ALPs remain in the model so that they can be applied if they are required in the future. Each wall of a room has three ALPs located in at its foot, mid–point and top [34]. Blower door test simulations at 50 Pa were used to validate each model.

2.2.2. Envelope air permeability

There are important differences in construction practices in houses built before and after 2008 (see [13]) and the groups are expected to perform differently. Therefore, two different distributions of air permeability were developed; one for each construction period.

Three is very limited data for Chilean houses and so the U.S multivariate regression model [36] is used to predict a distribution of Normalised Leakage, *NL*, for Chilean houses built either before or after 2008, using the empirical data presented in [13] and following the procedure used by Chan *et al.* [36]. Here, the natural log of *NL* is given by

$$ln(NL) = \beta_{area} \cdot Area + \beta_h \cdot H + \beta_{year} \cdot I_{year} + \\ \beta_{LI} \cdot I_{LI} + \beta_e \cdot I_e + \beta_{cz} \cdot I_{cz} + \beta_{floor} \cdot I_{floor} \quad (1)$$

where $Area$ is the house floor area (m^2), H is the house height (m), I_{year} is the house construction year category, I_{LI}, and I_e is the energy performance corresponding to low–income (LI) and energy efficient houses respectively, I_{cz} is the climate zone, and I_{floor} is the air leakage of the house floor.

Data for Chilean houses is used when they are known; for example, floor area and climate zone. The I terms are assigned a value that best represent the stock, or a value of 1 if true or 0 if not, where appropriate.

The distribution of floor areas is estimated using the number of rooms and their floor areas given in [13]. Here, only the median, $79\,m^2$, is retained as the best measure of central tendency. Building height is assumed to be 3 m (2.5 m + 0.5 m for roof space), following [36].

Due to the differences in construction practices and standards between Chile and the U.S. and the lack of information on the prevalence of energy efficient houses in the stock, I_{LI} and I_e are assumed to be 1 and 0, respectively, so that they are considered to be equivalent to low–income U.S. houses. β_{floor} is set to assume that all houses have a concrete slab due to the lack of reliable data.

Climate was one of the most influential parameters in the Chan model. The International Energy Conservation Code (IECC) classification is used to match the Köppen classification [12] for Chilean and the USA climate zones, and to associate a β_{cz} with each region; these are given in the Supplementary

Materials. In order to find the β_{year} coefficient that best fit the empirical data, all estimates of β given in [36] are retained except for β_{year}. The results of β_{year} (SE) for houses built before and after 2008 are 3.490 (0.719) and 1.469 (0.845), respectively.

To obtain a national distribution, floor areas for each climate zone were sampled in sets using the Monte Carlo method. The accuracy of the prediction improves with the sample size, and so the sample was increased incrementally by the size of each set until the absolute differences of the mean (μ) and standard deviation (σ) between one set of samples and the previous set was less than 1e−6. To compare two bands of construction year, two different data sets are used; one for *old* houses and one for *new*. The model predicts *NL* 95% CI [$9.91 - 106.59$] for old houses and 95% CI [$1.39 - 15.90$] for new houses. Figure 3 shows both distributions.

To generate separate distributions of *NL* for each climate zone, the same method is conducted by sampling random values from the normal distribution of each β coefficient shown in [36].

Finally, a comparison between the *NL* distributions predicted by the model and the empirical data presented in Molina *et al.* [13] is carried out to evaluate the performance of the model; see Figures 4a and 4b. For old houses, linear regression between the measured and predicted indicates a *strong* or *high* correlation (coefficient of determination R^2 of .62 and a correlation coefficient R of .79 [37], 95% CI [$.70 - .85$]). Similarly, *NL* predictions for new houses have R^2= of .57 and a R of .75, 95% CI [$.66 - .83$].

2.3. Indoor environment

Indoor air temperatures must be specified but there are no studies of indoor air temperatures found in Chilean houses of sufficient quality. Therefore, they are sampled from a normal distribution of $N(21.1°\text{C}, 2.5°\text{C})$, following Shipworth *et al.* [38] who determined these values from measurements made in a representative sample of 196 English dwellings, and because it is used elsewhere [8]. It is clear that there is significant uncertainty in this parameter, that these temperatures are likely to be different from those found in Chilean houses, and that they may overestimate them. Therefore, the appropriateness of its application is discussed generally in Section 3 and the sensitivity of the model to indoor air temperature is tested in Section 4.

Air temperatures are constant and identical in each room and so they are not included in the daily and weekly schedules; see Section 2.6.1.

2.4. Outdoor environment

Chile is divided into 15 geographical regions. The latitude and longitude, altitude, and the main climatic zone for each regional capital city are used as the location of each modelled archetype.

Weather data is obtained from Meteonorm files [39] and used to represent each region for the heating season. Atmospheric pressure (Pa) is calculated as a function of the altitude of the nearest capital city following [40].

The meteorological wind speed is modified by the location of each house and calculated following [40]. Because the terrain type is unknown, it is randomly sampled as a function of an urban to rural ratio [13]. Wind pressure coefficients, C_p, are calculated using the Swami and Chandra model [41], following [8, 7]. House orientation is an unknown and so is assumed to be a uniform random variable between 0° and 360°. The sensitivity of the model to orientation is tested in Section 4.

2.5. Pollutants

Cooking and space heating have been identified as primary sources of $PM_{2.5}$; see Section 1. Both the background and internal initial concentrations are assumed to be zero so that only the contribution of indoor sources to the total exposure are estimated.

CONTAM requires an emission rate and a deposition rate for all pollutants, and an emission rate schedule when they are not emitted continuously.

This study only accounts for the dynamic processes of aerosols and gases associated to emissions from primary sources and their deposition onto indoor surfaces. Losses are also possible through purpose provided openings, exfiltration, and mechanical extract fans without recirculation.

It is important to model the potential loss of indoor particles due to their deposition onto, or their reaction with, indoor surfaces. Therefore, a probability distribution of deposition rates reported in the literature are sampled from a normal distribution of 0.39 ± 0.16 (h^{-1}) [42], truncated at the origin.

Finally, we do not account for outdoor $PM_{2.5}$ and so they do not contribute to indoor concentrations. The consequences are discussed in Section 3.

2.5.1. Emission rates from cooking

A synthetic cumulative distribution function of emission rates is developed to model the uncertainty in the emission rates from cooking. Data from four studies [43, 44, 45, 46] reporting $PM_{2.5}$ emission rate means and standard deviations, or an emission rate for an individual test, is combined using bootstrapping by assuming that each rate is equally likely. This distribution can be updated in the future as more data becomes available.

Cooking emissions need to be matched to the daily activities of occupants. Therefore, meals are classified into two groups: emissions from *toasting bread* are classified as *breakfast*, and all other eating activities are classified as *main meals*.

The distribution of emission rates for main meals ($N = 15,650$) has $\tilde{\mu} = 2.56\,\mathrm{mg\,min^{-1}}$, $\sigma = 4.4\,\mathrm{mg\,min^{-1}}$, and 90% CI $[0.047 - 15.2]\,\mathrm{mg\,min^{-1}}$. The breakfast distribution ($N = 4,165$) has $\tilde{\mu} = 4.32\,\mathrm{mg\,min^{-1}}$, $\sigma = 7.42\,\mathrm{mg\,min^{-1}}$, and 90 % CI $[0.072 - 21.77]\,\mathrm{mg\,min^{-1}}$. Emission rates are assumed to be constant during each cooking event [15].

2.5.2. Emission rates from heating

Six common types of heaters are used in Chile that burn gas, paraffin, and wood [18]. Their $PM_{2.5}$ emission rates into the indoor environment have been measured by [18]; see Supplementary Material for emission rates and distribution in the stock. The prevalence of each type of heating fuel varies across the country [13] and so the emission rate is assumed to be a constant determined by the fuel type and measurements, following the probability of presence of each heater type allocated by region; see Supplementary Material.

The total number of heating hours per day corresponds to those where the outdoor temperature is below 16°C, an equilibrium temperature commonly used to derive degree–days. This is calculated using the Meteonorm 7.0 weather files [39]. If the indoor temperature in unheated houses is assumed to be approximately 3°C above the external temperature, following [8], then the average time the indoor temperature is below 16°C during the winter season can be calculated and is given in the Supplementary Material.

2.6. Occupancy and activity data

The duration of emissions and activities in houses is derived from the 2015 ENUT national survey [47, 13]. Heating and cooking activities are derived for week days and weekends. Nationwide, cooking activities have a mean duration of 1 h 06 min on weekdays and 1 h 12 min at weekends. Average sleep durations are used nationally but are also given regionally in the Supplementary Material.

ENUT allows activities to be related to different room, such as *cooking* to the kitchen or *sleeping* to the bedroom, so that the total time spent in each room can be calculated. The ratio of the time a household spends in the kitchen, bedroom, and family room is 10 : 38 : 52, respectively. A similar ratio has been used by modelling studies[10] and to adjust exposure estimates. The application of occupancy patterns is discussed in Section 2.8.1.

2.6.1. Activity schedules

Archetypes occupants generally comprise 2 adults and 0–3 children; see [13]. In the example shown in Figure 2a, the household comprises 2 adults and 2 children. A fixed daily schedule is developed for each room and source using the data presented in Section 2.6 to account for the use of different rooms and to calculate occupant exposures.

Sleep duration follows the national average and is applied to the entire Chilean population. Consequently, sleeping is scheduled from 11:00 pm to

6:21 am for weekdays and from 12 am to 7:57 am for weekends.

Meal preparation is assumed to have a duration of 1 h on week days and 1 h at weekends. Meals are eaten immediately after cooking and last for an hour.

The kitchen fan is considered *ON* when cooking a meal and remains on for one hour while the meal is eaten. The Bathroom fans is considered *ON* during *showering* and *dressing*. These activities are informed by ENUT [47].

The heater is considered to be *ON* from 7 am, and functions for the same number of hours every day during the winter season. The heating duration is a function of location; see the Supplementary Material.

