The Development of Archetypes to Represent the Chilean Housing Stock

Constanza Molina[1], Benjamin Jones[*1], Michael Kent[1] and Ian P Hall[2]

1 Department of Architecture and Built Environment,
University of Nottingham
Nottingham, NG7 2UH, UK.
**Corresponding author:*
benjamin.jones@nottingham.ac.uk

2 Division of Respiratory Medicine,
School of Medicine,
University of Nottingham
Nottingham, NG7 2UH, UK.

ABSTRACT

There are three common methods used to analyse Indoor Air Quality in buildings: in-site measurements, laboratory measurements, or the simulation of indoor spaces using a validated computational model. Each have their advantages, but computational models are generally used to predict air quality in a wide range of indoor environments because they are quick, cheap, and non-invasive. A wide range of inputs are required to accurately simulate airflow and pollutant transport. However, this information may not exist or may only exist in abstract forms. Furthermore, the collation and processing of data can be time consuming and can introduce systematic error when it is undocumented. A documented database containing a range of building archetypes with statistically representative values of related parameters can facilitate the simulation process. The archetypes might then be used to predict and evaluate the impacts of new policies on the indoor air quality of a stock of houses.

This paper describes a method to identify sets of archetypes that are statistically representative of the Chilean housing stock. The Santiago housing stock comprises 41% of the Chilean housing stock and is well documented (55% of all dwellings are surveyed), and so it used to represent the Chilean stock. All available data on the Santiago housing stock, including CENSUS data and annual building statistics reports, are utilised. The archetypes account for elements of building design that affect indoor pollutant concentrations, which can be used for indoor environment modelling. Some similarities are found between houses belonging to Santiago and the rest of country, and although a database that encompassed both Santiago and the remaining stock is desirable, the houses outside Santiago are more difficult to categorise and so an independent analysis is required. Dwellings were categorised according to relevant factors, such as geometry and total floor area, and allocated to weighted groups by dwelling type to form a series of archetypes. Notable architectural elements and values of relevant parameters are used to identify each archetype. These can be used to model the air quality found across the entire Chilean housing stock using validated tools, such as CONTAM or EnergyPlus.

The representative archetypes provide a better understanding of the current Chilean housing stock and will enable the testing of a range interventions designed to improve indoor air quality in Chilean houses. Existing databases are non-exhaustive and contain errors and so knowledge gaps are highlighted. This information can be used to inform future surveys of the Chilean housing stock.

KEYWORDS

Statistical Methods; Indoor Air Quality; Dwellings; Modelling

HIGHLIGHTS

- A list of archetypes is selected to statistically represent the Chilean housing stock.
- 3 sets of archetypes are proposed that are suitable for different studies using 13, 35 or 88 houses to represent 35%, 63% or 91% of the housing stock, respectively.
- Show areas where there is poor quality or a lack of data on the characteristics of Chilean houses, occupants, and their behaviour.

1 INTRODUCTION

Air Quality assessments are either conducted for outdoor ambient air or for indoor air. Air quality in indoor environments depends on a range of factors such as the building's design, use, and environment. Indoor exposures are then affected by three systems, which makes indoor air quality (IAQ) analysis a complex task because they involve: i) a heterogeneous, and sometimes peculiar, building stock; ii) a population with varied characteristics, composition and behaviours; and iii) the presence of mixtures of pollutants that may be difficult to identify and quantify. These three systems are constantly interacting with each other, and their parameters and characteristics need to be considered in order to analyse their relationship in a building stock. When considering different building types, dwellings are a focus of attention because people spend most of their lives in them (McCurdy & Graham, 2003). Also, concentrations of some pollutants indoors are found to be higher than those outdoors (Chen & Zhao, 2011; Cometto-Muñiz & Abraham, 2015), and so there is likely to be a higher degree of exposure to them indoors (e.g. de Bruin *et al.*, 2008; Guo *et al.*, 2004). Therefore, housing and public health interventions that aim to improve indoor air should account for the heterogeneity of houses and the air within them.

Modelling a housing stock is a quicker and more economic method of assessment when compared to *in-situ* measurements. Modelling approaches commonly use a set of reference buildings, known as *archetypes*, which represent classes of houses found in the residential sector. Generally, the larger the numbers of archetypes used to represent the whole stock, the more widespread are the conclusions derived from the results.

