A DATA ANALYSIS OF THE CHILEAN HOUSING STOCK AND THE DEVELOPMENT OF MODELLING ARCHETYPES

Constanza Molina[1,2]

Michael Kent[3]

Ian Hall[4]

Benjamin Jones[*1]

1. Department of Architecture and Built Environment, University of Nottingham, Nottingham, NG7 2RD, United Kingdom
2. Escuela de Construcción Civil, Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860, Santiago, 7820436, Chile
3. Berkeley Education Alliance for Research in Singapore Limited, Singapore
4. Division of Respiratory Medicine, School of Medicine, University of Nottingham, Nottingham, NG7 2UH, United Kingdom

*Corresponding author: benjamin.jones@nottingham.ac.uk

ABSTRACT

Chile is a South American country experiencing greater social and economic development than the majority of its neighbours and whose economy is *transitioning* from *developing* to *developed* status. It is committed to reduce its greenhouse gas (GHG) emissions by 35-45% by 2030, requiring national energy demand reduction. Chile's housing stock is responsible for 15% of its total final energy consumption and so its Government is regulating the construction of new dwellings, although it is difficult to know if this policy will succeed or lead to unintended consequences affecting energy demand, GHG emissions, and occupant health. Measuring the effects *in situ* is often time and cost prohibitive and so the simulation of archetypal buildings is a common method of investigation. A range of data sources, such as censuses and building permits, are used to develop archetypal Chilean houses with statistically significant representative values for design parameters related to energy demand and indoor air quality. It finds that 496 archetypes can represent 100% of the Chilean housing stock and only 90 can represent 95% of the stock. The archetypes can be used to predict and evaluate the impacts of policies on indoor air quality and the energy demand of a stock of houses, or to guide future data gathering exercises. The data analysis highlights knowledge gaps in categorical descriptors and occupant behaviours, and poor granularity of physical data. These gaps should be filled by augmenting national surveys and complimented by field work.

HIGHLIGHTS

- Archetypal Chilean houses are developed for energy and air quality modelling
- 496 archetypes represent 100% of the housing stock
- 90 archetypes represent 95% of the stock
- Future data gathering is required to fill knowledge gaps

KEYWORDS

indoor air quality, energy demand reduction, carbon emissions, climate change, regulation, housing typologies

1 INTRODUCTION

Chile is a South American country with a population of 17.6 million people [1] and a population growth rate of <1% [2]. In 2016, Chile's economy was ranked the 42nd largest in the world and considered *high income* by the World Bank [2]. Between 2003 and 2013, it experienced real economic growth at an average annual rate of 4.65%, attributable to "a sound institutional framework, countercyclical fiscal policy, tight monetary policy, and integration into the global economy" [3]. Over the same period, the carbon intensity of its economy (measured in tonnes of CO_2 equivalent per million US dollars of GDP) decreased from 198 tCO_2e/mUSD to 81 tCO_2e/mUSD [4], although latterly the decline has ceased in contrast to the decreasing trends of many International Energy Agency (IEA) member countries [5]. This is attributed to the increasing carbon intensity of electricity generation, which surpassed the IEA average in 2011. Over the same period, the proportion of imported energy has remained steady at around 65% [57] indicating that Chile has not actively sought to outsource carbon emissions by importing energy. Chile imports the vast majority of its oil and petroleum, and the coal and natural gas used to generate electricity [58]. The largest consumer of electricity is the Industry sector (62.6% of the total consumption) followed by the Transport (17.6%) and Residential (17.4%) sectors. The same three sectors are also responsible for the majority of the nation's *Total Final Consumption* (i.e., the aggregate of all of the energy used to provide energy services): Industry (43.0%); Transport (33.6%); and Residential (15.4%). Here, the Residential sector's end-uses are defined as space heating and cooling, water heating, lighting, cooking, appliances, and any other energy use [6].

Chile's recent social and economic development no longer qualify it for development aid from the Organisation for Economic Co-operation and Development's (OECD) Development Assistance Committee [7]. With Uruguay it is one of only two South American countries that could be considered *transitional*. When compared to other South American countries, Chile's more advanced economy and ongoing development makes it a case-study that they can learn from, and so is worthy of study.

In January 2017, Chile ratified the Paris Agreement and set two energy-related targets as part of its Nationally Determined Contribution: these were (i) to reduce the carbon intensity of its economy to 30% below its 2007 level by 2030, and (ii) to reduce all green-house-gas (GHG) emissions by approximately

35 to 45%. To achieve this reduction new mitigation measures are defined for the residential sector: (i) to increase the number of thermal solar panels for dwellings, (ii) to update thermal regulations to reduce stock energy demand to a 70% by 2025, (iii) to promote the national the Energy Qualification for Dwellings scheme, (iii) to develop Sustainable Construction Standards and a Sustainable Construction label and (iv) to continue the weatherisation subsidises for insulating dwellings that do not fulfil existing thermal requirements to reduce heating demand and to reduce the risk of internal condensation.

Whilst pragmatic policies designed to reduce national energy demand and GHG emissions might first be directed at the Industry and Transport sectors because their energy demands and emissions are the most significant, the IEA [5] believes that building energy codes and standards are “essential to improve their energy performance and comfort”, and to provide long term energy savings and health benefits for householders. The first statutory energy codes for buildings were established in 1929, but the most significant regulations have been implemented since Chile’s return to a democratic political system in 1990 following a period of autocracy. These regulations solely concern the housing stock, and energy efficiency standards now exist for social and private dwellings, applying to around 80% and 20% of all newly constructed dwellings respectively [5]. However, the IEA [5] notes that compliance is currently only verified at the design stage and not following construction. Furthermore, many of the regulations and standards are duplicates of those used by other countries and have not been specifically developed for the Chilean housing stock and climates. Accordingly, it is difficult to know whether they will achieve their aims, or if there will be any unintended consequences that affect energy demand, GHG emissions, and occupant health [8]. To ameliorate these issues, a full understanding of a dwelling’s energy and ventilation systems is required, which can be scaled up to consider a stock of dwellings [9]. Ideally, data for existing dwellings should be obtained from large-scale field studies, but a modelling approach is required when this is time or cost prohibitive, or when modelling future interventions. Such approaches have been used by many countries to estimate changes in the energy demand and indoor environment quality attributable to one or more policies in both individual houses and housing stocks [9, 10]. There are neither previous analyses of the Chilean housing stock nor any significant modelling programs, and so Chile can learn from the data gathering and modelling successes and failures elsewhere. In time, the techniques applied to analyse Chilean houses could be implemented by other South American countries.

The United Kingdom (UK) has a 25 year history of building models to support policy development and so it is used for comparison herein. Its models of dwelling energy and ventilation systems are summarised by Sousa *et al.* [9] – who show that they are generally categorised as *top-down* or *bottom-up*, working at aggregated and disaggregated levels, respectively. For example, a top-down energy demand model might predict the changes in demand from houses associated with a number of factors, such as variations in weather or energy price, but it cannot explain those changes in detail. However, a bottom-up model might use empirical data from surveys to describe each component in a physical model of a dwelling and so could be used to assess the impacts of a policy on each component [11]. Both methods can be sub-categorised by their application of statistical and physical modelling methods. The majority of UK housing stock energy models (HSEMs) are bottom-up simplified steady-state models of physical phenomena. Many suffer from over-simplification, a lack of model and data transparency, and poor modularity [9]. However, other recent models have applied advanced sampling methods that generate distributions of predictions and quantify the uncertainty in them, and global sensitivity analyses that identify key drivers that can be targeted for remediation or can be the subject of future field surveys; see [12, 13].