Bedroom doors are open during the day and closed at night. The kitchen door is closed except when cooking following [48]; see Section 2.6.1. Doors are never partially opened.

2.7. Sampling method

The sampling method follows that described by [7, 8, 11]. The model requires input variates that are specified deterministically, or are described by discrete or continuous probability distributions. They are applied to `CONTAM`, which then predicts pollutant concentrations at time intervals of 10 minutes during the winter season. These are used to calculate the winter average concentrations in the kitchen, bedrooms, and in the family room, which are weighted by the daily time the cook householder spends in each of them using the room ratio defined in Section 2.6 to give a room–weighted average (RWA) pollutant concentration. The RWA is then used to check for convergence. By systematically varying the variates and running multiple simulations, distributions of output variables are generated that quantify uncertainty in them.

There are 8 probabilistic inputs: Block aspect ratio, Δ temperature, relative north, air permeability, n exponent, $PM_{2.5}$ deposition rate, breakfast emission rate, and cooking meal emission rate. The values of each probabilistic input are obtained using Latin Hypercube Sampling (LHS) and bespoke *R* code[2] [49]. LHS is used because it improves the stratification of a samples over the probability space [50] and reduces the number of simulations required to reach convergence. They generate a value between 0 and 1 for each input, which are then applied to their inverse cumulative distribution functions (CDF) to generate an input.

Ten sets of these input variates are chosen at a time, following [7]. The total sample size increases incrementally by the set size. After each set of predictions is made, the overall μ and σ of the RWA for all sets of samples are calculated. When the change in μ and σ from the addition of one set of samples to the next is $\leq 0.5\%$ the total number of samples is deemed to have converged, and the *stopping criteria* met. This stopping criterion is chosen to reflect the lower limit of accuracy of a good Indoor Air Quality (IAQ) sensor following [8]. Simulations were run for each of the 8 archetypes in the 15 geographic regions for the window scenarios until they converge. This gives $8 \times 15 \times 2 = 240$ sets of converged data.

2.8. Post processing the model predictions

Three metrics are computed from the predictions, median ventilation rates, total $PM_{2.5}$ exposure levels, and total airflow heat losses, giving a CDF for each output for each of the 15 geographic regions. To determine national CDFs, such as Figure 6, a bootstrapping technique is used to sample from the regional CDFs weighted by the proportion of the stock located in each region.

2.8.1. Exposure analysis

An indirect approach is used to quantifying exposures by predicting indoor pollutant concentrations over time in each room, and by making assumptions about occupant behaviour; see Section 2.6. Average hourly $PM_{2.5}$ concentration profiles for winter and the contact times of the cook householder are used for the exposure assessment, following the population–weighted method of [42]. Composite hourly concentration profiles are used to produce time–weighted averages (TWA) based on the behaviour of the cook householder. For example, bedroom concentrations are used when the occupants are asleep, bathroom concentrations when washing, kitchen concentrations when cooking, and living room concentrations at all other occupied times.

2.8.2. Ventilation and heat loss

Hourly average and median airflow rates were calculated for each dwelling by combining infiltration and ventilation rates. The associated heat loss as a function of time, $H(t)$ (kW), is then calculated to be

$$H(t) = \int \dot{V}(t) \cdot \overline{\rho}(t) \cdot c \cdot \Delta T(t) \cdot dt \tag{2}$$

Here, $\dot{V}$ (m^3/s) is airflow rate, $\overline{\rho}$ (kg/m^3) is the mean of the indoor and outdoor air densities, c $(kJ/kg/K)$ is the specific heat capacity of air, and ΔT (K) is the difference between the indoor and outdoor temperatures when the indoor temperature is greater, otherwise it is assumed to be 0°C. Equation 2 is integrated over the winter to estimate the total heat loss, H (kWh).

2.9. Sensitivity analysis

The model is non-linear and the distributions of inputs vary. Therefore, it is difficult to state *a priori* the types of relationships that exist between the model inputs and outputs and their strength. Thus, a global sensitivity analysis (SA) is used to test the dependence of the three outputs on the twelve inputs. However, a fundamental requirement of the SA is that all the tested inputs are independent of one other, and so any that are themselves correlated are combined. Therefore, nine inputs are used directly and three are scaled using house characteristics to avoid multicollinearity.

All inputs and outputs are unique for each house, except for the heater emission rate and envelope area to volume ratio because they relate to a specific household appliance and archetype, respectively. To compute representative values for the wind speed, the median wind speed scaled at house height is used (see Section 2.4), and ΔT is the difference between the indoor air temperature and the median outdoor temperature.

We follow the method of Jones *et al.*[3] [8, 7], which tests for linear, monotonic, and non-monotonic relationships between the inputs and outputs. The tests for linear relationships are: (i) Kendall's τ rank, (ii) Pearson's r product moment correlation coefficient, and (iii) linear regression. Monotonic relationships are tested using: (iv) Spearman's ρ rank correlation coefficient, (v) rank-transformed standardised variables. Non–monotonic relationships are tested using: (vi) Kolmogorov–Smirnov and (vii) Kruskal–Wallis quantile tests.

[3] The code was used under a creative commons license and obtained from DOI: 10.13140/RG.2.2.21670.88644

Depending on the statistical method, the test coefficients are useful for identifying the inputs that are more important (two or more sample-comparison methods), more related (by using correlation-based methods), and/or contribute the most to the outputs (using the regression-based models); see [7] for a more detailed description of each test and the procedure. The methods applied estimate the total effect of each element of an input on each element of an output, where the hypothesis is that there is a relationship between an input and an output.

The input and output data are not transformed, and all outliers are retained. Data for both window scenarios are merged and are tested together. Coefficients and p–values are obtained for each test, and the inputs are ranked according to the magnitude of the coefficient. The p–values can be used to determine whether a result is statistically significant at a predefined level of significance. We use a 5% level herein.

2.10. Statistical tests

Most statistical tests present a coefficient and a p–value, which indicates the probability of obtaining results at least as extreme as those obtained during a test, assuming that the null hypothesis is true. The *null hypothesis* is a general statement that there is no relationship or association between groups. CHAARM generates a significant number of data points, see Section 3, which can make a p–value meaningless because the probability of significance increases with the sample size [51]. Furthermore, the 5% significance threshold used herein (see Section 2.9) is arbitrarily, and so we focus on the nature and the magnitude of any effect [52] where possible.

To test the occurrence of an effect in the the medians of categorical variables between archetypes and regions a Kruskal–Wallis H test is applied, where the null hypothesis is that all samples originate from the same population. Then, *post–hoc* pairwise multiple comparison tests are used to determine the location of the difference and to identify which pair of samples differ significantly, following [53]. A Levene statistic is used to test the homogeneity of variance using medians; the null hypothesis is that the variances in different groups are equal [53]. And finally, effect sizes are used to identify the magnitude of the difference between two samples, following Ferguson [29] and using Cohen's d. Effect sizes are useful and objective estimates of the magnitude of an effect that is not influenced by the sample size, thus providing a better measure of the magnitude of the effect between two samples [54, 53], and so they are used to identify groups that need to be assessed separately. Thresholds are used to label the effects where $d < 0.2$ corresponds to a *negligible* effect size, $0.2 \leq d < 0.5$ to a *small* effect size, $0.5 \leq d < 0.8$ to *medium* effect size, $0.8 \leq d < 1.3$ to a *large* effect size, and $d \geq 1.3$ corresponds to a *very large* effect size.

The coefficient of variation, C_V, is the quotient of the standard deviation and the mean, σ/μ. It is a descriptive statistic used to measure the variability of any value and can be used to compare different distributions because it is dimensionless.

Values of kurtosis and skewness are used to characterize the variability of the data and to identify a central value that best describes it. Kurtosis is a measure of the size of the tails relative to a normal distribution. The kurtosis for normally distributed data is three but is adjusted to zero using an *excess* kurtosis, and is applied here. Therefore, data with high kurtosis has large tails and a large number of outliers. The skewness for normally distributed data is zero and positive values indicate data that are right–skewed with a long right tail.

3. Model predictions

Approximately $2,100$ simulations were required per archetype per scenario to achieve convergence. They are aggregated over hourly, daily, seasonal, periods so that they can be compared against relevant benchmarks. Table 1 gives a summary statistics at the national scale for the two window opening scenarios over the winter period. The 90% confidence intervals show the lower and upper limits of the predicted exposure levels, ventilation rates, and heat loss, and exclude those that are unlikely to occur. The coefficient of variation, C_V, (see Section 2.10) shows that winter exposures are more variable than the ventilation rates or heat loss, and the difference in variation between the *windows closed* and *windows open* scenarios is similar for all outcomes. The lowest variability is seen in ventilation rates for the *window open* scenario ($C_V = 0.50$) because increasing opening areas increases the magnitude of the smallest airflow rates more than the highest, whereas the largest is variability in exposures for the *window open* scenario ($C_V = 1.85$) because the higher airflow rates increase dilution and so the magnitude of the smallest exposures decreases substantially more than the largest.

Table 1 gives skewness and kurtosis statistics (see Section 2.10) for the three outcomes, which indicate that their distributions are all positively skewed and heavily–tailed. This suggests that the use of the median, instead of the mean, is more appropriate for policy–making or benchmarking.

Figure 6 shows CDFs of predicted daily mean $PM_{2.5}$ concentrations nationally. This graph can be used to visualise the boundaries of the Chilean problem and to see the maximum impact of window opening behaviour on hourly exposures.

Table 2 presents a statistical summary of the predicted median exposures and ventilation rates by archetype for the *windows closed* scenario. Archetype properties are given in Table 4 of the Supplementary Material of [13]. Figure 7 shows the distributions of the median hourly exposures to $PM_{2.5}$; the concentrations that occupants are exposed to half of the time. Median hourly exposures for the *windows closed* scenario are generally higher in archetypes representing newer and more airtight houses. This is unsurprising given the significant difference in the distributions of *NL* between old and new dwellings; see Figure 3. Conversely, the *windows open* scenario shows that there are negligible differences between archetypes, and exposures will be close to ambient concentrations. This indicates that ventilation via windows and fans is independent of the archetype, and that windows are an effective mitigation method against exposure to $PM_{2.5}$ emitted by heaters and cooking, although not necessarily a pragmatic one.