The aim of this study is threefold. Firstly, to increase the knowledge of the Chilean housing stock, its composition and characteristics, from an air quality viewpoint; see Section 2.1. Secondly, to consider the variables that should be taken into account by both modelling and field studies so that they can be developed into a useful database; see Section 2.2. Finally, to use the housing classifications to choose a suitable number of archetypes that can be used for further analysis of the housing stock; see Section 2.3. These archetypes can then be used to represent the Santiago housing stock for IAQ and exposure analyses. Although the study is confined to the capital region of Santiago de Chile, it is possible to extend it to cover other locations in Chile.

2 MATERIALS AND METHODS

There are two approaches to modelling building stock energy demand: the bottom-up and the top-down approaches (Kavgic *et al.*, 2010). This study selects a series of archetypes to model the air quality in the Santiago building stock using a bottom-up approach. This allows the modeller to work with disaggregated data, which provides a better understanding of how individual components (e.g., house typology, an architectural element), or changes to them, impact predictions.

This study follows 4 steps defined by Ghiassi and Mahdavi (Ghiassi & Mahdavi, 2017): i) Choose variables that most affect model predictions and their categories, ii) classify dwellings into groups, iii) decide on an appropriate number of dwelling archetypes, and iv) determine the stochastic representation of other properties and components that affect pollutant concentrations and that will allow the re-diversification of the stock. It also considers the development and application of archetypes in other countries for both energy consumption (e.g. Ballarini *et al.*, 2014; Swan & Ugursal, 2009) and air quality-related studies (e.g. Persily *et al.*, 2006; Shi *et al.*, 2015).

2.1 Characteristics of dwellings in Chile and Santiago

Chile is divided into 15 regions located in 7 different climate zones. Its housing stock comprises 6m houses whose characteristics vary according to the local weather conditions and the availability and the affordability of building materials. Chile is over 4,300km long, yet 41% of the population live in the metropolitan and capital region of Santiago de Chile whose stock is estimated to comprise 2.4m houses.

2.2 Sources of information on the building stock

Three databases containing information on the buildings' characteristics are used: CENSUS 2002 (INE, 2003), CENSUS 2012 (INE, 2013), and the Annual Building Permit Reports (INE, 2016). CENSUS surveys are conducted every 10 years, and they collect basic self-reported information on each house and its occupants, hereon termed *cases*. The Annual Reports data is compiled and updated monthly, and contains information about building projects at the design stage.

CENSUS 2002 data provides information on basic geometry (i.e. apartment or house), the number and use of rooms, and information on occupants. It does not include information about the year of construction, the type of house (whether it is detached, semidetached or terraced), the size of the building, or the number of floors. For the purpose of this study, the cases are selected by occupancy (i.e. occupied houses), number of homes (equal to 1), and location (Santiago). By recoding the variables they are easier to control and missing data can be removed, including houses with absent occupants, empty houses (6.6%) and cases with more than 1 home per building (5.8%). The maximum number of occupants and enclosed areas are arbitrarily set to 6 and 7, respectively. The total remaining cases in Metropolitan Santiago is 1,369,314, representing a 68% of the Santiago stock.

Similar information is available from the last CENSUS 2012. However, this database is rejected due to inconsistencies in the data collection methods and to changes in the categories collected (structural differences). A new CENSUS 2017 is ongoing and the data may be available for future updates.

The last and most detailed database is the Annual Building Permit Reports, which contains statistical information of more than 103,146 construction projects or 1,337,639 houses built between 1990 and 2016, and corresponds to 55% of the Santiago stock. This information is reported by the owner and the architect of the project when applying for a construction permit and part of the information is compiled and registered by the National Institute of Statistics. The figures are published in an Annual Report. The data includes the total constructed area, number of floors, building type (detached, semidetached, terraced or apartment), and building materials. Although this database gives more detailed information on the buildings, it contains significant and obvious deficiencies that appear to be transcription errors. Therefore, this database is excluded from the classification process but is instead used to provide other peripheral values, such as building type and year of construction. The remaining cases have been collated and can be used to statistically determine parameters of interest for each archetype, such as floor area, or the number of storeys.

Consequently, the CENSUS 2002 database is used to classify the building stock according to key parameters and to select archetypes, and the Building Annual Reports aggregated databases are used to provide additional information on the archetypes.