A wide range of inputs are required to accurately simulate heat and mass transfer in houses. However, this information may not exist or may only exist at an aggregated level. Furthermore, the collation and processing of data can be time consuming and can introduce systematic errors, such as translational and typing errors, when they are not tracked. Therefore, this paper seeks to develop a documented database for Chile containing a range of building archetypes with statistically representative values of design parameters related to energy demand and indoor air quality (IAQ) that can facilitate the simulation process. The archetypes can then be used to predict and evaluate the impacts of new policies on the energy demand, GHG emissions, air quality, and occupant health for stocks of houses. We also highlight areas of data paucity, thus informing future surveys and research.

2 SOURCES OF DATA

This section identifies available sources of data, relevant for modelling buildings, held by the various Ministries and Institutions of the Chilean government, or relevant non-governmental organisations. We searched for available data to ensure we included all relevant parameters required for modelling energy and mass transport in houses. They include population scale censuses carried out at large intervals, a tri-annual socioeconomic survey of a sample of the population, information on all newly constructed houses, a nationwide network of measurements of indoor environment parameters made in houses, and weather data that represents national climate variations.

Table 1 shows the three most relevant data sources used in the categorisation process, the survey frequency, the size of the dataset, and the resolution of parameters. They are also described in Sections 2.1-2.3. In addition, four other sources are described in Sections 2.4-2.7, which are used to provide supplementary information on the current status of the stock, its residents, and environmental conditions, and are given in the Supplementary Materials; see Section 4.1. The variables are selected because they are thought to be the most relevant for understanding IAQ and energy performance issues. Other variables are also given in the Supplementary Materials.

2.1 National Census

A census aims to record information about an entire population rather than a sample of it, and so it is an accurate and valuable source of data. The Chilean Population and Housing Census has been conducted approximately every 10 years since 1952. The data is summarised for public dissemination by the National Institute of Statistics (INE) and the raw data is available upon request. We analyse both the summary documents and the raw data from the 2002 and 2012 censuses [14, 15], and Table 2 gives the total entries for each. The analysis excludes 0.7m unoccupied and empty dwellings because their data was not recorded. However, it includes occupied dwellings with absent residents during the census.

Table 2 shows that the 2002 census has higher coverage than the 2012 census and the percentage of omissions increased from 3.8% to 9.6%. There are also significant differences in the methods and questionnaires used by the two censuses. Therefore, a technical audit [16] was carried out immediately after the 2012 census to analyse the data and evaluate whether it achieved its aim. Several inconsistencies were identified, such as significant changes in the gender ratio that was inconsistent with birth and death rates, which suggested that the fieldwork was deficient and the data is of limited use [16]. A further audit found the 2012 census to be unhelpful in developing new public policies or for estimating the demand on public services. To understand the evolution of the housing stock, it is important to able to compare censuses to identify changes. However, the two censuses cannot be merged because of differences in the categorical data recorded and types of questions included for some variables. For example, in the 2012 census the categories for the *type of building* variable were *flat/apartment with elevator*, *flat/apartment without elevator*, *detached*, *attached* (semi-detached or terraced), or *other*, whereas the 2002 census used categories of *flat/apartment* or *house*. Herein, the term *apartment* is used. Furthermore, the 2012 census asks fewer questions about the number of rooms in the dwelling and their use. Consequently, the 2002 census is considered to be the more robust of the two. A new census was conducted in 2017 to correct the problems with the 2012 census, although the data is not yet publically available. This new census used categories and questions that were similar to those used in 2002 and so the analyses done with these data may be replicated.

The censuses ask a number of questions that help to characterise each dwelling and household and are useful for understanding the stock (see Sections 3) or for modelling it (see Section 4).

For each dwelling the number of rooms and their use, the presence of bathrooms (including the number of showers) and a kitchen, the dwelling type (*house*, *apartment/flat* or *other*), and services present (boiler, water supply, sewage, heating system (only in the 2012 census), photo-voltaic solar panels, and appliances, such as a cooker (with fuel type), washing machine, or fridge) are all recorded. All Chilean censuses use the *de facto* definition of residency so that all occupants of a dwelling are interviewed at the time of the census. Occupants identified as guests are counted in their own household rather than the surveyed household. If a dwelling is empty, it is considered as *occupied but absent*. In addition to the number of people, their gender and age, employment status, general health (including disabilities) are also recorded. Social-economic index (i.e., using a scale ranging from one for the first decile or most vulnerable houses to 10 corresponding to the most wealthy) is assigned to each household as a function of their education and employment, the properties of their house, and their appliances. Regional information is included so that the distribution of the population and housing around the country are known thus helping to government to provide adequate utilities and public services.

2.2 Building Permits

More detailed building information is contained within the Building Permits' datasets [17]. This data has been collected since 1929 in Santiago and since the 1980s for the rest of the country. The raw dataset obtained for this study only covers the period between 1990 and 2015 for the entire country and provides information on 3.1m new houses, approximately 54% of the current national housing stock. Information about a proposed building is submitted to the local government (municipality) when requesting a construction permit. It is then forwarded to the INE, which compiles and publishes summary statistics annually.

Although this dataset provides additional information, some values of physical dimensions suggest that either the data gathering or the transcription processes introduced errors. Therefore, the modeller must decide whether to exclude these data from the study or to clean the data set. In this study, all the questionable data was removed and the variables were re-coded so that they can be used in conjunction with other databases. Approximately 2% of records were removed so that the number of dwellings in the Santiago Metropolitan region after cleaning was 1,278,328, which corresponds to 57% of the current housing stock.

2.3 National Socioeconomic Characterisation Survey

The National Socioeconomic Characterisation Survey (CASEN) is a cross-sectional survey of education, employment, income, and health status that has been conducted by interview every two or three years since 1987 by the Ministry of Social Development. In 2015, the sample size was 83,887 households and 266,968 people living in 15 regions of the country, and is considered nationally representative [18, 19]. Of particular interest is its reporting of dwellings (housing materials and conditions), the household (family size, number of children, education, employment, family composition, hours worked, income, living conditions, cooking fuels, and household air pollution), utilities (electricity, household heat, and water supply), and occupant health (general health status and mobility, and instances of asthma, chronic obstructive pulmonary disorder, cancers, and ischemic heart disease).