Levene and Kruskal–Wallis tests are used to compare predictions by archetype and by region. All tests are found to be significant with $p \ll .001$ and $p \ll .05$, respectively, indicating that there is a significant difference in the variances of predictions for each archetype and region, although we note the problems when interpreting p–values described in Section 2.10. Effect sizes are calculated using median values and Cohen's d thresholds and show high variability in effect size (negligible, small, medium, and large) between pairs of archetypes and regions, and for the two window scenarios. Thus, the statistical significance of the tests, and the magnitude of the effects, generate confidence in the use of the archetypes for analysing different types of house. It shows that it would not be appropriate to aggregate the entire Chilean housing stock into a single archetype. Some *negligible* effects sizes are seen between some neighbouring regions for the three outcomes (exposure, ventilation rate, and energy demand), suggesting that they could be joined by proximity and analysed together. This amalgamation would save time and computational resources. It also means that interventions can be targeted at all houses where similar consequences should be expected. All test statistics are given in the Supplementary Material.

3.1. Exposure

Generally, $PM_{2.5}$ concentrations are found to be high, especially in the kitchen during cooking periods, but also in the living room and bedrooms. Doors are assumed to be open when cooking activities are taking place, which will contribute to the spread of pollutants. The impact of door opening on indoor air quality should be an area of future research.

To obtain a more representative estimation of occupant exposures to indoor $PM_{2.5}$, TWAs are used over RWAs; see Sections 2.8.1 and 2.7. RWAs are found here to be around 20% lower than TWAs confirming that there is a significant difference between the two metrics. A TWA is dependent on an occupant's presence in each room over time. An analysis of activity patterns is required to develop TWAs that account for actual occupant behaviour and to answer related social research questions. Here, this pattern is assumed to be the same for all houses because there is no understanding of uncertainty in this parameter.

The limits of the mean TWA $\mathrm{PM}_{2.5}$ winter exposures for the two window scenarios are $6.61 \leq \mu \leq 134.47\,\mu\mathrm{g/m}^3$ and bound the mean values of exposure predicted by Das *et al.* [7] for the English stock. Das *et al.* considered kitchen windows to be open between 0.01–10 times the during of the cooking period, giving $\mu = 12.7\,\mu\mathrm{g/m}^3$, $\sigma = 12.6\,\mu\mathrm{g/m}^3$, and $P_{50} = 8.0\,\mu\mathrm{g/m}^3$. Although this is reassuring, these values are not directly comparable because Das *et al.* weighted the hourly concentrations they predicted for each room by their volume to calculate a mean dwelling concentration for the heating season.

The World Health Organization (`WHO`) recommends that mean $\mathrm{PM}_{2.5}$ concentrations in ambient air are less than $10\,\mu\mathrm{g/m}^3$ per year and $25\,\mu\mathrm{g/m}^3$ per day [19]. These guidelines are also applicable to the indoor environment because there is not yet any convincing evidence of a difference in the hazardous nature of particulate matter from indoor and outdoor sources [55]. Here, the `WHO` guideline daily mean value is used to determine the *acceptability* of IAQ [56], although we acknowledge that outdoor $\mathrm{PM}_{2.5}$ would contribute to the indoor concentration and that our evaluation systematically underestimates total the total exposure to $\mathrm{PM}_{2.5}$ and that any future evaluation of health effects using CHAARM would have to account for this. However, the difference is not a simple addition of indoor and outdoor $\mathrm{PM}_{2.5}$ because the transfer of outdoor $\mathrm{PM}_{2.5}$ is dependent on the *penetration coefficient* of a building's fabric, a non—dimensional parameter between 0 and 1 that represents its filtering effect [57, 11, 58]. Nevertheless, Figure 6 shows that 34% of Chilean dwellings are predicted to have unacceptable daily $\mathrm{PM}_{2.5}$ concentrations if their windows are closed at all times, and so their occupants have an elevated risk of experiencing negative health outcomes, such as tracheal, bronchial, and lung cancers, Chronic Obstructive Pulmonary Disease, ischaemic heart and cardiovascular diseases, and lower respiratory infections. The remaining dwellings must keep their windows open all winter, when indoor concentrations will tend towards outdoor $PM_{2.5}$ concentrations because the penetration factor is 1, or be *leaky* when outdoor $PM_{2.5}$ will have a lower effect on indoor $PM_{2.5}$ because the penetration factor is between 0.7 and 0.9 [11].

3.2. Ventilation and total heat loss

Figure 8 shows that a ventilation rate of $>13\,\mathrm{h}^{-1}$ is required to ensure that 95% of the stock is below the `WHO`'s guideline daily mean value of $25\,\mu\mathrm{g/m}^3$. However, the associated energy demand is predicted to be 33.4 MWh for the winter season, which is cost and carbon prohibitive. Clearly, it is sub–optimal to prescribe a single ventilation rate for all houses to meet the `WHO` $PM_{2.5}$ guideline value. Table 1 shows that $13\,\mathrm{h}^{-1}$ can never be achieved by infiltration alone and so some window opening is always required. Then, the `WHO` guideline value can be achieved in around 50% of dwellings. Providing general ventilation is not always enough, and so removing pollutant sources, such as gas, paraffin, and wood heaters, or installing and using targetted ventilation, such as cooker hoods (also known as *range* hoods), are important remediation measures that can help to simultaneously provide acceptable air quality and minimize energy demand. The use of cooker hoods has not been considered by CHAARM because there is little information about their implementation in Chilean homes and about their *capture efficiency*, a metric that describes their ability to extract pollutants before they mix in the kitchen. However, as this information becomes available, it is straightforward to implement in `CONTAM` by following the method of O'Leary *et al.* [11].

A ventilation rate of $0.5\,\mathrm{h}^{-1}$ is threshold rate used by some European countries because some negative air quality related health effects are thought to increase below it, although there is significant uncertainty in this value [8]. Hourly ventilation rates for the *windows closed* scenario are predicted to be $< 0.5\,\mathrm{h}^{-1}$ 39% of the time, which is less than the estimated times for dwellings in the USA (57% of the time), England, and Beijing. This is partly because the US and Chinese studies do not consider the use of mechanical ventilation in bathrooms and kitchens, but also because because Chilean houses are less airtight; see Figure 5. However, it does suggest that Chilean dwellings need more air during the winter than is provided by a combination of infiltration and bathroom and kitchen fans. If the Chilean government was to seek to reduce the energy demand of its stock by increasing its airtightness, the proportion of houses with a mean winter airflow rate of $\leq 0.5\,\mathrm{h}^{-1}$ will inevitably increase and could cause negative health effects unless additional ventilation is provided.

Table 1 shows that the heat loss in a Chilean house during the winter is estimated to be $0.25 \leq H \leq 42.3\,\mathrm{MWh}$ with 90% confidence. It is possible to determine the uncertainty in the mean energy demand of all 5.8 m Chilean dwellings [13] attributable to airflow in winter using the distribution of energy demand. This is done by repeatedly sampling from it in sets of 5.8 million until the mean of the means is normally distributed [59]. The mean total heat loss attributable to airflow is estimated to be 8.2 TWh for the *windows* *closed* scenario and 124.0 TWh for the *windows open* scenario.

The Chilean Technology Development Corporation (CDT) [60] estimates the total energy demand of the housing stock to be 53.8 TWh per year, and so the *windows closed* and *windows open* scenarios account for 15%–230% of the estimated total, respectively. These values are broadly plausible given that the scenarios explore extreme conditions. However, the *windows open* scenario is the least plausible because occupants are unlikely to simultaneously heat their houses and leave their windows permanently open, which explains the significant over–estimation of energy demand for this scenario. A field survey is required to understand the window opening behaviour of occupants. Furthermore, CHAARM's energy demand predictions do not account for the efficiency of heating systems because their stock–wide distribution is unknown, and so they are not directly comparable with the CDT value. If it was possible to predict the primary and secondary energy required to provide space heating, the range for the two scenarios would increase significantly. This calculation has been done by Jones *et al.* [8], who explore the *windows closed* scenario without fans for the UK housing stock. They estimate that infiltration is responsible for 11–15% of UK housing stock energy demand and account for the efficiency of heating systems. Removing the fans from the CHAARM model would reduce the lower limit, but only slightly because they run infrequently. Accounting for the efficiency of heating systems would increase the lower limit significantly. Directly applying a UK stock–average heater efficiency of 76% increases the lower limit to 20%, indicating that Chilean dwellings have higher infiltration rates that UK dwellings.

4. Model Sensitivity

The assertions made in the previous sections about the CHAARM's predictions are dependent on the assumptions made in Section 2. Therefore, the SA described in Section 2.9 is used to determine the relative importance of its inputs. Tables 3–5 rank the inputs for each test, but the test statistics used to determine the ranks and their p–values are given in the Supplementary Material. A rank of 1 indicates the most sensitive input, and Table 3 shows that $PM_{2.5}$ winter exposures are most strongly correlated with the $PM_{2.5}$ emission rate from cooking, followed by the permeable envelope area, A_{perm}, and the heater emission rate. Table 4 shows that ventilation rates are most strongly affected by A_{perm} and ΔT. And finally, Table 5 shows that the total heat loss is also most sensitive to A_{perm}, but ΔT is the second–ranked input, and dwelling air permeability Q_{50} is the third.

There is uncertainty in $PM_{2.5}$ emission rates from cooking because they are a function of many factors [15]. Although empirical data from North America and Europe was used to derive a PDF of emission rates, it is impossible to say whether it is representative of $PM_{2.5}$ emissions from Chilean cooking without corroborative measurements. Nevertheless, the importance of this metric to exposure suggests that cooker hoods should be installed in all new Chilean houses and should also be installed in any existing house whose airtightness is improved. The cooker hood should extract cooking pollutants directly outside and should not recirculate [11]. The $PM_{2.5}$ emission rate of heaters that burn gas, paraffin, and wood is also an important determinant of exposure, and so they should be targetted for removal from Chilean houses in the near future. Table 2 shows that they should not be installed in new homes.