2.3 Selection of relevant variables and categorization process

The building design variables used during the categorization process were selected by their relevance to the final indoor pollutant concentration, based on information collected from the literature. Two types of study are considered: those using field measurements and those modelling indoor pollutants. The former uses indoor measurements of concentrations of selected pollutants in order to identify common air quality issues and to relate them to a building's characteristics (Gilbert *et al.*, 2008; Langer & Bekö, 2013; Langer *et al.*, 2016). Some focus on sensitive groups of the population, showing the relationships between health outcomes and indoor variables (Chin *et al.*, 2014; Singleton *et al.*, 2017). The latter model one or more residences that represent the housing stock (A. Persily *et al.*, 2010; Shi *et al.*, 2015).

Studies that use housing stock models to predict energy demand are also included in the literature review due to the similarity of their methods used to analyse data, although there some differences between the key variables that area considered. Studies of IAQ using measurements of indoor pollutant concentrations relate their results to air change rates and the type of ventilation system, the location of the dwelling, the year of construction (correlated with emission rates), pollutant sources and their locations, occupancy levels and behaviour, and environmental parameters (i.e. temperature and relative humidity).

Other modelling studies, especially those investigating air infiltration rates, generally require similar variables: building type, geometry and size, permeability of the envelope, year of construction, number of floors, building orientation (Jones *et al.*, 2015; A. Persily *et al.*, 2010; Shi *et al.*, 2015). CONTAM is a multi-zone airflow and transport pollutant modelling tool (Dols & Polidoro, 2015) widely used to model indoor the pollutant concentrations found in different building types (Ng *et al.*, 2013), the analyses of strategies that might affect IAQ, and the parameters that most affect occupant exposures. It requires the definition of indoor zones and the interactions between them, and outdoor environmental parameters, such as wind speed and direction. These parameters are used to inform inputs to the database.

The resulting data is then used to select key variables for a categorization process based on house geometry or type, building size, and construction period. Ventilation and infiltration in Chilean houses is largely natural via windows and doors opening, and through the envelope. Therefore, mechanical ventilation systems are not expected to be influential and so the type of ventilation is not included as a variable. Following categorization of the dwellings into groups, average values of each group were related to each archetype, such as floor area, number of storeys, and number of occupants (see Table 2).

2.3.1 Geometry

The CENSUS 2002 shows that 95.44% of the dwellings located within the urban area of the capital city are classified as a house or an apartment. Therefore, the housing stock can be divided approximately into those two categories. Other categories, such as the *conventillo* or *ruca* vernacular building types, correspond to building types that were built before 1935 and do not comply with the current building codes. By classifying in this way, the same dataset is then aggregated according to the building size.

2.3.2 Building Size

Dwelling floor area is not included in the CENSUS surveys and so the number of rooms and bedrooms are used to approximate it. The maximum number of zones is arbitrarily set up to a threshold of 7 and the number of bedrooms ranges from 0 (i.e. in cases such as studio rooms) to 6 or 7 in apartments and houses, respectively. This gives 105 separate groups, hereon termed *cells*, where each dwelling is assigned to only one. The buildings are categorised according to their size within each cell, and distributed into smaller sub-cells based on the type of house and construction period.

2.3.3 Type of House

The dwelling type is given by the Building Permit datasets. Three categories can be considered according to the number of attachments or shared walls: i) detached, ii) semi-detached and end-terrace, and iii) mid-terrace. None of the databases include information on the position of a dwelling within a terraced block of houses (e.g. end or mid-terrace) or the number of houses in a block. Therefore, the proportions of end- and mid-terrace buildings are estimated according to the probability of obtaining a given housing type. For example, the maximum number of houses in a block is arbitrarily limited to 20, and when a terrace contains 3 houses, 1 is classed as terraced whereas the other 2 houses are classed as semi-detached.

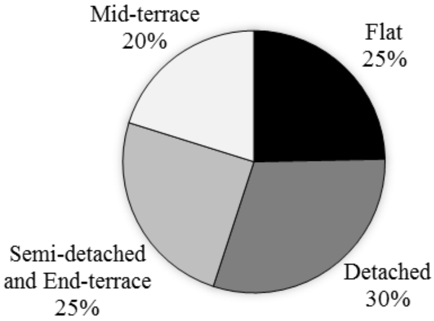

Figure 1: Proportion of apartments and houses in Santiago considered in the classification process.