2.4 Use of Time National Survey

The way people behave in a dwelling is usually more significant in determining energy demand than either the size of the dwelling or the household size [20]. Accordingly, it is important to understand how they use their time. The Use of Time National Survey (ENUT) database is a nationwide cross-sectional study. The 2015 dataset comprises a representative and randomly sampled group of 15,312 dwellings (34,575 persons >12 years of age) located in urban areas. The sample represents the housing stock at a regional level, and is sampled from cities where 85% of the Chilean population lives using a self-reported questionnaire [21]. A limited number of activities and their duration during one normal week and weekend day are recorded. It also surveys socioeconomic status and derives quintiles. However, personal location is not recorded and must be assumed by an activity. For example, the nationwide average time people spend at home on a working day is 14.0 hours (58.2% of the day), the time *in-transit* is 2.5 hours (10.4%) and the time at work is 8.2 hours (34.2%). The survey only focusses on an activity and so the duration does not add up to 24 hours and the percentages to 100%. Time activity data can be further aggregated so that specific activities can be related to a room, such as *cooking* to the kitchen or *sleeping* to the bedroom, so that the total time spent in each room can be calculated. For example, questions specifically survey the time spent preparing and heating meals. Nationally, cooking activities have a mean duration of 1 hour and 6 minutes on weekdays and 1 hour and 12 minutes at weekends. This information can be used to schedule pollutant and moisture emissions in kitchens. Two further questions survey sleeping durations for weekdays and weekends, and can be used to determine occupancy. The total time ratio of a family cook spending time in the kitchen, bedroom, and family room is 10:38:52, respectively. This ratio can be used to inform data analyses of modelling or measurement studies; see [10].

Longitudinal data is required to understand how personal activities vary throughout a year and the determinants or explanatory variables that affect decisions. If the season, location, and weather conditions are given, it is possible to describe occupant profiles more accurately and thus estimate more realistic exposures to pollutants and their impacts on health. Therefore, in the short term, the cross-sectional ENUT data must be adapted to infer longitudinal activity patterns that can be used by modellers [22].

2.5 National Housing Monitoring Network

The Ministry of Housing, known colloquially as the Ministerio de Vivienda y Urbanismo (MINVU), is running an ongoing program of *in situ* indoor and outdoor environment monitoring in houses since May 2017, known as the National Housing Monitoring Network or colloquially as the Red Nacional de Monitoreo de Viviendas (RENAM) [23]. RENAM has installed real-time sensors in 299 houses (all concurrently) located in the north ($N = 28$), south ($N = 60$) and centre ($N = 211$) of the country in five cities. Most ($N = 150$) are located in the metropolitan and capital region of Santiago de Chile, which is inhabited by 41% of the population and whose stock is estimated to comprise 2.4 million houses [14]. The monitoring is preceded by a survey that provides information about the location (region, city, and commune), the dwelling (type, construction year, storeys, floor area, orientation, envelope materials, number of windows, glazing properties, and heating system), the householders (income, energy bills, health characteristics issues), and behaviours (heating months, weekday and weekend occupancy, smoking). The platform is accessible online [23] to registered users and contains the approximate location of each sensor and all surveyed and measured data.

Three different devices are used. The first is the *Netatmo* weather station [24], which comprises indoor and outdoor modules. Both record air temperature and relative humidity, but the indoor module also measures sound level, and CO_2. The majority of sensors are located in the living room, although some have been placed in a bedroom, and measurements are time-averaged over 30-minute periods. The second is a Plantower [25] which is capable of measuring the concentration of particles in air with a diameter of 0.3-2.5μm (known as $PM_{2.5}$), and the third is a Wenu Work smart meter that records the electricity demand in houses that use electric heaters exclusively ($N = 50$).

There are known limitations of *consumer-grade* IAQ monitors; see [26]. All of the sensors require annual calibration. Most optical CO_2 sensors are prone to zero-drift and so the Netatmo periodically self-calibrates using a predefined ambient concentration when the measured concentration is both low and steady. Therefore, its readings are imprecise and heavily dependent on the differences between the actual ambient, defined calibration, and measured concentrations. There are no studies quantifying the deterioration in accuracy of low-cost $PM_{2.5}$ sensors over time, but their precision depends on the difference between the size fractions they are sensitive to and those emitted by a source. This is compounded by differences in the refractive index of particles from varying sources, and so for accurate measurements a source must be identified and a calibration factor applied, which can range from 0.016 to 12 [27, 28]. Therefore, the outputs of these devices can be considered indicative rather than exact [29].

The sample is randomly selected but cannot be said to be statistically representative of the wider housing stock, but statistical methods can be applied to generalise findings. The data is expected to identify broad trends in indoor environment quality and highlight areas worthy of more detailed investigations. Furthermore, a top-down modelling approach could use the categorical survey data to identify trends in specific groups of dwellings defined by their type, geography, or the socioeconomic status of their occupants.

2.6 Airtightness

The permeability of a dwelling's thermal envelope is conventionally assessed using a *blower-door*, which artificially and systematically increases the difference between the internal and external air pressures ΔP (Pa) and measures the airflow rate through adventitious openings located in it $\dot{V}$ (m^3/h); see [13]. These parameters are related by a power law:

$$\dot{V} = C\{\Delta P\}^n \tag{1}$$

where C ($m^3h^{-1}Pa^{-n}$) is a flow coefficient and n is a flow exponent. It is common to report $\dot{V}$ at 50Pa, interpolated from measurements, when it is known as an air leakage rate, $\dot{V}_{50}$ (m^3h^{-1}). In order to compare the air leakage rates in different dwellings, it is normalized by a common parameter, such as the thermal envelop area when it becomes an Air Permeability, Q_{50} ($m^3h^{-1}m^{-2}$), or the building's volume to give N_{50} (h^{-1}). These values of *airtightness* are often used to infer an *infiltration* rate, the rate unconditioned ambient air passes through adventitious openings located in the thermal envelope of a dwelling.

Measurements of airtightness are not a legal requirement in Chile and so there is no national database equivalent to that compiled in the U.S. [30], which contains more than 160,000 measurements.

The most significant number of measurements have been made in 187 dwellings built in between 2007 and 2010 [31], immediately after the implementation of national regulation of fabric thermal performance [32] (for a description of Chilean thermal regulations, see Section 3.6), and so this sample was calculated to minimise the probability of making type I and II errors. . Dwellings were selected by common parameters, such as year of construction, main structural material, and dwelling type. This project has access to the measurements made in Santiago ($N = 65$), where Q_{50} and n are reported with dwelling geometry and structural materials.

A second dataset comprises 58 social houses located in the centre-south and extreme south of the country where space heating is often required [33]. These houses are a statistically representative sample of 15,000 low-income and uninsulated houses that received insulation and improved airtightness interventions from an ongoing weatherisation program that started in 2009 and is subsidised by the Chilean Government. Only N_{50} values were reported. Most houses have a significantly reduced N_{50} after the intervention (median (M_{dn})=36.0h^{-1}; mean (μ)=41.9h^{-1}; standard deviation (σ)=23.9h^{-1}) than before it (M_{dn}=39.0h^{-1}; μ=51.2h^{-1}, σ=34.6h^{-1}). When performing a sign test, the results show that the difference between the two groups medians is statistically significant: $Z(58)=-2.54$; $p \leq 0.05$; Cohen's, $\delta = 0.39$ (small). Although the effect size (Cohen's δ) is of a small magnitude – as denoted by the thresholds given by Cohen [34], it still has a practically relevant influence on the outcome of the analysis.