The A_{perm} parameter may differ from the thermal envelope area if the floor is solid and if party walls are assumed to be impermeable; see Section 2.2.1. Here, all party walls are assumed to be impermeable and all ground floors solid. Party wall permeability can only be determined by *guarded zone* blower door tests [34]. These are non–standard tests that are rarely conducted and, to the best of our knowledge, are not mandated by the regulatory authority of any country. This will remain an uncertain parameter, although the effects of party wall impermeability could be tested in the future using CHAARM by following the method of Jones *et al.* [8]. The ground floor type should be added to the Building Permit database for new houses, and be a recorded parameter in future surveys of existing dwellings; see [13].

Indoor air temperature was highlighted as a parameter with high uncertainty in Section 2.3. Its magnitude affects the ventilation rate and is incorporated into the ΔT parameter of Equation 2 to predict total heat loss, H. Tables 4 and 5 show that the ventilation rate and H are both sensitive to ΔT and so empirical data is urgently required. Recently, the Chilean government released a database of measurements of indoor air temperatures made in nearly 300 homes [61], which will be processed and incorporated into a future version of the model.

The air permeability of houses is discussed in Section 2.2.2 and shows that there are limited measurements of air leakage rates in Chilean houses because they are not yet a legal requirement, although this is expected to be changed in the near future. Field work is required to measure the airtightness of the most common archetypes.

Finally, Section 2.4 shows that dwelling orientations are unknown, but the SA shows that all 3 outputs are insensitive to it at stock scale, which is consistent with [31, 7], and [8]. However, coupling CONTAM with a dynamic thermal model might show a different effect in the outcomes.

The understanding of the CHAARM model provided by the SA and the discussion of its outputs (see Section 3) shows that there are many areas that should be improved by gathering more data. Furthermore, the outputs, in the form of probability distribution functions, are useful tools that policy makers can use to make informed decisions about the energy demand of Chilean houses and its relationship with indoor air quality.

5. Conclusions

This paper presents the Chilean Housing Archetypes AiR quality Model (CHAARM) and a stochastic framework for predicting uncertainties in indoor pollutant concentrations, ventilation and infiltration rates, and their associated energy demand during the winter season. Pollutant sources are $PM_{2.5}$ emitted by cooking and gas, paraffin, and wood heaters. Outdoor $PM_{2.5}$ are not considered and so the exposure analysis is restricted to indoor $PM_{2.5}$, leading to a systematic underestimation of total exposures. Because window opening behaviour in Chilean houses is not understood, two extreme scenarios are considered; a *windows open* at all times scenario and a *windows closed* at all times scenario. A distribution for each output is produced for each scenario. They show that 66% of Chilean dwellings are predicted to have a daily mean $PM_{2.5}$ concentration below the WHO 24 hour guideline value of $25\,\mu g/m^3$, even if their windows are closed at all times. This suggests that most houses are not airtight. This is confirmed by a synthetic distribution of air permeabilities for houses built before 2007, representing 66% of the stock, which shows that over 90% of them have $Q_{50} > 10\,m^3\,h^{-1}\,m^{-2}$ (95% CI $[9.91 - 106.59]\,m^3\,h^{-1}\,m^{-2}$). Therefore, many of these houses require remediation measures to improve their airtighness and reduce their annual space heating demand. However, to avoid negative health effects from exposure to $PM_{2.5}$ from cooking and heaters, cooker hoods should be installed and the heaters should be replaced.

Ventilation provided by windows and fans is found to be independent of dwelling archetypes and an effective mitigation method against exposure to $PM_{2.5}$ emitted by heaters and cooking, although not necessarily a pragmatic one. Moreover, a recent study has shown that there is a law of diminishing returns in the relationship between effective area and opening angle, which may influence the impact of window opening related inputs to the three outcomes [62]. Heat loss in a Chilean house during the winter is estimated to be $0.25 \leq H \leq 42.3$ MWh with 90% confidence, and to account for 15%–230% of the estimated total energy demand of the housing stock, although this interval does not account for the collective efficiency of the heat sources. The implausibility of the *windows open* scenario makes the lower limit more likely, as people tend to keep windows closed during the heating season to safe energy. Note that this paper presents the two extreme scenarios. Although this binary assumption contributes to our understanding of the uncertainties in the three outcomes, it is clear that a better description of this and other inputs highlighted here would improve the representation of the actual condition of the stock.

CHAARM used 8 archetypes, which is found to be appropriate, and the stock cannot be represented by a single archetype. However, the model is a work in progress and will be updated as more data becomes available. A sensitivity analysis shows that there is a pressing need for knowledge of indoor air temperatures, dwelling air permeabilities, and occupant behaviour.

Declaration of competing interests

None.

Acknowledgements

This work was supported by the National Agency for Research and Development of Chile, ANID (Grant number 72160504).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at DOI: 10.13140/RG.2.2.12221.61920.

Tables

Table 1: Statistical summary of winter $PM_{2.5}$ exposures, ventilation rates, and heat losses, nationwide. C_V: coefficient of variation. P_n: nth percentile.

	Exposures (μg/m^3)		**Ventilation rates (h^{-1})**		**Heat loss (kWh)**	
Statistic/Window	Closed	Open	Closed	Open	Closed	Open
Mean, μ	134.47	6.64	0.89	15.80	1283.85	19526.08
Median, P_{50}	58.65	2.30	0.75	13.57	947.44	16681.37
Standard deviation, σ	210.28	12.26	0.79	7.91	1190.02	12556.3
90% CI	[2.58; 548.72]	[0.08; 29.81]	[0.08; 2.40]	[8.20, 29.75]	[252.6; 3471.2]	[5736; 42342]
P_{10}	5.20	0.22	0.13	8.85	319.3	7614
P_{25}	19.06	0.77	0.31	10.42	508.8	11053
P_{75}	154.20	6.77	1.16	19.14	1624.5	24686
P_{90}	367.06	16.94	1.84	24.75	2585.7	34295
C_V	1.56	1.85	0.89	0.50	0.93	0.64
Skewness	4.12	4.26	2.17	2.47	3.00	2.18
Kurtosis	34.62	26.36	9.19	11.31	16.07	9.48

Table 2: Statistical summary of median hourly $PM_{2.5}$ exposures (μg/m^3) and median ventilation rates (h^{-1}) for the *windows closed* scenario aggregated by archetype ID. See [13] for archetype details. Bold IDs indicates those built after 2007.

ID	$PM_{2.5}$ exposure (μg/m^3)			Ventilation rate (h^{-1})		
	P_{50}	μ	σ	P_{50}	μ	σ
27	2.75	9.94	9.14	0.8	0.9	0.3
36	2.22	6.12	5.90	1.1	1.0	0.3
91	2.81	12.68	11.88	0.7	0.7	0.2
100	1.09	4.24	4.32	1.0	1.0	0.3
275	17.52	25.46	18.45	0.1	0.1	0.0
35	0.99	5.44	5.31	1.0	1.1	0.3
19	4.10	11.76	11.02	0.8	0.8	0.2
284	7.61	16.20	12.68	0.2	0.2	0.1

Table 3: Sensitivity analyses of the input parameters on the exposure to $PM_{2.5}$ over the winter. Bold ranks, $p < .05$.

Input	Kendall	Pearson	Spearman	Regression	Rank Regression	KS	KW P_2	KW P_5	KW P_{10}	KW P_{20}
$L:W$	**10**	**9**	**10**	9	10	**11**	**11**	**11**	**10**	**12**
Orientation	**11**	**12**	**11**	12	11	**10**	**10**	**10**	**12**	**10**
Q_{50}	5	3	5	3	5	6	5	5	6	6
n	6	5	6	5	6	9	6	7	8	8
k	7	4	7	4	7	8	7	9	7	9
G breakfast	**12**	**11**	**12**	11	12	**12**	**12**	**12**	**11**	**11**
G cooking	1	1	1	1	1	1	1	1	1	1
ΔT	8	7	8	7	8	7	9	6	5	5
Wind speed	4	**10**	4	10	4	5	3	4	4	4
G heater	3	6	3	6	3	4	4	3	3	3
$S:V$	9	8	9	8	9	3	8	8	9	7
A_{perm}	2	2	2	2	2	2	2	2	2	2

$L:W$, length to width ratio; Q_{50}, air permeability; k, $PM_{2.5}$ deposition rate; n, flow exponent; G, emission rate;
$S:V$, surface area to volume ratio; A_{perm}, permeable area.

Table 4: Sensitivity analyses of the input parameters on ventilation rate. Bold ranks, $p < .05$.

Input	Kendall	Pearson	Spearman	Regression	Rank Regression	KS	KW P_2	KW P_5	KW P_{10}	KW P_{20}
$L:W$	7	5	7	5	7	7	**8**	7	7	7
Orientation	**8**	**8**	**8**	8	8	**8**	**7**	**8**	**8**	**8**
Q_{50}	2	7	2	7	2	2	2	2	2	2
n	6	6	6	6	6	6	4	6	6	6
ΔT	3	4	3	4	3	5	3	5	5	5
Wind speed	4	2	5	2	5	4	6	3	3	3
$S:V$	5	3	4	3	4	3	5	4	4	4
A_{perm}	1	1	1	1	1	1	1	1	1	1

$L:W$, length to width ratio; Q_{50}, air permeability; k, n, flow exponent; $S:V$, surface area to volume ratio; A_{perm}, permeable area.

Table 5: Sensitivity analyses of the input parameters on heat loss. Bold ranks, $p < .05$.