2.3.4 Construction Period

Table 1: Proportion of dwellings by construction period. Sources: Annual Reports (INE, 2017).

Construction period	Number of Dwellings	Percentage
1990 - 2007	890,889	66.6%
From 2008	446,750	33.4%
Total	**1,337,639**	**100%**

Table 2: Variables considered in the categorization process.

Variables	Categories: Percentage	Source of information	Cumulative Number of cells
Geometry:	Apartments: 22.6% Houses: 77.4%	CENSUS 2002	2
Building size: Number of zones: from 1 to 7 and Number of bedrooms: from 0 (studio) to 6 (apartments) or 7 (houses)	-	CENSUS 2002	105
Type of house:	Detached: 40.92% Semi-detached/ end-terraced: 32.58% Mid-terrace: 26.50%	Building permits	217
Construction period:	1990 - 2007: 66.6% > 2007: 33.4%	Building permits	434
Total cells:			**434**

The envelope structure and design of a house affects its thermal performance and airtightness. In Chile, dwellings built before 2000 had no thermal requirements and are generally considered as non-insulated, although some may have been. Dwellings designed between 2000 and 2007 are required to meet a maximum thermal transmittance for the roof, but not for the walls and windows. After 2007, building codes were upgraded and all new dwellings must meet a maximum thermal transmittance for all envelope components in contact with the ambient air. The requirement varies along the length of the country depending on the number of heating degree days. New changes to the building codes are about to be adopted, and so the database will need to be updated in the near future and new archetypes may be required.

In this study, the housing stock is divided into two age-related groups: those constructed before 2008 when little or no insulation was required by law, and those built thereafter with high levels of insulation; see Table 1. Houses constructed before 2008 that have been weatherised are not included. The Annual Reports are used to obtain the proportion of the stock belonging to each group because they contain the total number of dwellings registered per year (INE, 2017).

2.3.5 Floor Area, Number of Storeys, and Number of Occupants

These values are given by the Building Permits aggregated database. The number of zones reported in the database can be associated with those of each archetype, and the mean calculated from each subset. The number of storeys and the number of occupants can be determined by the mode value in each cell. The relevant factors considered for the grouping process are given in Table 2.

3 RESULTS AND DISCUSSION

3.1 Selection of Archetypes

Each house is classified into one of 434 archetypes and the most representative can be selected using a range of statistical tests. A frequency plot can be used to rank the number of times each archetype appears in the database from highest to lowest; see Figure 2.