These are very small samples, which do not capture the variability of N_{50} across the stock and are inevitably biased toward the building types tested. Infiltration is responsible for a significant proportion of the energy demands and GHG emissions of national housing stocks; for example, infiltration is responsible for 11-15% of UK housing stock energy demand, and 10-14% of its carbon emissions [13]. Therefore, future field work is required to develop a database that can be used to identify construction quality, and infiltration's contribution to Chilean national energy demand and GHG emissions.

2.7 Weather

Chile is an inter-continental territory with a north to south length of 4,300km and an average width of 177km. It is bounded on the West by the Pacific Ocean and on the East by the Andes mountain range, and so elevations range from sea level to 6.8km. The mean annual air temperature varies by up to 6°C laterally and by more than 15°C longitudinally, and differences in relative humidity are similarly well defined [35].

Weather data for Chile is available from four sources. The first is the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) International Weather for Energy Calculation version 2.0 (IWEC2) database [36], which provides weather data for 14 locations in Chile. The data is averaged over 12 to 25 years giving wind speed and direction, sky cover, visibility, dry-bulb temperature, dew-point temperature, atmospheric pressure, and liquid precipitation at hourly intervals. A second database is provided by the Department of Meteorology of Chile (DMC), which contains climate averaged data from 24 weather stations for a 30-year period between 1970 and 2000, organised by month and year. A third source is the Meteonorm meteorological database that uses Meteonorm software [37] to stochastically generate time-series data for typical years. It does this by interpolating from long term monthly means for a location by combining a database of ground station measurements with those from five geostationary satellites. The root mean square error for ambient air temperatures is 1.2°C [37]. Finally, the Climate.OneBuilding platform gives weather files derived from a number of public sources to support building simulations and data is available for 14 regions [64].

There are currently no weather files that can be used to predict future climate change scenarios, although guidance could be sought elsewhere; for example, the future weather files of the Chartered Institution of Building Services Engineers [38].

3 STOCK CHARACTERISATION

This section draws on the data source described in Section 2 and uses it to characterise the housing stock and its occupants.

3.1 Dwelling Quantity, Type, and Geometry

The Chilean housing stock comprises 6.5 million residential units, where 79.5% are houses and 17.5% are apartments [1]. The remaining 3% are classified as *other* types and includes mobile homes, uninsulated timber *Mediaguas* used for temporary or emergency accommodation, and self-constructed (using local materials) *Ranchos*. The characteristics of the stock vary according to the local weather conditions and the availability and the affordability of building materials. This variation is considerable because of Chile’s varying geography (see Section 2.7). Anecdotally, there is known to be some variance in dwelling shapes and layouts by geographical region, but these variables are not recorded by any known survey. Other information, such as household size and socioeconomic status, is parsed by geographical region in the Supplementary Materials; see Section 4.1.

3.2 Year of Construction

The age of buildings is currently undocumented. An estimation can be made using a Buildings Permit date, which is the date when the building was approved for construction, rather than the date of actual construction. This is an area of epistemic uncertainty and future surveys should aim to estimate the age of dwellings. Building age can be used to infer levels of insulation (see Section 3.6) and some construction practices, energy demand, envelope air permeability, and sources of indoor pollution.Number of Rooms and Floor Areas

Floor area is a key parameter and independent variable with relevance to both energy and mass transfer models. Building Permits register the gross lettable area for every new dwelling, and the CASEN survey categorises the useful floor area into six bands. This difference in area types is a source of epistemic uncertainty and may result in an overestimation of room volumes. To obtain a volume, the ceiling height must be assumed. Here, the Chilean building code suggests a height of 2.3-2.4m [32]. Therefore, it would be useful if future surveys could clearly define the floor areas they use and record ceiling heights.

Buildings Permits [17] issued between 1990-2015 show that the average floor area of new dwellings has increased by around 40% over the last 27 years from 57m^2 in 1990 to 82m^2 in 2015, with a strong Pearson’s r correlation coefficient of 0.86 for these two variables; see Figure 1. This is important because floor area is often correlated with energy demand [20], and suggests an increase in energy demand and carbon emissions attributable to the housing stock over the same period. However, the national mean floor area of μ=65.5m^2 (σ=46.1m^2; M_{ed}=54m^2; 98%CI: [14-252 m^2]; see Figure 1 box plot) is modest when compared to those of many other nations; see Figure 2. The relationship between house size and the number of rooms has remained relatively stable over time, but has started to increase over the past decade. Therefore, the energy demand and carbon emissions of the housing stock are very likely to continue to increase in the future if house sizes approach those found elsewhere around the world, unless steps are taken to mitigate them.

3.3 Occupancy

The number of occupants, their location and activities are required when modelling IAQ and energy demands to correctly allocate pollutant sources and sinks, determine occupant exposures, and to determine the demand of services. The mean number of occupants per dwelling is 3.64 persons [14], which is high when compared to the UK's mean of 2.35 occupants per dwelling. The mean occupancy density in Chilean dwellings is 18.2m^2 per person.

The total number of rooms in a dwelling is commonly used to determine overcrowding. In Chile, overcrowding is defined as more than 2.4 persons per bedroom, and occurs in 11% of dwellings surveyed by the 2015 CASEN, in 15% in 2002 census, and 7% in 2017 census. The European Union, the UK Government, and the American Crowding Index deem a house to be overcrowded when there is more than 1 person per room, where a room is defined as any enclosed space with habitable conditions [39], and thus excludes sanitary rooms and circulation spaces. Using this metric, overcrowding occurs in 19.6% of Chilean dwellings that contain only one household, which is high when compared to 3% in England, where English households comprise approximately 80% of all UK households [40].

Some dwellings are unoccupied, either temporarily (unfurnished and ready for sale or rent or furnished but occupants are absent) or permanently (abandoned). It is important to understand the type of occupancy of a dwelling because it affects its energy demand and the health of its occupants. However, only the total number of houses corresponding to each occupancy types is reported by the censuses (see Section 2.1) and so this information is not available for individual dwellings.

Ventilation in Chilean houses is predominantly naturally driven, occurring through windows and doors, and is dependent on occupant behaviours and preferences. For example, windows opening is associated with the time of day and outdoor air temperatures [62,63]. Therefore, knowledge of indoor activities that relate to ventilation is required to help develop accurate models of airflow rates and pollutant transport

3.4 Cooking and Heating

The combustion of fuels used for heating and cooking are significant contributors to the total energy demand of houses and their carbon emissions. The most common heating systems are stoves and are used by 80% of the stock, where 45% are fuelled by wood, 25% by bottled LPG, propane, and butane gases, and 10% by kerosene/paraffin [18]. This diversity of fuel sources is not found in the UK where 90% of dwellings have a gas fired central heating system with a mean efficiency of 82.5% [9, 13]. The modern stoves used in Chile are relatively airtight to minimise indoor emissions [65]. However, smoke exhausted outside may re-enter a house through its infiltration and ventilation openings. In this specific circumstance, low ventilation rates might benefit both energy demand and indoor air quality.

Heating and cooking systems, and the cooking of foods, are known sources of pollutants associated with elevated incidences of asthma, wheeze, airway obstruction, and lung cancer [41]. The most common cooking fuel used in Chilean houses are bottled gas (86%) and wood (13%) [14].