Input	Kendall	Pearson	Spearman	Regression	Rank Regression	KS	KW P_2	KW P_5	KW P_{10}	KW P_{20}
$L:W$	**7**	**6**	**7**	6	7	**7**	**8**	**8**	**7**	**8**
Orientation	**8**	**8**	**8**	8	8	**8**	**7**	**7**	**8**	**7**
Q_{50}	3	5	3	5	3	3	2	2	2	2
n	6	**7**	5	7	5	6	5	6	6	6
ΔT	2	2	2	2	2	5	3	3	3	3
Wind speed	4	3	4	3	4	4	4	4	4	4
$S:V$	5	4	6	4	6	2	6	5	5	5
A_{perm}	1	1	1	1	1	1	1	1	1	1

$L:W$, length to width ratio; Q_{50}, air permeability; k, n, flow exponent; $S:V$, surface area to volume ratio; A_{perm}, permeable area.

Figures

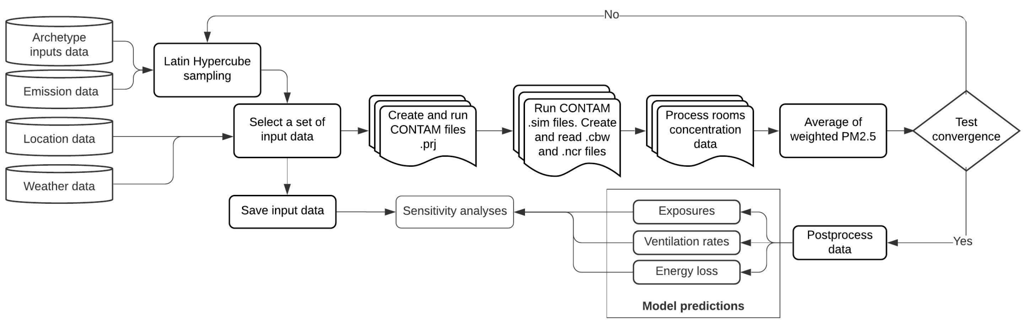
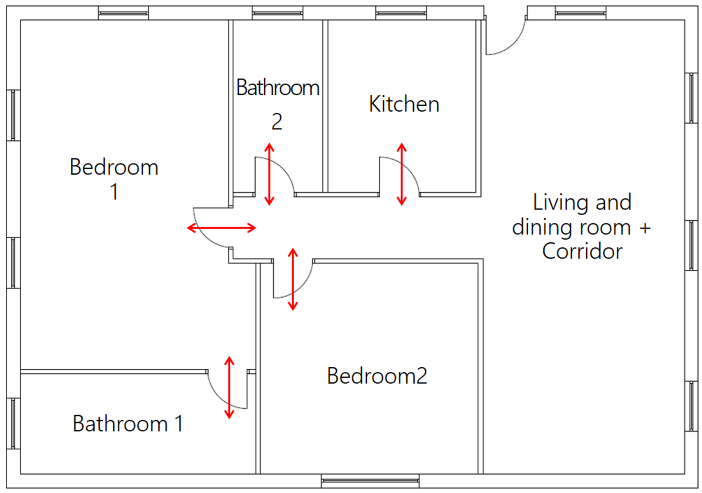
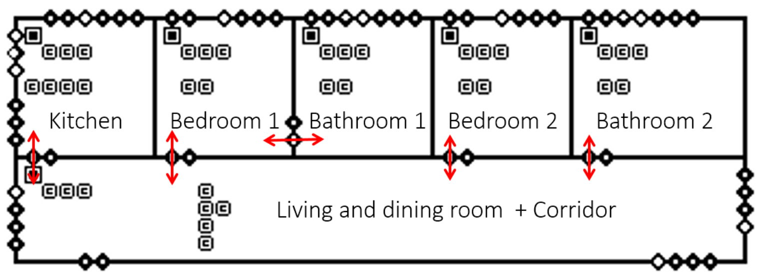


Figure 1: Flow diagram of the CHAARM model.



(a)



(b)

Figure 2: Archetype 27 (a) layout and (b) CONTAM model.
See [13] for archetype details.

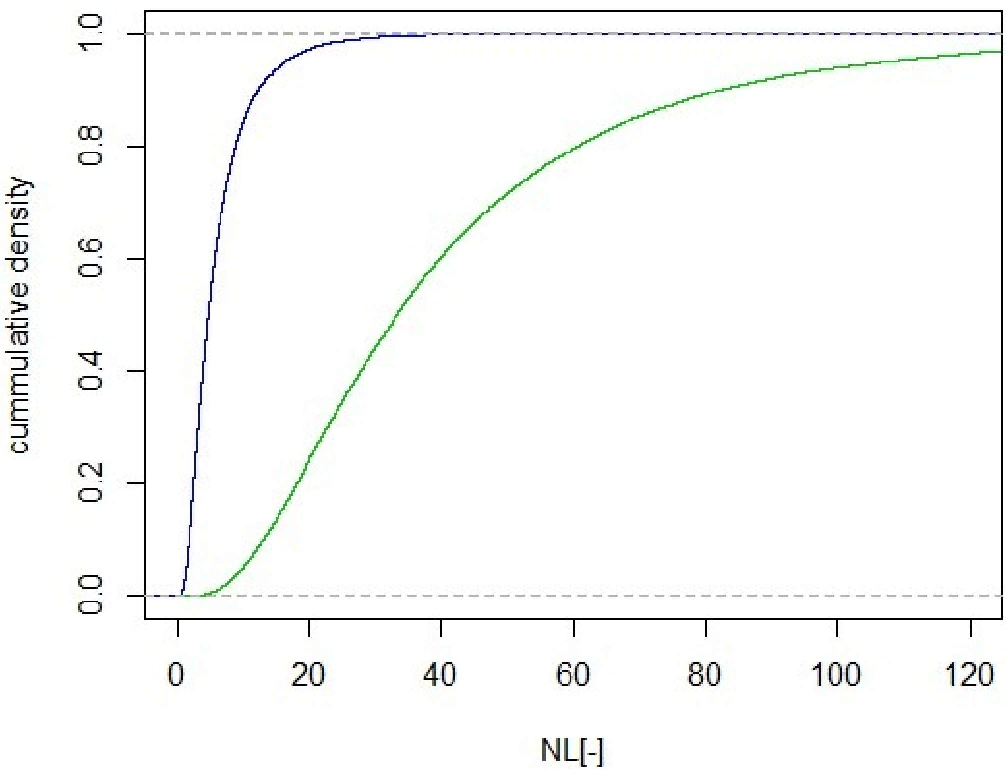
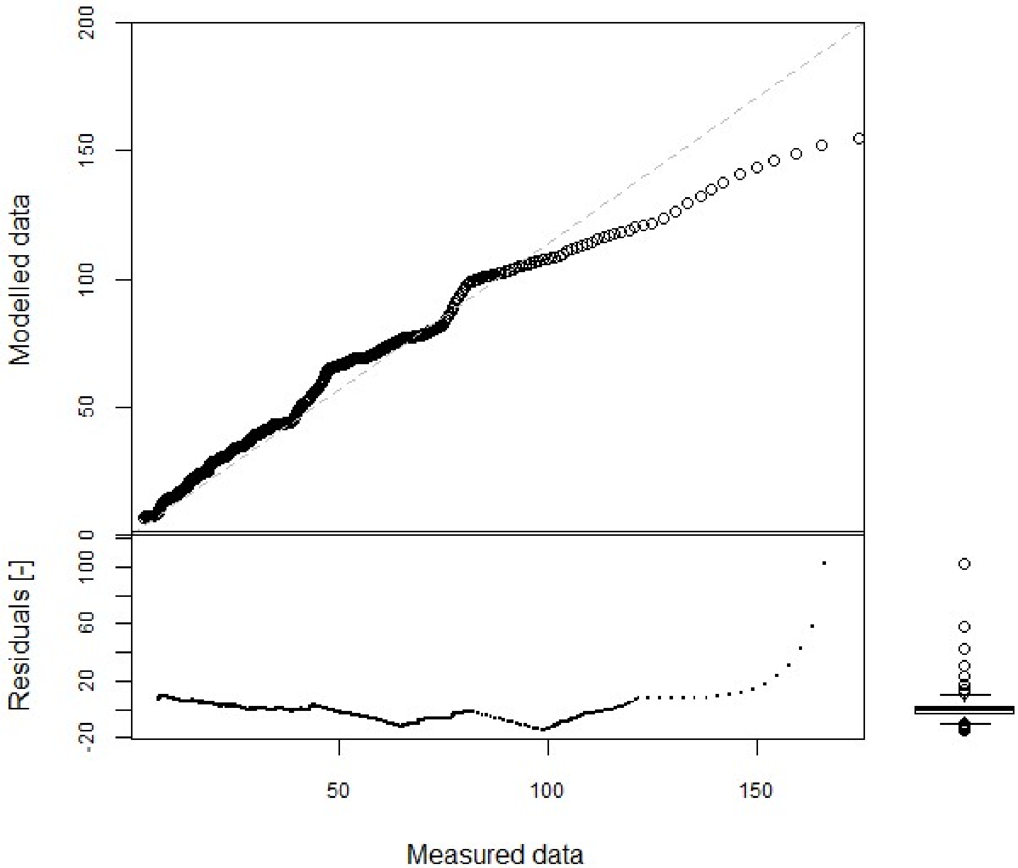
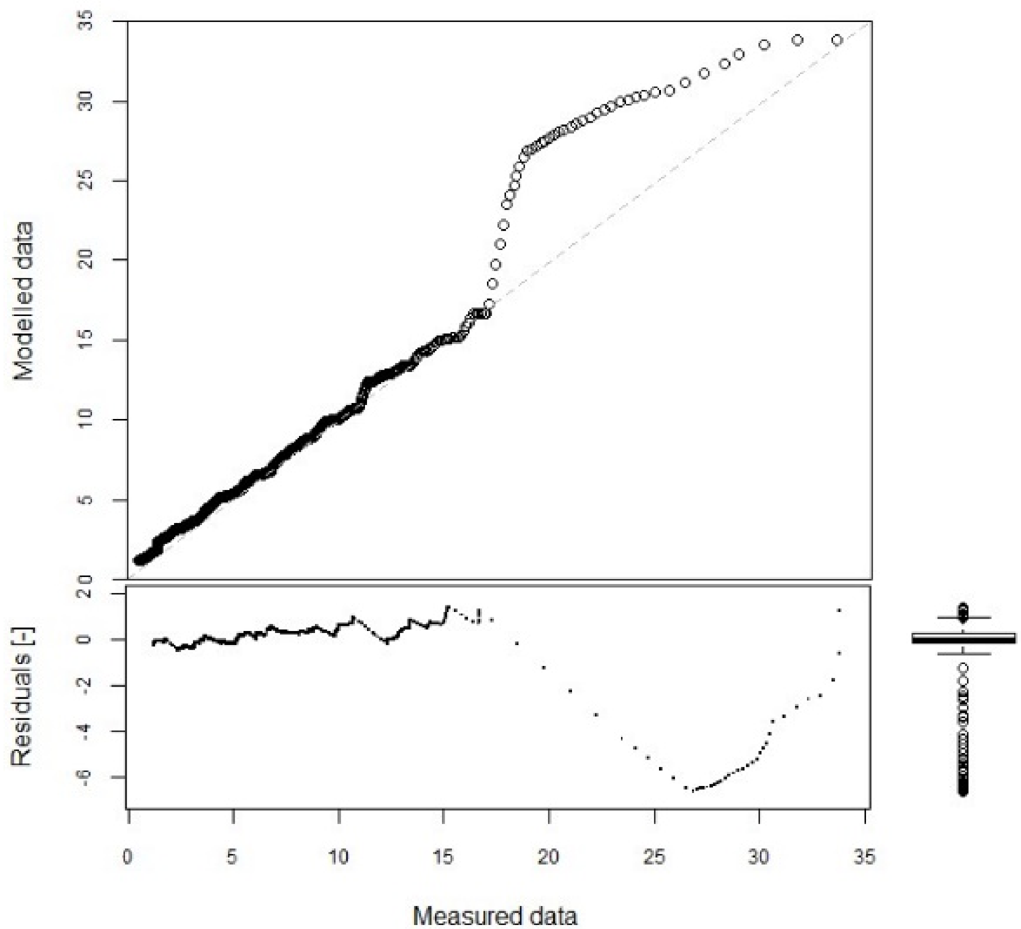