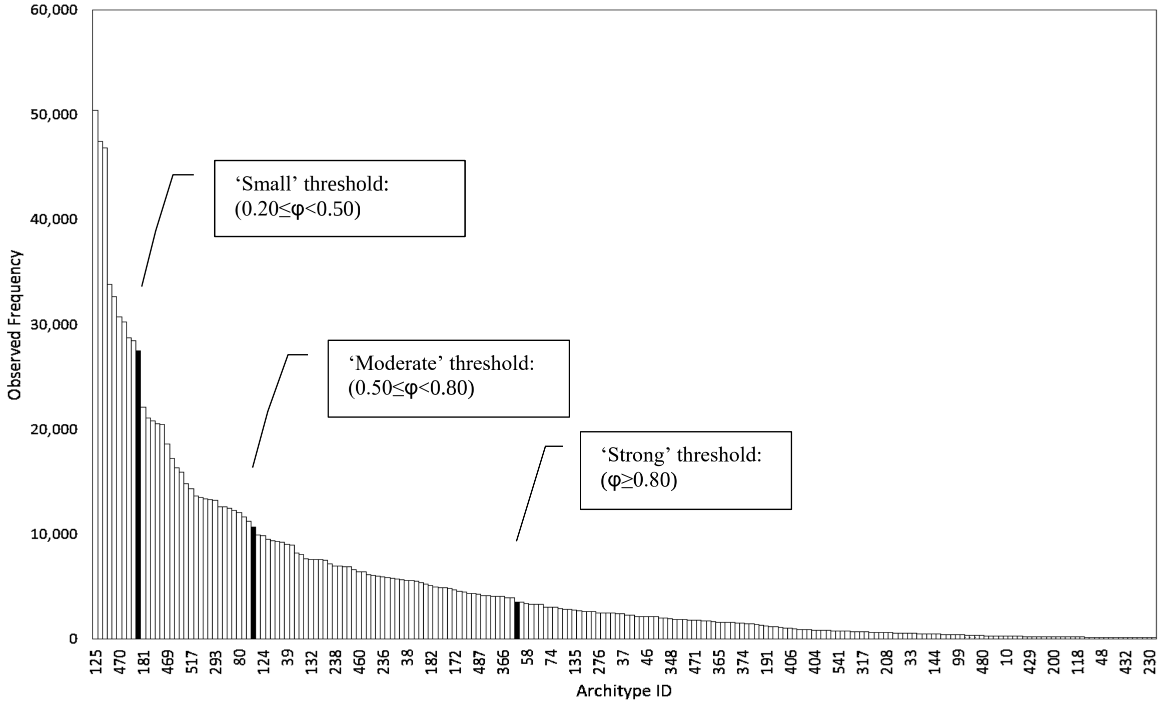

Figure 2: Frequency distribution of the archetypes.

Figure 2 shows that the archetype with the highest frequency has an ID of 125. Chi-squared (χ^2) tests of statistical significance can be performed to compare the difference in frequency of each archetype to a lower ranked archetype (Field, 2013). For each comparison, effect sizes (φ) are used to obtain a standardised measure of the difference that is independent from the sample size (Cohen, 1990, 1992). Estimations of effect sizes across the whole stock can be quantified using the tables of Ferguson (2009) to show those considered be *small* ($0.20 \leq \varphi < 0.50$), *moderate* ($0.50 \leq \varphi < 0.80$) and *strong* ($\varphi \geq 0.8$) in magnitude, and are given in Table 3. Therefore, this approach accounts for both descriptive (percentages) and inferential (effect size) statistics. By comparing each archetype against the highest ranked archetype, those that have an effect sizes smaller than a particular threshold can be considered representative and those with larger differences can be excluded. Figure 3 shows the number of archetypes required to represent each centile of the building stock for each effect size threshold.

Table 3: Effect size of the archetypes classification. Effect size moderate or greater than 0.5 was considered relevant to this study.

Ferguson's benchmark for Effect Sizes	Number of archetypes (retained)	Percentage of the stock represented
Small: $\varphi = 0.2$	13	35.3%
Moderate: $\varphi = 0.5$	35	62.8%
Strong: $\varphi = 0.8$	88	90.5%

