Evaluation of housing stock indoor air quality models: A review of data requirements and model performance

Tha'er Abdalla and Chengzhi Peng*
School of Architecture, University of Sheffield, Arts Tower, Western Bank, Sheffield S10 2TN, UK
Tkmabdalla1@sheffield.ac.uk; c.peng@sheffield.ac.uk*
*Corresponding Author

Abstract

To develop and evaluate housing strategies and interventions for improving indoor air quality (IAQ) at a city or national scale, housing stock IAQ models are required. Without such models, reliable prediction of residential indoor pollution concentrations and exposures at various spatial-temporal scales cannot be achieved. To address the areas not covered by past review articles in housing stock performance modelling, this paper presents the first comprehensive review of the housing stock IAQ models published during 2012-2020. Our review was carried out to achieve three outcomes: (1) to identify and summarise the fundamental IAQ modelling approaches including model assumptions, (2) to review the housing stock IAQ modelling methods (engineering and/or statistical), sampling methods, data sources in use and the underlying computation assumptions, and (3) to propose a descriptive framework and a performative matrix. The review resulted in a set of 11 key model attributes with which housing stock IAQ models can be evaluated in terms of data requirements and model performance. To improve robustness and accuracy in concentration and exposure predictions, future housing stock IAQ models should be developed to account for the dynamic interaction between heat transfer, inter-zone airflow and indoor contaminant transport through IAQ-Energy co-simulation.

Keywords: indoor air quality (IAQ), housing stock IAQ models, CONTAM, EnergyPlus, IAQ-Energy co-simulation, data requirements, model performance

Nomenclature

Symbol	Description	Unit
t	Time	*hour (h)*
C_i	Concentration of particles in indoor air at time *t*	μg/m^3
$C_{sources}$	Sum of concentration gain from all sources	μg/m^3
C_{sinks}	Sum of concentration loss from all sinks	μg/m^3
C_o	Concentration of particles in outdoor air at time *t*	μg/m^3
Q_S	Mechanical supply flow rate	m3/h
Q_N	Natural Ventilation flow rate	m3/h
Q_L	Leakage (Infiltration) flow rate	m3/h
Q_F	Indoor air particle control flow rate	m3/h
η_S	Filter with single pass removal efficiency	*dimensionless*
η_F	Filter with single pass removal efficiency	*dimensionless*
P	Penetration fraction of particles	*dimensionless*
β	Deposition Coefficient	h-1
E	Emission Rate of Particles	μg /h
M_{space}	Fluid mass of room/building air	kg

Q_{space}	Heat energy of room/building air	*Joules*
F_{ji}	Mass airflow rate from zone j to zone i	kg/s
m_i	Mass flow through an airflow path connecting zone i to zone j	kg
T_i	Temperature in zone i	K
R	Gas constant of air = 287.055	J/kg.K
P_i	Pressure in zone i	Pa
ρ	Air density	kg/m^3

1. Introduction

Due to the rising levels of air pollution in cities around the world, poor urban outdoor and indoor air quality has received much attention in recent years because of its detrimental impact on the environment and population health. In the UK, outdoor air pollution is estimated to contribute to about 40,000 premature deaths a year [1]. Globally, the number of premature deaths from outdoor air pollution is expected to increase from 3 million in 2010 to a total of 6~9 million in 2060, with the highest increase in countries not affiliated with the Organisation for Economic Co-operation and Development [2]. The fact that urban populations stay long periods of time indoors, the full extent of the health burden of air pollution is yet to take into account indoor air quality (IAQ). As a gross estimate, 99,000 deaths a year have been attributed to exposures to indoor air pollutions across Europe [3].

The world is currently undergoing the most accelerated urbanisation in history. More than half of the global population is now concentrated in urban areas, and by 2060, new floor areas totalling 230 billion m^2 are expected to add to the global building stock [4]. Notwithstanding this unprecedented global growth in both urban population and the building sector, approximately two-thirds of the current building stock will continue to exist in 2050 [5]. Previous studies have shown that building stocks in some developed countries are responsible for a significant portion of the energy demand and greenhouse gas emissions [6], meanwhile, the research into building stock energy modelling at different scales have increased in the past decades [7]. More recently, the extent and effects of increased airtightness due to policy-guided measures to improve building energy efficiency have been investigated [8–10]. The attention to airtightness is of a particular interest and concern since people, on average, spend 85-90% of their time indoors [11–13]. Living in houses with enhanced airtightness, dwellers' exposures to air pollutants may increase where adequate ventilation systems (natural or mechanical) are not installed or poorly maintained [14,15]. To be able to assess indoor air environment with accuracy, the complex interactions of multiple factors such as envelope leakage, occupant's behaviour, pollutant emissions, ventilation rates, building materials, and others remain to be better understood [16].