Figure 3: Cumulative distribution national of *NL* for old (green line) and new houses (blue line).



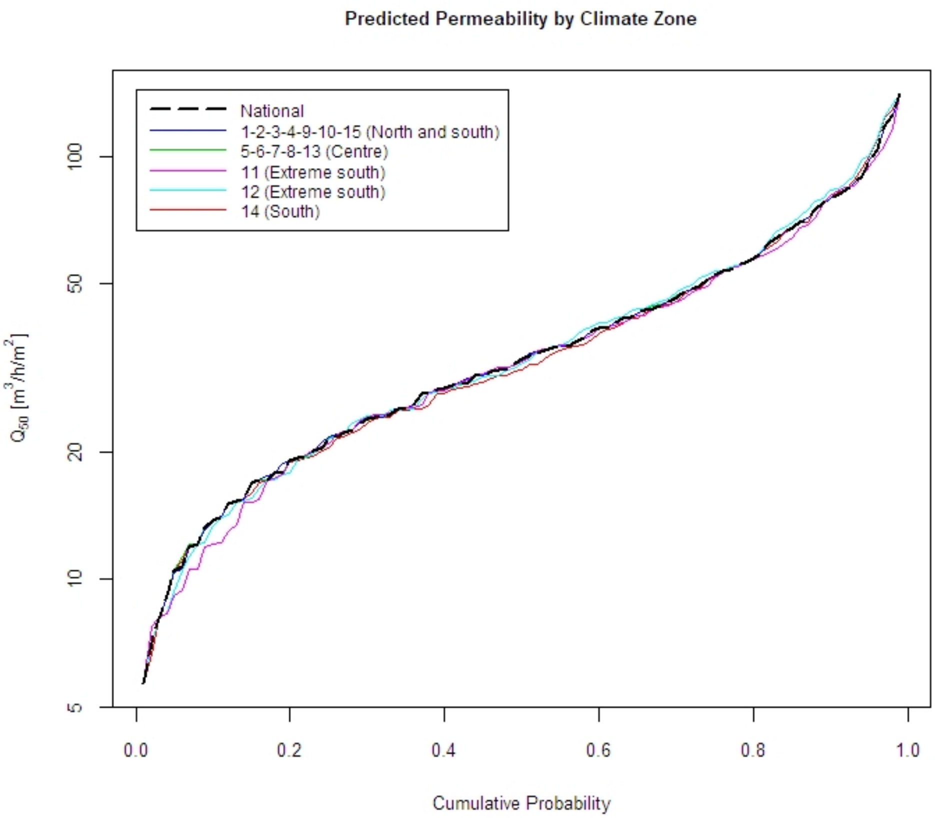
Modelled data
200
150
100
50
0
Residuals [-]
100
60
20
-20
50
100
150
Measured data

(a)

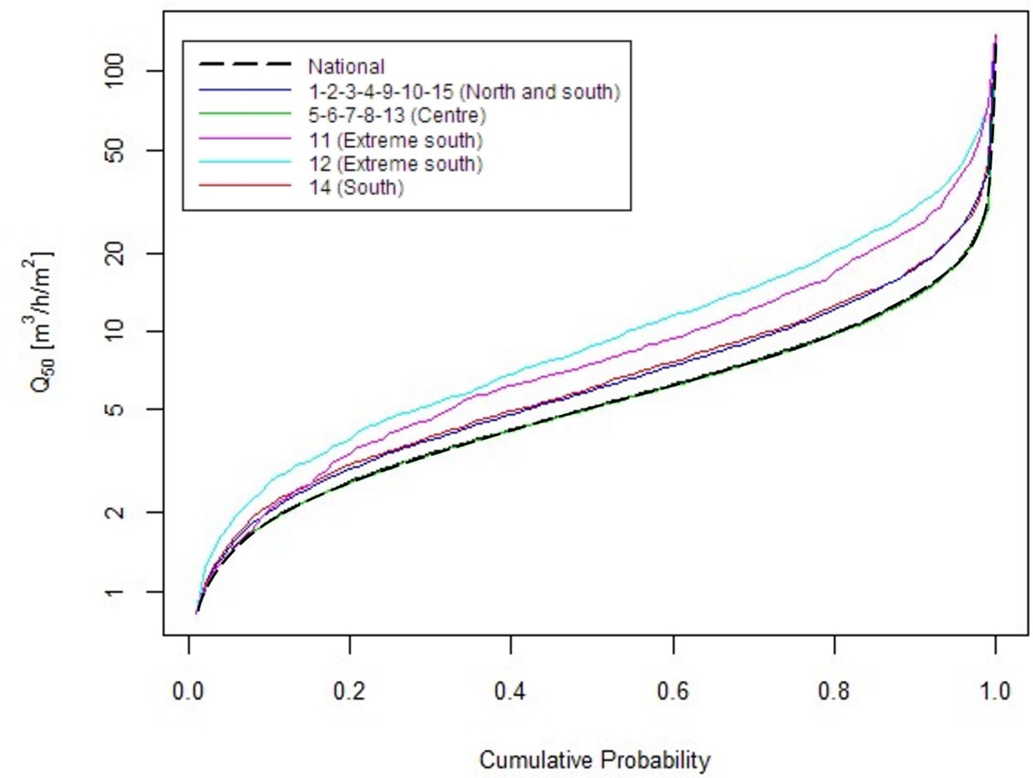


(b)

Figure 4: Predicted normalised leakage (NL) distribution versus empirical data for **(a)** old and **(b)** new Chilean houses. Boxplots show the residuals, with a $\tilde{\mu} = -0.42$ and $\sigma = 6.54$ for old houses, and $\tilde{\mu} = -0.086$ and $\sigma = 1.14$ for new houses.


Predicted Permeability by Climate Zone
National
1-2-3-4-9-10-15 (North and south)
5-6-7-8-13 (Centre)
11 (Extreme south)
12 (Extreme south)
14 (South)
Q_{50} [$m^3/h/m^2$]
5
10
20
50
100
0.0
0.2
0.4
0.6
0.8
1.0
Cumulative Probability

(a)



(b)

Figure 5: Predicted air permeability, $\dot{Q}_{50}$, grouped by climate zone and nationwide for (a) old and (b) new Chilean houses. Legend numbers are geographic regions.