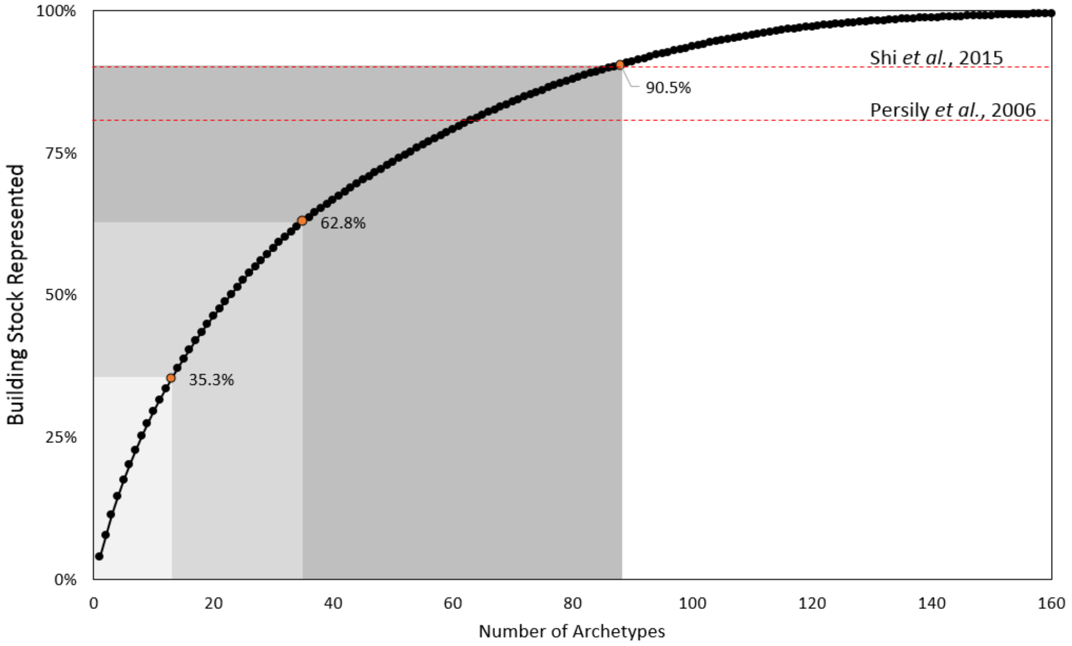

Figure 3: Cumulative distribution of the archetypes according to the percentage of the housing stock represented by the effect size thresholds and two thresholds obtained from the review of the literature (dashed lines).

Figure 3 shows that 99.5% of the Santiago de Chile housing stock can be represented by 160 archetypes. However, limiting factors, such as time and effort, may require a more expedient number of archetypes. The gradient of the line in Figure 3 decreases as the number of archetypes increases, and so there is a law of diminishing returns where adding more archetypes to a set does not significantly increase the proportion of the building stock represented. Therefore, Figure 3 can be used to choose an appropriate number of archetypes for any given modelling task. Table 3 and Figure 3 show that an effect size of *moderate* magnitude ($0.50 \leq \varphi < 0.80$) requires the first 35 archetypes given in Figure 2 to represent 63% of the stock. Table 4 shows the characteristics of these 35 archetypes. Similarly, an effect size of *small* magnitude ($0.20 \leq \varphi < 0.50$) requires the first 13 archetypes.

Table 4: Characteristics of the selected 35 archetypes.

Variable	Category, Value	Frequency
Type of building	Apartment	10
	Detached	10
	Semidetached / End-terrace	8
	Mid-terrace	7
Number of Zones	3-6	
Number of Bedrooms	1-3	
Age period	Before 2008	24
	From 2008	11

3.2 Examples of Archetype Parameters

Table 5 presents the first 13 ranked archetypes identified in Figure 2, which represent 35.3% of the housing stock of Santiago de Chile. The ID Number indicates the initial position of each archetype among all 434. No statistical data on individual room heights are available to compute volumes, and so they are obtained from building codes, which specify a minimum of 2.4m.

Table 5: Architectural elements and values of relevant parameters to be used when modelling the archetypes values obtained either from the databases or assumed from other sources.

ID Number	Weighting factor	Construction Period	Type of dwelling	Number of Zones	Number of Bedrooms	Mean floor area [m^2]	Number of storeys
125	48338	1	D	4	2	66.58	1
134	45455	1	D	5	3	150.56	1
24	44917	1	A	4	2	70.32	4-storey building
237	38490	1	S+E	4	2	55.05	1
246	36195	1	S+E	5	3	103.53	2
32	32387	1	A	5	3	104.33	4-storey building
349	31312	1	M	4	2	50.64	1
358	29444	1	M	5	3	146.79	1
133	27545	1	D	5	2	150.56	1
31	26356	1	A	5	2	104.41	4-storey building
181	24242	2	D	4	2	70.37	1
190	22796	2	D	5	3	157.59	1
73	22526	2	A	4	2	73.08	4-storey building

D: Detached; **A**: Apartment; **S+E**: Semidetached and End-terrace; **M**: Mid-terrace.
1: built between 1990-2007 period; **2**: Built during or after 2008.

4 CONCLUSIONS

This paper uses available data contained within national censuses and building reports to identify sets of archetypes that are statistically representative of the housing stock of the metropolitan and capital region of Santiago de Chile. This is considered a proxy of the entire Chilean housing stock because 41% of the population is located there. The proportion of the stock represented increases with the number of archetypes, but there is a law of diminishing returns. Therefore, the number of archetypes can be chosen based on the research question and available resources.

The descriptive parameters of each archetype are informed by the elements of building design recorded by IAQ field studies and required as inputs to energy, ventilation, and pollutant transport modelling tools. Therefore, they can be used to model indoor environments and to test a range interventions designed to improve IAQ in Chilean houses.

The archetypes increase our understanding of the Chilean housing stock, highlight its heterogeneity, and show all available data on their characteristics. Existing databases are non-exhaustive and contain errors and so areas are highlighted where there is poor quality, or a lack of data, on the characteristics of houses, and occupants and their behaviour. This information can be used to inform future surveys of the Chilean housing stock.