Pollutant types and emission rates associated with heating and cooking are functions of the fuel, the condition and design of the heater or cooking stove, food types and cooking methods, and the frequency and period of heating and cooking events [28]. The presence and use of fans is not recorded by any known public source, and so would be an area of uncertainty in any model designed to predict ventilation rates in dwellings. Therefore, it would be advantageous if they were documented by future surveys to inform models and show compliance. Here, Chilean Standard DS No66 [66] states that vents must be located in any room with a combustion appliance and that its cross-sectional area must be >100cm^2.

3.5 Construction and Finishing Materials

Information on predominant structural materials used in walls, roofs and flooring are reported in the censuses. The prevalence, availability, and affordability of materials varies by region; for example, concrete block construction is more common in the north and centre of the country, brick work is most frequently used in the centre and centre-south, and wood is most commonly used in the south and extreme south. Roofs are predominantly constructed using zinc or fibre-cement boards, whereas floors are mainly constructed from wooden planks [14]. More recently, new construction typologies and technologies have been increasingly incorporated into the stock, and traditional or vernacular architecture has given way to light-weight and mixed systems. More detailed information is given in the building permits (see Section 2.2), which describes up to three types of finishing materials per building element and their percentage of use.

Dwellings built before 2000 did not have to conform to any thermal requirements and are generally considered un-insulated. Those built between 2000 and 2007 were required to meet a maximum thermal transmittance for roofs only, with values varying between 0.25 and 0.84 W/m^{2}K depending on their longitude and the number of local heating degree days. After 2007, envelope components in contact with the ambient air had to meet threshold thermal transmittance values that varied depending on their location; for example, walls varied between 0.6 and 4.0 W/m^{2}K. Furthermore, the maximum allowable glazed area ranged between 12-50% for single glazing windows, and 37-80% for double glazed. Here, the housing stock is divided into two groups based on their construction year: those constructed before 2008 when little or no insulation was required by law, and those built afterward that are insulated to some degree. This information can be used to inform energy and mass transfer models, and to highlight sections of the stock that may require interventions to improve the energy performance of their fabric.

4 REPRESENTATIVE ARCHETYPES

It is quicker and cheaper to model the performance of a stock of dwellings than it is to measure it *in situ*. Furthermore, a model can be used to consider the consequences of future changes to the stock. However, the size and diversity of large stocks of dwellings often makes modelling individual buildings impractical, and so they can be grouped together by common factors into *archetypes* to make the dataset and the number of models more manageable. The unique properties of each dwelling are replaced by representative values when they are allocated to an archetype, which increases uncertainty in the stock model and makes tracking data sources important. The performance of the whole stock is then considered by extrapolation [9].

The common factors used to create archetypes are determined by the model's outputs or performance indicators derived from them. The change in the size, composition, and characteristics of the stock over time requires archetypes that can be updated quickly and easily. There are no archetypes reported in the literature that could be used to represent the Chilean houses at stock-scale, or for any other South American country.

The information given in Sections 2 and 3 can be used to apply bottom-up techniques to model building physics using a set of nationally representative archetypal dwellings. These are commonly developed using *clustering* techniques that group dwellings so that the houses in each group are more similar to each other than those in other groups. For example, the English housing stock (a subset of the UK stock) of 22.3m dwellings is represented by over 14,000 archetypes using a statistically representative cross sectional survey of the stock and a clustering method. Each archetype is weighted where the sum of all weights is the number of dwellings in the stock. It has been applied by a seven independent models [9] to investigate energy related research questions, and recently reduced to just over 1000 archetypes by considering 8 parameters of interest and used to investigate housing stock decarbonisation strategies [42]. Similarly, over 200 archetypes are used to represent 80% of the US housing stock [43] by considering 10 parameters of interest, and 593 archetypes are developed for four European countries [44]. Further examples are given by Geyer *et al.* [45]. However, the location data for surveyed houses is currently too aggregated to apply geographic information system techniques, such as those employed by Ghiassi and Mahdavi [46].

4.1 Method

This section uses the terminology of Persily *et al.* [43] and broadly follows the methods of Persily *et al.* [43], Mata *et al.* [44], and Shi *et al.* [47]. The majority of the data given in Sections 2 and 3 are either discrete (number of rooms) or categorical (building type), and so a set of archetypes are developed by the aggregation of relevant characteristics, known as *factors*. Each factor may have a number of discrete *levels* (number of rooms, storeys) or may be continuous (total floor area). By allowing some factors to be varied within known limits, the uncertainty in the dataset can be assessed when simulating the models using an appropriate sampling method, such as Monte Carlo; see [48]. Factors are chosen according to their influence on both indoor air pollution concentrations and the energy demand of a dwelling, by their availability in the dataset, and from information given in the literature. Reports of field measurements and modelling generally relate the results to air change rates and the type of ventilation system, the location of the dwelling, the year of construction, pollutant sources and their locations, occupancy levels and behaviours, and environmental parameters (including temperature and relative humidity), building type, geometry and size, air permeability, envelope thermal transmittance, construction year, number of floors, and building orientation [13,62,47]. Predominant factor levels or values in the housing stock are selected to aggregate the data entries into clusters, or *cells*, where houses share common characteristics. Some factor values occur more frequently than others in the dataset causing the formation of larger cells.

When selecting data sources, those considered reliable in Section 3 were given priority. Data entries from the 2002 census (see Section 2.1) and Building Permits (see Section 2.2) were aggregated and used to develop the archetypes. Key descriptive data from the 2002 census data, dwelling geometry (2 levels: *apartment* or *house*), number of zones (8 levels), and number of bedrooms (i.e., levels 0-7 for houses and 0-6 for apartments, whereby 0 is a studio), were used to create 120 cells in a factorial manner, comprising 8×8=64 for houses and 8×7=56 for apartments. The same method was applied to Building Permits to determine their type (3 house levels: *detached*, *semi-detached*, *terraced*; 1 apartment level), construction period (2 levels: ≤2007, >2007), and building size (8 levels: *number of zones*) was used to create a total of 4×2×8=64 cells. Following the aggregation of both databases, the census data was given priority because it is a more comprehensive data source (see Section 2.1). This was done by weighting the aggregated data by the dwelling type and construction period so that $(64 \times 3 \times 2) +$

$(56 \times 1 \times 2) = 496$ cells were defined (see Section 4.2). The analysis was performed using bespoke R (Version 3.5.0) code and the `aggregate` and `subset` functions, which is also available to download[1].

The *building size* parameter is contained in both datasets and, although they each use different descriptive factors, it can be used to unify them. Here, the 2002 census gives the *number of rooms* whereas the Buildings Permit gives the *total floor area*. The number of rooms is used as an estimator of dwelling size. Means, standard deviations, and median values found within each cell are used to define levels; see Supplementary Material[1] for the values assigned to each number of rooms. The *number of bedrooms, living rooms, and kitchen* were assigned according to the total number of rooms. The *number of floors*, *occupants* (from 1 to 5), and *bathrooms* were selected from the mode.