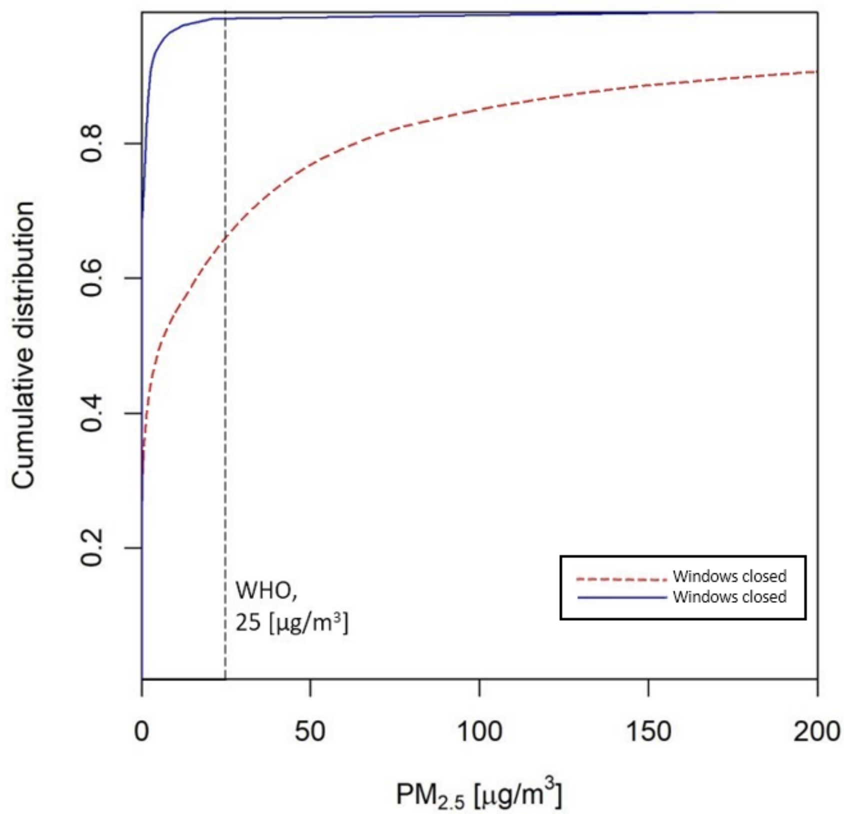
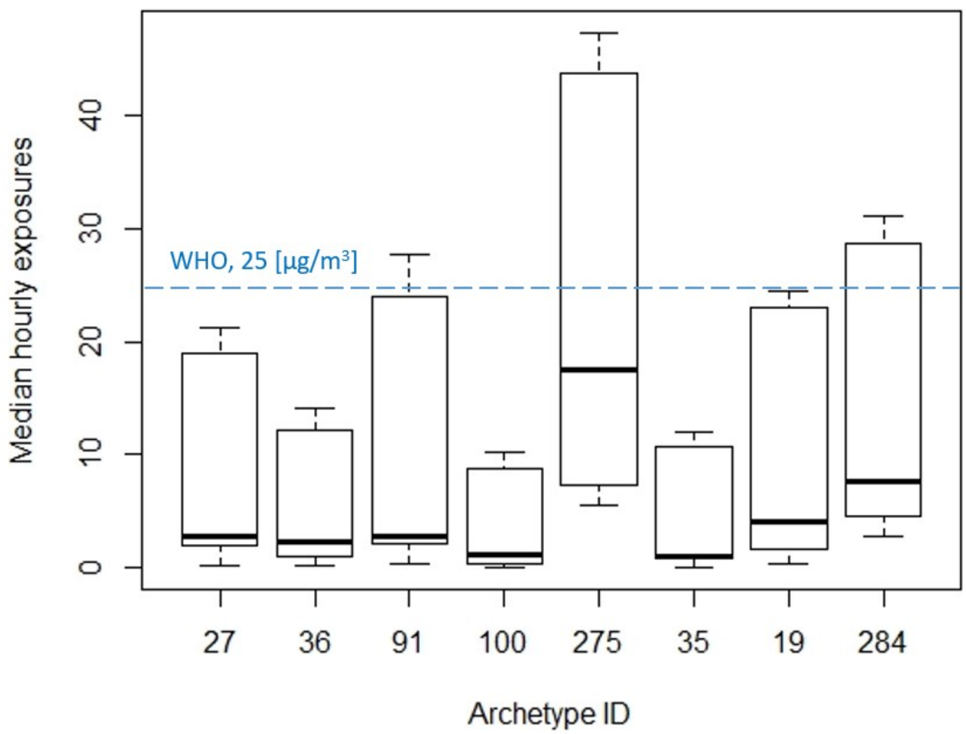
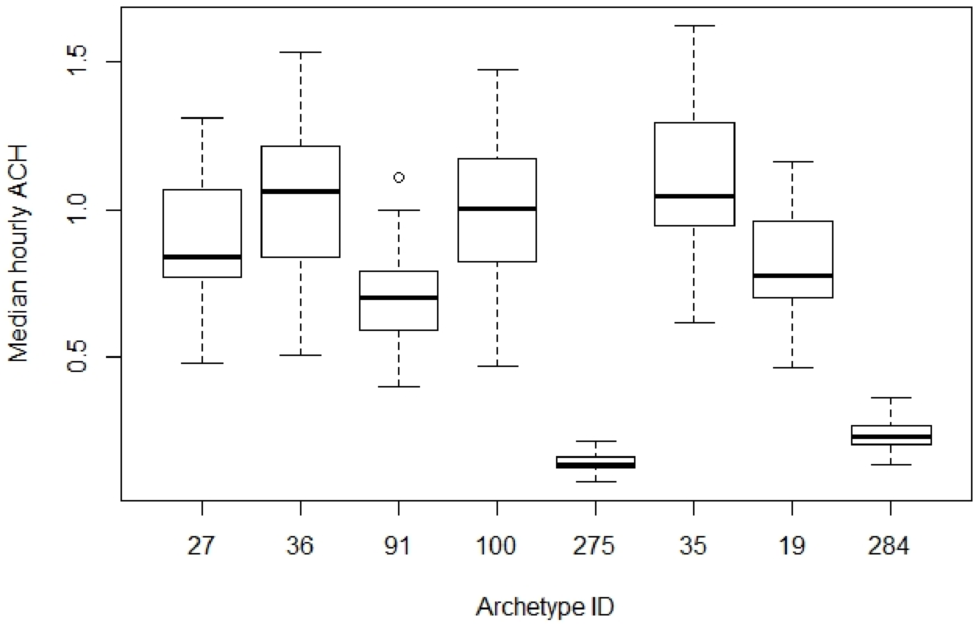


Figure 6: Nationwide daily exposures to $PM_{2.5}$ during the winter season.
Red dashed line, *windows closed* scenario; Blue, *windows open* scenario; Vertical dashed line, the WHO's 24 h guideline value.


WHO, 25 [µg/m³]
Median hourly exposures
0
10
20
30
40
27
36
91
100
275
35
19
284
Archetype ID

(a)



(b)

Figure 7: *Windows closed* scenario. Distribution of the medians of: (a) hourly $PM_{2.5}$ exposures; and (b) median hourly ventilation rates (h^{-1}) of all houses by archetype. Dashed line, the WHO's 24 h guideline value of 25 $\mu g/m^3$. See [13] for archetype details.

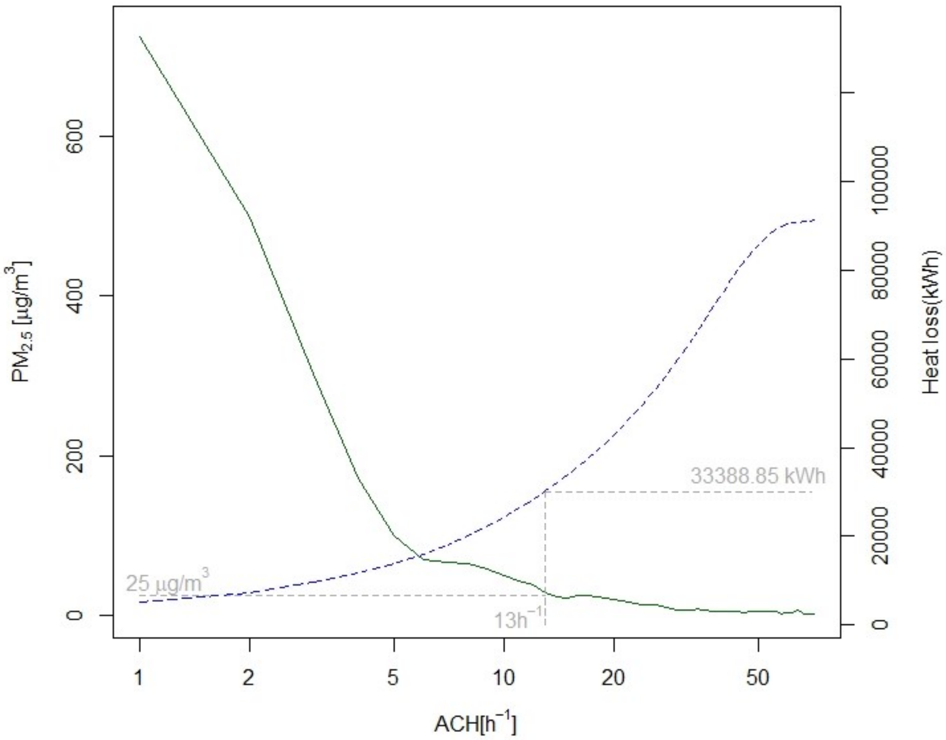


Figure 8: P_{95} of the predicted winter exposures to $PM_{2.5}$ (in green) and the total heat loss (in dashed blue) versus the ventilation rates. Windows closed and open combined. The gray dashed lines show the WHO's annual recommendation of $25\,\mu g/m^3$ and the related heat loss.

References

[1] T. Salthammer, Critical evaluation of approaches in setting indoor air quality guidelines and reference values, Chemosphere 82 (11) (2011) 1507–1517. doi:10.1016/j.chemosphere.2010.11.023.

[2] D. Lader, S. Short, J. Gershuny, The time use survey, 2005: how we spend our time, Report, Office for National Statistics (2006).

[3] T. McCurdy, S. E. Graham, Using human activity data in exposure models: analysis of discriminating factors, Journal of Exposure Science and Environmental Epidemiology 13 (4) (2003) 294.

[4] M. Davies, T. Oreszczyn, The unintended consequences of decarbonising the built environment: A UK case study, Energy and Buildings 46 (2012) 80–85.

[5] G. Sousa, B. M. Jones, P. A. Mirzaei, D. Robinson, A review and critique of UK housing stock energy models, modelling approaches and data sources, Energy and Buildings 151 (2017) 66–80.

[6] T. Fazli, B. Stephens, Development of a nationally representative set of combined building energy and indoor air quality models for us residences, Building and Environment 136 (2018) 198–212.

[7] P. Das, C. Shrubsole, B. Jones, I. Hamilton, Z. Chalabi, M. Davies, A. Mavrogianni, J. Taylor, Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling, Building and Environment 78 (2014) 171–182.

[8] B. Jones, P. Das, Z. Chalabi, M. Davies, I. Hamilton, R. Lowe, A. Mavrogianni, D. Robinson, J. Taylor, Assessing uncertainty in housing stock infiltration rates and associated heat loss: English and UK case studies, Building and Environment 92 (2015) 644–656.