Both the database and the archetypes will be shared in a public domain in the future.

5 REFERENCES

Ballarini, I., Corgnati, S. P., & Corrado, V. (2014). Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy Policy, 68*, 273-284.

Chen, C. & Zhao, B. (2011). Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor. *Atmospheric Environment, 45*(2), 275-288.

Chin, J. Y. *et al.* (2014). Levels and sources of volatile organic compounds in homes of children with asthma. *Indoor Air, 24*(4), 403-415.

Cohen, J. (1990). Things I have learned (so far). *American psychologist*, *45*(12), 1304-1312.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 115-159.

Cometto-Muñiz, J. E., & Abraham, M. H. (2015). Compilation and analysis of types and concentrations of airborne chemicals measured in various indoor and outdoor human environments. *Chemosphere, 127*, 70-86.

de Bruin, Y. B. *et al.* (2008). Characterisation of urban inhalation exposures to benzene, formaldehyde and acetaldehyde in the European Union. *Environmental Science and Pollution Research, 15*(5), 417-430.

Dols, W. S., & Polidoro, B. (2015). *CONTAM User Guide and Program Documentation*.

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.

Ghiassi, N., & Mahdavi, A. (2017). Reductive bottom-up urban energy computing supported by Multivariate Cluster Analysis. *Energy and Buildings, 144*(5), 372-386.

Gilbert, N. L., Guay, M., Gauvin, D., Dietz, R. N., Chan, C. C., & Lévesque, B. (2008). Air change rate and concentration of formaldehyde in residential indoor air. *Atmospheric Environment, 42*(10), 2424-2428.

Guo, H., Lee, *et al*. (2004). Risk assessment of exposure to volatile organic compounds in different indoor environments. *Environmental Research, 94*(1) 57-66.

INE. (2003). Censo de Población y Vivienda 2002. Santiago de Chile: Instituto Nacional de Estadísticas.

INE. (2013). Censo de Población y Vivienda 2012. Santiago de Chile: Instituto Nacional de Estadísticas.

INE. (2016). Bases Edificación. Santiago de Chile: Instituto Nacional de Estadísticas.

INE. (2017). Número de viviendas autorizadas total país y región metropolitana (Jan 2017 ed.). Santiago de Chile: Instituto Nacional de Estadísticas.

Jones, B. *et al.* (2015). Assessing uncertainty in housing stock infiltration rates andassociated heat loss: English and UK case studies. *Building and Environment, 92,* 644-656.

Kavgic, M. *et al*. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment, 45*(7), 1683-1697.

Langer, S., & Bekö, G. (2013). Indoor air quality in the Swedish housing stock and its dependence on building characteristics. *Building and Environment, 69*, 44-54.

Langer, S., Ramalho, O., Derbez, M., Ribéron, J., Kirchner, S., & Mandin, C. (2016). Indoor environmental quality in French dwellings and building characteristics. *Atmospheric Environment, 128,* 82-91.

McCurdy, T., & Graham, S. E. (2003). Using human activity data in exposure models: Analysis of discriminating factors. *J Expo Anal Environ Epidemiol, 13*(4), 294-317.

Ng, L. C., Musser, A., Persily, A. K., & Emmerich, S. J. (2013). Multizone airflow models for calculating infiltration rates in commercial reference buildings. *Energy and Buildings, 58*, 11-18.

Persily, A., Musser, A., & Emmerich, S. (2010). Modeled infiltration rate distributions for US housing. *Indoor Air, 20*(6), 473-485.

Persily, A. K., Musser, A., & Leber, D. D. (2006). *A collection of homes to represent the US housing stock*: US Department of Commerce, Technology Administration, National Institute of Standards and Technology.

Shi, S., Chen, C., & Zhao, B. (2015). Air infiltration rate distributions of residences in Beijing. *Building and Environment, 92*, 528-537.

Singleton, R., Salkoski, A. J., Bulkow, L., Fish, C., Dobson, J., Albertson, L., . . . Ritter, T. (2017). Housing characteristics and indoor air quality in households of Alaska Native children with chronic lung conditions. *Indoor Air, 27*(2), 478-486.

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews, 13*(8), 1819-1835.