The 496 cells were used to classify the stock as dwelling archetypes and weighted using the number of dwellings in each cell. However, this is a significant number of archetypes. Sousa *et al.* [42] use a high performance computer to simulate their 1,000 archetypes, whereas Persily *et al.* [43] and Mata *et al.* [44] only use a few hundred archetypes making simulation computationally less expensive. Therefore, it may be convenient to use fewer archetypes and acknowledge the loss of resolution. To select groups of archetypes to be used, the proportion of the stock represented by each archetype and a Null Hypothesis Significance Testing were used. The differences in observed frequencies between two archetypes were compared using the chi-square (χ^2) test of statistical significance [49]. Emphasis was placed on the magnitude of the differences to represent a set of archetypes. This method allows each archetype to be removed from the whole set, by comparing the largest (i.e., archetype with the highest observed frequency) to other archetypes in a descending order of frequency. When comparisons made against the largest archetype frequency were of a practically relevant effect size (as denoted by classified

---

[1] All code and supplementary archetype information are available from ResearchGate. See project: *A data analysis of the Chilean housing stock and the estimation of uncertainty in indoor air quality in Chilean houses.*

Full database DOI: 10.13140/RG.2.2.16242.15041

Supplementary material DOI: 10.13140/RG.2.2.13227.49440

Code DOI: 10.13140/RG.2.2.16058.64961

thresholds), all comparisons made prior were used to represent the overall stock. Hence, less importance was placed on the statistical significance. The chosen effect size index, φ for group comparison, is given by [50]

$$\varphi = \sqrt{\frac{\chi^2}{n}} \tag{2}$$

where n is the total number of observations considered for each statistical comparison. Thresholds were then used to interpret the magnitude of the effect size and are given by Ferguson [51] as *small* ($\varphi \geq 0.2$), *moderate* ($\varphi \geq 0.5$), or *strong* ($\varphi \geq 0.8$). Effect sizes of $\varphi < 0.2$ are considered to be negligible (i.e., no practically relevant difference).

4.2 Results

Figure 3 shows the number of archetypes required to represent each centile of the building stock and the number of required to meet each effect size classification. The archetypes are ranked by their observed frequency in descending order, beginning with the most common archetype, which is used as a reference for comparison. Figure 3 shows that a set of 496 archetypes can be used to represent the entire national housing stock and 90 of these archetypes represent 95% of the stock (marked by the blue data point), indicating a large number of small outlier archetypes. The number of archetypes reflects the level of resolution provided by the sources of information and the characteristics of the chosen parameters. So, if the method was applied to test different outcomes or research questions, such as a life cycle assessment or thermal comfort, then other parameters might be included in the model changing the total number of archetypes required to represent the stock.

A χ^2 test was used to compare the change in effect size between the most common archetype and each lower ranked archetype. Using the effect size classifications defined in Section 4.1 and Table 3, sets of 2, 8, and 29 archetypes are shown to represent 13%, 35% and 70% of the entire stock with *small*, *moderate*, and *strong* effects, respectively (see the orange data points in Figure 3). The gradient of the line shows a law of diminishing returns as archetypes with ever decreasing weights are added because they do not significantly increase the proportion of the building stock represented. Therefore, Figure 3 can be used to choose a number of archetypes that balances the proportion of the stock represented with the time taken to create and simulate models.

Table 4 shows key factors for the first eight archetypes, which are given in order of their representativeness. The first two archetypes represent a *small* effect size, whereas all eight represent a *moderate* effect size. This latter group comprises six single-story detached houses and two semi-detached houses with floor areas of 41-104m^2. Further information is given in the Supplementary Material. The most abundant archetype represents a 6.8% of the stock, and is detached single-storey uninsulated house, constructed with clay bricks in central cities of the country, with prefabricated panels in northern cities, and with wooden panels in southern cities. It has two bedrooms, two bathrooms and a separate kitchen. Its heating and cooking fuel is generally gas. Its mean household is 3.4 persons whose average socioeconomic status ranges from the 2nd decile (by population) in the centre-south of Chile to the 7th decile in the capital region.

The distributions of some variables vary across the country and between archetypes. Each archetype is further described by the predominant structural materials used in their walls, roofs and flooring, their cooking fuel and the socioeconomic status of their occupants. Variables were assigned according to the observed frequency in each region and data is presented in Supplementary Material. Table 4 and the Supplementary Material both show that the most factor with the greatest variance between archetypes are the floor area and composition of the rooms (number and type). At the regional level, the factor with the greatest variance is the socioeconomic status.

5 DISCUSSION

When defining a housing stock with archetypes, it is expected that there is some variability within each archetype. The extent of this variability is shown in Table 4. The uncertainty in each parameter can be explored by varying it between known limits and running multiple simulations to give a range of outputs, following the approach described by Das *et al.* [48], Jones *et al.* [13], and Sousa *et al.* [42]. The outputs can then be analysed to identify those that are the most common and those that are extreme and unlikely to occur most of the time. This enables policy decisions to be made that affect an acceptable proportion of the stock where some intervention is financially and logistically achievable, rather than all of it. A sensitivity analysis can determine relationships between each of the inputs and the outputs to identify those inputs that are most influential. They can then be targeted for attention when designing new dwellings or by future data gathering exercises when there is epistemic uncertainty in influential model inputs; see Section 3 for examples. These techniques are computationally expensive and so there is a trade-off between computational power and time. To reduce computational time and increase utility, meta-modelling techniques can be used to simplify the model with the penalty of predictive accuracy and simulation flexibility [48]. Models of the archetypes require calibration after the simulation process to minimise their predictive error. This can be done using available *in situ* measurements, such as those described in Section 2.5, following Booth *et al.* [52], Cerezo *et al.* [53], and Balaras *et al.* [54].

The development of a Chilean housing stock model is the subject of future work. However, it is possible to determine metrics from the available data that indicate living standards and that can be compared against those of other housing stocks to benchmark them. Chilean dwellings were shown to have a high occupancy density when compared to European and USA norms [39, 55], and 20% are overcrowded, which could affect occupant health, particularly by exacerbating communicable disease transmission. Furthermore, floor areas are comparatively smaller than those found in many countries, although the average floor area of new dwellings has increased by around 40% over the last 30 years. This suggests that dwellings size is likely to increase if recent economic growth is maintained and the general population becomes wealthier. In many countries, floor area is correlated with the energy demand of a dwelling and so housing stock energy demand and carbon emissions will simultaneously increase unless the energy performance of dwelling fabric, services, and white and grey goods are well regulated. This may prove challenging in the short-term because there is significant diversity in heating and cooking systems and also extensive use of solid fuels, which decreases both indoor and outdoor air quality. This makes it difficult to know where to target interventions first to have the greatest impact, and is dependent on the priorities of policy makers. The optimal approach requires trade-offs between competing needs, which can be explored using techniques, such as multi-objective optimization [48]. Furthermore, interventions designed to improve the energy performance of dwellings can have unintended consequences on occupant health; for example, adverse effects on respiratory health from increased exposure to $PM_{2.5}$ attributable to increased airtightness without an increase in ventilation [8,59,60,61].

The efficacy of interventions can be investigated using models of the archetypal dwelling derived in Section 4.