[9] E. Oikonomou, M. Davies, A. Mavrogianni, P. Biddulph, P. Wilkinson, M. Kolokotroni, Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London, Building and Environment 57 (2012) 223–238.

[10] I. Hamilton, J. Milner, Z. Chalabi, P. Das, B. Jones, C. Shrubsole, M. Davies, P. Wilkinson, Health effects of home energy efficiency interventions in England: a modelling study, BMJ open 5 (4) (2015) e007298.

[11] C. O'Leary, B. Jones, S. Dimitroulopoulou, I. Hall, Setting the standard: The acceptability of kitchen ventilation for the english housing stock, Building and Environment (2019) In-doi:https://doi.org/10.1016/j.buildenv.2019.106417.
URL `http://www.sciencedirect.com/science/article/pii/S0360132319306274`

[12] M. Peel, B. Finlayson, T. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. 11 (2007) 1633–1644.

[13] C. Molina, M. Kent, I. Hall, B. Jones, A data analysis of the Chilean housing stock and the development of mod-

elling archetypes, Energy and Buildings 206 (2020) 109568. doi:https://doi.org/10.1016/j.enbuild.2019.109568.

[14] J. M. Logue, P. N. Price, M. H. Sherman, B. C. Singer, A method to estimate the chronic health impact of air pollutants in US residences, Environmental Health Perspectives 120 (2) (2011) 216–222.

[15] C. O'Leary, Y. de Kluizenaar, P. Jacobs, W. Borsboom, I. Hall, B. Jones, Investigating measurements of fine particle ($PM_{2.5}$) emissions from the cooking of meals and mitigating exposure using a cooker hood, Indoor Air (2019).

[16] A. Schueftan, A. D. González, Proposals to enhance thermal efficiency programs and air pollution control in south-central chile, Energy Policy 79 (2015) 48–57.

[17] CASEN, National Socioeconomic Characterization Survey (CASEN) (2015).
URL http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/casen_2015.php

[18] CENMA, Evaluación de impacto atmosférico de sistemas de calefacción domiciliaria (2011).

[19] WHO, WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide, Report, Word Health Organization (2005).

[20] W. S. Dols, B. J. Polidoro, Contam user guide and program documentation version 3.2, Tech. rep., NIST (2015).

[21] W. S. Dols, S. Emmerich, B. J. Polidoro, Coupling the multizone airflow and contaminant transport software contam with energyplus using co-simulation, Building simulation 9 (4) (2016) 469–479.

[22] J. C. Chang, Z. Guo, Modeling of the fast organic emissions from a wood-finishing product—floor wax, Atmospheric Environment. Part A. General Topics 26 (13) (1992) 2365–2370.

[23] L. C. Ng, A. Musser, A. K. Persily, S. J. Emmerich, Airflow and indoor air quality models of doe reference commercial buildings, Gaithersburg, MD, National Institute of Standards and Technology 163 (2012).

[24] W. Chen, A. K. Persily, A. T. Hodgson, F. J. Offermann, D. Poppendieck, K. Kumagai, Area-specific airflow rates for evaluating the impacts of voc emissions in us single-family homes, Building and Environment 71 (2014) 204–211.

[25] A. Bastani, F. Haghighat, J. A. Kozinski, Contaminant source identification within a building: toward design of immune buildings, Building and Environment 51 (2012) 320–329.

[26] S. Yu, L. He, G. Feng, The transient simulation of carbon dioxide emission from human body based on contam, Procedia Engineering 121 (2015) 1613–1619.

[27] L. Underhill, M. Fabian, K. Vermeer, M. Sandel, G. Adamkiewicz, J. Leibler, J. Levy, Modeling the resiliency of energy-efficient retrofits in low-income multifamily housing, Indoor Air 28 (3) (2018) 459–468.

[28] J. Taylor, C. Shrubsole, P. Biddulph, B. Jones, P. Das, M. Davies, Simulation of pollution transport in buildings: the importance of taking into account dynamic thermal effects, Building Services Engineering Research and Technology 35 (6) (2014) 682–690.

[29] C. J. Ferguson, An effect size primer: A guide for clinicians and researchers, Professional Psychology: Research and Practice 40 (5) (2009) 532.

[30] C. Chen, B. Zhao, X. Yang, Y. Li, Role of two-way airflow owing to temperature difference in severe acute respiratory syndrome transmission: revisiting the largest nosocomial severe acute respiratory syndrome outbreak in hong kong, Journal of the Royal Society, Interface 8 (58) (2011) 699–710.

[31] S. Shi, C. Chen, B. Zhao, Air infiltration rate distributions of residences in beijing, Building and Environment 92 (2015) 528–537.

[32] D. Etheridge, M. Sandberg, Building Ventilation: Theory and Measurement, Vol. 1, John Wiley and Sons, 1996.

[33] SEC, Decreto supremo n^o 66 reglamento de instalaciones interiores y medidores de gas. (2007).

[34] B. Jones, P. Das, Z. Chalabi, M. Davies, I. Hamilton, R. Lowe, J. Milner, I. Ridley, C. Shrubsole, P. Wilkinson, The effect of party wall permeability on estimations of infiltration from air leakage, International Journal of Ventilation 12 (1) (2013) 17–30.

[35] M. H. Sherman, D. J. Dickerhoff, Air-tightness of us dwellings, Transactions-American Society of Heating Refrigerating and Air Conditioning Engineers 104 (1998) 1359–1367.

[36] W. R. Chan, J. Joh, M. H. Sherman, Analysis of air leakage measurements of us houses, Energy and Buildings 66 (2013) 616–625.

[37] R. Taylor, Interpretation of the correlation coefficient: a basic review, Journal of diagnostic medical sonography 6 (1) (1990) 35–39.

[38] M. Shipworth, S. K. Firth, M. I. Gentry, A. J. Wright, D. T. Shipworth, K. J. Lomas, Central heating thermostat settings and timing: building demographics, Building Research & Information 38 (1) (2010) 50–69.

[39] Meteonorm, Meteotest (ed.) version 7 ed. switzerland (2017).

[40] M. W. Liddament, A guide to energy efficient ventilation, Air Infiltration and Ventilation Centre Coventry, 1996.

[41] M. V. Swami, S. Chandra, Procedures for calculating natural ventilation airflow rates in buildings, ASHRAE Final Report FSEC-CR-163-86, ASHRAE Research Project (1987).

[42] H. Oezkaynak, J. Xue, R. Weker, D. Butler, P. Koutrakis, Particle team (pteam) study: Analysis of the data. final report, volume 3, Tech. rep., Harvard Univ., Boston, MA (United States). School of Public Health (1996).

[43] P. J. Dacunto, K.-C. Cheng, V. Acevedo-Bolton, R.-T. Jiang, N. E. Klepeis, J. L. Repace, W. R. Ott, L. M. Hildemann, Real-time particle

monitor calibration factors and $PM_{2.5}$ emission factors for multiple indoor sources, Environmental Science: Processes & Impacts 15 (8) (2013) 1511–1519.

[44] C. He, L. Morawska, J. Hitchins, D. Gilbert, Contribution from indoor sources to particle number and mass concentrations in residential houses, Atmospheric environment 38 (21) (2004) 3405–3415.

[45] D. A. Olson, J. M. Burke, Distributions of $PM_{2.5}$ source strengths for cooking from the research triangle park particulate matter panel study, Environmental science & technology 40 (1) (2006) 163–169.

[46] C. O'Leary, Personal communication (2018).

[47] INE, Encuesta nacional sobre uso del tiempo (2016).

[48] J. Milner, C. Shrubsole, P. Das, B. Jones, I. Ridley, Z. Chalabi, I. Hamilton, B. Armstrong, M. Davies, P. Wilkinson, Home energy efficiency and radon related risk of lung cancer: modelling study, Bmj 348 (2014) f7493.

[49] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2018). URL `https://www.R-project.org/`

[50] J. C. Helton, F. J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliability Engineering & System Safety 81 (1) (2003) 23–69.

[51] G. Gigerenzer, Mindless statistics, The Journal of Socio-Economics 33 (5) (2004) 587–606.

[52] N. Fenton, M. Neil, Risk assessment and decision analysis with Bayesian networks, Crc Press, 2012.

[53] A. Field, J. Miles, Z. Field, Discovering statistics using R, Sage publications, 2012.

[54] R. Coe, It's the effect size, stupid, in: Paper presented at the British Educational Research Association annual conference, Vol. 12, 2002, p. 14.

[55] WHO, WHO guidelines for indoor air quality: selected pollutants, Report, Word Health Organization (2010).

[56] ASHRAE, ANSI/ASHRAE standard 62.2 – ventilation and acceptable indoor air quality in residential buildings, Report, The American Society of Heating, Refrigerating and Air–Conditioning Engineers (2016).

[57] W. Ott, A. C. Steinemann, L. A. Wallace, Exposure analysis / edited by Wayne R. Ott, Anne C. Steinemann, Lance A. Wallace, Boca Raton London : CRC Press, Boca Raton London, 2007.

[58] B. Stephens, J. A. Siegel, Penetration of ambient submicron particles into single-family residences and associations with building characteristics, Indoor Air 22 (6) (2012) 501–513.

[59] I. Hughes, T. Hase, Measurements and their uncertainties: a practical guide to modern error analysis, Oxford University Press, 2010.

[60] CDT, Estudio de usos finales y curva de oferta de la conservacion de la energía en el sector residencial. (2010).

[61] C. Molina, A. Jackson, B. Jones, The evaluation of real-time indoor environment parameters measured in 297 Chilean dwellings, in: AIVC 2019: *From energy crisis to sustainable indoor climate. 40 Years of the AIVC.*, 2019, pp. 662–671.

[62] P. Sharpe, B. Jones, R. Wilson, I. C, What we think we know about the aerodynamic performance of windows, Energy and Buildings *In review* (2020).