The accuracy of the predictions of any model is a function of the quality of the input data. Section 2 shows that there are a number of data sources that describe Chilean houses, and their occupants, and that the quality of the data is good enough to derive archetypes. However, there are some areas where the data can be improved. There is currently inconsistency between questionnaires and so a collaborative approach between surveying organisations would make linking their outputs easier. Surveys can be improved to reduce uncertainty in some parameters. There is a paucity of information about dwelling ages and types, occupancy patterns and behaviours, ownership, air permeability (especially for dwellings built before 2007), and local environmental conditions, such as sheltering and housing density. Future field work is required to develop a database that can be used to identify construction quality, and the contribution of infiltration to the Chilean national energy demand and GHG emissions. The archetypes can should used to target the most common houses first.

There are also a number of parameters and metrics that are not surveyed by the census that would be useful to know in order to construct basic steady-state energy demand models, similar to the UK's Cambridge Housing Model (see Sousa *et al.* [9] and Jones *et al.* [13] for more details). These include dwelling properties, such as geometries (floor area and volume), window area and glazing type, year of construction, number of floors, insulation level, internal air temperature (perhaps indicated by a thermostat setting), orientation, heating system fuel, and the presence of mechanical ventilation systems. It would also be useful to understand some the type and frequency of occupant activities, such as ventilation behaviour, system and appliance use, cooking, or tobacco consumption, to improve estimations of energy demand, GHG emissions, IAQ, and occupant health risks. However, these parameters may not fall within the scope of the census, and so further data sources are required.

6 CONCLUSIONS

A range of data sources are identified that give an understanding of the Chilean housing stock and have been used to derive archetypal dwellings by considering descriptive parameters relevant to indoor air quality and energy demand. We find that 496 archetypes can be used to represent 100% of the Chilean housing stock and only 90 to represent 95% of the stock. Furthermore, sets of 2, 8, and 29 archetypes represent 13%, 35% and 70% of the stock, respectively, corresponding to small, moderate, and strong effect sizes. The most common archetype represents around 7% of the stock and is a single-storey detached house located in the centre-south of Chile. The archetypes can be used for modelling purposes to predict energy demand and indoor pollutant concentrations and to target data gathering exercises. They can be updated whenever new or better information becomes available.

There is always uncertainty in data. Therefore, when using the archetypes for modelling, a stochastic approach is required to capture both stock variability and parametric uncertainty. To reduce uncertainty in the data in the future, the organizations responsible for surveying houses should collaborate to make linking their outputs easier. Their surveys should also be augmented to fill key knowledge gaps. There is a paucity of information about categorical descriptors, such as dwelling ages and types, ownership, year of construction, and heating and cooking fuels. There is also insufficient granularity in physical, such as dwelling floor area and volume, window area and glazing type, air permeability, and number of floors, insulation level, and orientation. Very little is known about occupant behaviours, such as occupancy patterns, appliance use, cooking, indoor air temperatures and thermostat settings, or tobacco consumption.

The approaches used here to analyse data and develop housing archetypes can be applied to other stocks of buildings in comparable economically transitioning, or developed, countries.

ACKNOWLEDGEMENTS

This work was financed by CONICYT-Chile (Commission of Scientific and Technologic Research of Chile), through the “Becas Chile” Doctoral Fellowship Programme, 2015; Grant No. 72160504. The research described in this article has not been subjected to any review by the CONICYT funding institution so it does not necessarily reflect their views, and therefore no official endorsement should be inferred.

The authors are grateful to Professor Max Sherman for his comments on this paper.

REFERENCES

[1] INE. Censo de Población y Vivienda 2017. Instituto Nacional de Estadísticas, Santiago, Chile. 2018.

[2] World Bank. World Bank Indicators. 2016. https://data.worldbank.org/indicator. [Accessed on: 6 March 2018]

[3] OECD (Organisation for Economic Co-operation and Development). Better Life Index. http://www.oecdbetterlifeindex.org/countries/chile/ [Access date: 7 March 2018]

[4] CDIAC (Carbon Dioxide Information Analysis Center). Available online at: http://cdiac.ess-dive.lbl.gov/trends/emis/tre_coun.html. [Access date: 6 March 2018].

[5] IEA. Energy Policies Beyond IEA Countries; Chile. International Energy Agency. 2018.

[6] IEA. Energy Efficiency Indicators: Essentials for Policy Making. International Energy Agency. 2014.

[7] DAC. History of DAC Lists of aid recipient countries. Development Assistance Committee. 2019. Available from: http://www.oecd.org/dac/stats/historyofdaclistsofaidrecipientcountries.htm [Accessed on: June 07, 2019]

[8] Davies M, Oreszczyn T. The unintended consequences of decarbonising the built environment: A UK case study. Energy and Buildings. 2012;46:80-5.

[9] Sousa G, Jones BM, Mirzaei PA, Robinson D. A review and critique of UK housing stock energy models, modelling approaches and data sources. Energy and Buildings. 2017;151:66-80.a

[10] Hamilton I, Milner J, Chalabi Z, Das P, Jones B, Shrubsole C, et al. Health effects of home energy efficiency interventions in England: a modelling study. BMJ Open. 2015;5.

[11] Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M. A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (7) (2010) 1683–1697.

[12] Das P, Chalabi Z, Jones B, Milner J, Shrubsole C, Davies M, Hamilton I, Ridley I, Wilkinson P. Multi-objective methods for determining optimal ventilation rates in dwellings. Building and Environment. 2013 Aug 1;66:72-81.

[13] Jones B, Das P, Chalabi Z, Davies M, Hamilton I, Lowe R, Mavrogianni A, Robinson D, Taylor J. Assessing uncertainty in housing stock infiltration rates and associated heat loss: English and UK case studies. Building and environment. 2015 Oct 1;92:644-56., doi: 10.1016/j.buildenv.2015.05.033 .

[14] INE. Censo de Población y Vivienda 2002. Instituto Nacional de Estadísticas, Santiago, Chile. 2003.

[15] INE. Censo de Población y Vivienda 2012. Instituto Nacional de Estadísticas, Santiago, Chile. 2013.

[16] INE. Auditoría técnica a la base de datos del levantamiento censal año 2012. Instituto Nacional de Estadísticas, Santiago, Chile. 2014.

[17] INE. Bases Edificación. Instituto Nacional de Estadísticas, Santiago, Chile. 2016.

[18] CASEN. Chile National Socioeconomic Characterization Survey (CASEN). Ministerio de Desarrollo Social. 2015.

[19] GHDx. Global Health Data Exchange. http://ghdx.healthdata.org/ [Access date: 14 March 2018]

[20] Palmer J, Cooper I. United Kingdom Housing Energy Fact File. Department of Energy & Climate Change; 2013.

[21] INE. Encuesta Nacional sobre Uso del Tiempo: Principales resultados de la ENUT 2015. Instituto Nacional de Estadísticas Santiago, Chile. 2016.

[22] McCurdy T, Graham SE. Using human activity data in exposure models: Analysis of discriminating factors. J Expo Anal Environ Epidemiol. 2003;13(4):294-317.

[23] RENAM (Red Nacional de Monitoreo de Viviendas). National Housing Monitoring Network. http://renam.cl/site/acerca-de [Access date: 15/03/2018]

[24] Netatmo. Personal weather station, https://www.netatmo.com/en-GB/product/weather/weatherstation (accessed 02 July 2018). 2015.

[25] Yong Z. Product data manual of PLANTOWER, PMS5003 series. Digital universal particle concentration sensor version 2.3. 2016.

[26] Singer BC, Delp WW. Response of consumer and research grade indoor air quality monitors to residential sources of fine particles. Indoor air. 2018 Jul;28(4):624-39.

[27] Fischer SL, Koshland CP. Field performance of a nephelometer in rural kitchens: effects of high humidity excursions and correlations to gravimetric analyses. Journal of Exposure Science and Environmental Epidemiology. 2007 Mar;17(2):141.

[28] O'Leary C, Kluizenaar Y, Jacobs P, Borsboom W, Hall I, Jones B. Investigating measurements of fine particle ($PM_{2.5}$) emissions from the cooking of meals and mitigating exposure using a cooker hood. Indoor Air. 2019;0(0):1-16.

[29] Jones, B, Phillips G, O'Leary C, Molina, C, Hall I, Sherman M. Diagnostic barriers to using PM2. 5 concentrations as metrics of indoor air quality. In: Proceedings of the 39th AIVC Conference, 2018.

[30] Chan WR, Nazaroff WW, Price PN, Sohn MD, Gadgil AJ. Analyzing a database of residential air leakage in the United States. Atmospheric Environment. 2005 Jun 1;39(19):3445-55. doi:10.1016/j.atmosenv.2005.01.062

[31] Citec UBB & Decon UC. Informe hito 3: Evaluación experimental de muestra representativa, Grupo de trabajo Proyecto FONDEF D10I1025, Concepción, Chile. 2013.

[32] MINVU. Ordenanza General de Urbanismo y Construcciones. Ministerio de Vivienda y Urbanismo. 2007.

[33] Fissore A, Colonelli P. Evaluación Independiente del Programa de Reacondicionamiento Térmico. Informe Final. ArqEnergía and Enersolutions, Chile. 2013.

[34] Cohen J. A power primer. Psychological bulletin. 1992;112(1):115-59.

[35] Castillo, C. Estadística climatológica. Dirección meteorológica de Chile. Santiago, Chile. 2001.

[36] ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). International Weather for Energy Calculation version 2.0 (IWEC2). http://ashrae.whiteboxtechnologies.com/ [Accessed on: 15/03/2018]

[37] METEOTEST. Meteonorm. In: METEOTEST (ed.) Version 7 ed. Switzerland: Meteotest. 2017.

[38] CIBSE. Technical Manual 49: Design Summer Years for London. Chartered Institution of Building Services Engineers, Balham, London. 2014.

[39] Eurostat, 2012. Eurostat statistics database. Available from: http://ec.europa.eu/eurostat. [Access on: Jan 22, 2018]. 2012.

[40] Wilson W, Barton C, Overcrowded housing (England). Briefing Paper 1013. House of Commons Library. 2018.

[41] Borsboom W, De Gids W, Logue JM, Sherman MH, Wargocki P. TN 68: Residential Ventilation and Health. Air Infiltration and Ventilation Centre; 2016.

[42] Sousa G, Jones BM, Mirzaei PA, Robinson D. An open-source simulation platform to support the formulation of housing stock decarbonisation strategies. Energy and Buildings. 2018;172:459-77.

[43] Persily AK, Musser A, Leber DD. A collection of homes to represent the US housing stock: US Department of Commerce, Technology Administration, National Institute of Standards and Technology; 2006.

[44] Mata É, Kalagasidis AS, Johnsson F. Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. Building and Environment. 2014 Nov 1;81:270-82.

[45] Geyer P, Schlüter A, Cisar S. Application of clustering for the development of retrofit strategies for large building stocks. Advanced Engineering Informatics. 2017:31. 32-47.

[46] Ghiassi N, Mahdavi A. Reductive bottom-up urban energy computing supported by multivariate cluster analysis. Energy and Buildings. 2017 Jun 1;144:372-86.

[47] Shi S, Chen C, Zhao B. Air infiltration rate distributions of residences in Beijing. Building and Environment. 2015 Oct 1;92:528-37.

[48] Das P, Shrubsole C, Jones B, Hamilton I, Chalabi Z, Davies M, et al. Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling. Building and Environment. 2014;78:171-82.

[49] Field A, Miles J, Field Z. Discovering statistics using R. Sage publications; 2012 Mar 31.

[50] Kim H. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restorative Dentistry & Endodontics, 2017: 42(2): 152-155.

[51] Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Professional psychology: Research and practice. 2009 Oct;40(5):532.

[52] Booth AT, Choudhary R, Spiegelhalter DJ. Handling uncertainty in housing stock models. Building and Environment. 2012 Feb 1;48:35-47.

[53] Cerezo C, Sokol J, AlKhaled S, Reinhart C, Al-Mumin A, Hajiah A. Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City. Energy and Buildings. 2017;154:321-34.

[54] Balaras CA, Dascalaki EG, Droutsa KG, Kontoyiannidis S. Empirical assessment of calculated and actual heating energy use in Hellenic residential buildings. Applied Energy. 2016;164:115-32.

[55] Blake, K.S., Kellerson, R.L., Simic, A.. Measuring overcrowding in housing. Washington, DC: Department of Housing and Urban Development, Office of Policy Development and Research. 2007.

[56] Moura MC, Smith SJ, Belzer DB. 120 years of US residential housing stock and floor space. PloS one. 2015 11;10(8):e0134135.

[57] World Bank. World Bank Indicators. 2016. https://data.worldbank.org/indicator. [Accessed on: 12 September 2019]

[58] IEA. International Energy Statistics. 2016. https://www.iea.org/statistics

[59] Shrubsole C, Macmillan A, Davies M, May N. 100 Unintended consequences of policies to improve the energy efficiency of the UK housing stock. Indoor and Built Environment. 2014 May;23(3):340-52.

[60] Hamilton IG, Davies M, Ridley I, Oreszczyn T, Barrett M, Lowe R, Hong S, Wilkinson P, Chalabi Z. The impact of housing energy efficiency improvements on reduced exposure to cold—the ‘temperature take back factor’. Building Services Engineering Research and Technology. 2011 Feb;32(1):85-98.

[61] Wilkinson P, Smith KR, Davies M, Adair H, Armstrong BG, Barrett M, Bruce N, Haines A, Hamilton I, Oreszczyn T, Ridley I. Public health benefits of strategies to reduce greenhouse-gas emissions: household energy. The Lancet. 2009 Dec 5;374(9705):1917-29.

[62] Persily A, Musser A, Emmerich SJ. Modeled infiltration rate distributions for US housing. Indoor Air. 2010 Dec;20(6):473-85.

[63] Dubrul C. Inhabitant behaviour with respect to ventilation-a summary report of IEA annex VIII. Oscar Faber; 1988.

[64] Onebuilding project, Dru Crawley and Linda Lawrie. climate.onebuilding.org. Last Accessed 1 September 2019.

[65] Schueftan A, González AD. Proposals to enhance thermal efficiency programs and air pollution control in south-central Chile. Energy Policy. 2015 Apr 1;79:48-57.

[66] SEC. 2007. Decreto Supremo No66 Reglamento de instalaciones interiores y medidores de gas.