IAQ modelling has long been an essential topic in the field of indoor air science. Compared to field measurements of indoor air pollution, well-developed IAQ models can be cost-effective tools for estimating levels of indoor pollutant concentration and occupant exposure. Past attempts at developing physical and chemical IAQ models have identified a number of underlying processes and factors in quantifying IAQ outcome: emissions [17–20], infiltration and ventilation [21–23], chemical reactions

[24,25], and surface interactions (sorption and deposition) [26,27]. Moreover, the physical-chemical processes affecting IAQ have been modelled in the dynamic contexts of indoor environmental conditions (temperature and humidity), building energy balance, and ambient atmospheric conditions (e.g., outdoor temperature, humidity, wind, and ambient pollutant concentrations) [16].

Over the last decade, there has been a growing demand of predictive housing stock IAQ models for several reasons: (1) to predict spatiotemporal variations in indoor concentrations of outdoor-infiltrated and indoor generated pollutants for policies and guidelines developments, (2) to quantify resultant indoor personal exposures from policy-driven interventions (e.g. improvements in buildings energy efficiency and changes in ambient pollution levels, and (3) to unravel the interactions between IAQ and other building performance indicators (e.g. domestic energy use and overheating risk) at the building stock level.

As obtaining field measurements of IAQ at a large scale can be time-consuming and prohibitively expensive, housing stock IAQ modelling has turned to computational methods that work with multiple types of data representative of a building stock at different locations and scales (i.e., neighbourhood, city, regional or national). In general, strategies of analysing a building stock are essential to perform building stock modelling. Such strategies have been developed to account for changes in both energy demand and indoor environment quality following in one or more policy changes in renovating domestic building stocks [28,29].

In a recent review of the UK's housing stock energy models, Sousa et al. [28] showed that the models developed could be generally categorised as 'top-down' or 'bottom-up' to work in an aggregated or disaggregated manner. Top-down models primarily predict the macroeconomic performance of a building stock based on the statistical relationships between historical aggregated data and socioeconomic determinants such as the gross domestic product, population, climate conditions and fuel prices [30]. Since the top-down models rely on historical data, they are less capable of testing the performance and impact of new policies and technologies.

Bottom-up models use empirical data sourced from a hierarchical level less than the sector/stock as a whole. The bottom-up approaches account for the performance of individual end-uses, individual buildings, or groups of buildings and extrapolate the sector/stock performance with weightings of each modelled dwelling or groups of dwellings based on their representations of the sector/stock [30]. The bottom-up methods can be statistical, or engineering (physical) based, or a combination of two. As an example, the majority of UK's housing stock energy models are bottom-up models developed from simplified steady-state representations of physical phenomena [28]. To overcome simplification and achieve satisfactory prediction accuracy and consistency, high-resolution housing stock data was used to characterise the stock constitution in terms of building geometry and construction, environmental systems (ventilation, sources and flow paths), and occupancy (patterns of presence and behaviour).

Furthermore, advanced statistical methods such as probabilistic sampling, Gaussian processes and sensitivity analysis are increasingly applied to quantify uncertainties encountered in stock modelling.

Existing housing stock IAQ models have varied significantly in model development and data used. As such, we are motivated to present an up-to-date understanding of the existing housing stock IAQ models and how they may be improved. In the sections followed, an evaluative study of eight housing stock IAQ models published during 2012-2020 is presented. Although these models were developed in specific national contexts (namely, US, UK, and Chile), the underlying methods, tools, sources of data and issues of model performance can be of interest to housing stock IAQ models to be developed elsewhere. In Section 2, the fundamental principles and equations commonly used in IAQ modelling and the scope of our literature search is summarised. Section 3 elaborates on the key questions and tasks in developing housing stock IAQ models identified from the literature including data sources, sampling methods, computational techniques and tools. A descriptive framework and a performative matrix are presented in Section 4, with which the eight housing stock IAQ models were evaluated. Finally, the paper concludes with the gaps in housing stock IAQ modelling required of further research and development.

2. Materials and methods

2.1. Modelling indoor air quality of a building

At the most basic level, there are two approaches to measuring or quantifying air quality inside a building: direct or indirect. Direct, or *in-situ*, approaches involve the deployment of either stationary or mobile sensors combined with data processing facilities. If resourced properly, direct methods can report personal exposures and indoor contaminant concentrations with accuracy given known instrument and measurement limitations. Although time consuming and potentially costly, direct methods can obtain specific IAQ measurements of real species and magnitudes (extremes) as well as the sources of pollutants and emission rates. Indirect approaches, on the other hand, utilise computational modelling and statistical methods to predict indoor contaminant concentrations and personal exposures.

The direct approaches seldom capture the complex dynamic interactions of air particles and transient behaviours within a building or a group of buildings. This is due to the limitations imposed by either the instrumental factors (e.g., device selection, calibration, and reliability) or the sampling methods (e.g., measurement location, sampling frequency and time-averaging period) [31,32]. Moreover, there can be uncertainties associated with individual or a network of IAQ sensors, resulting in measurements that may be potentially misleading [33]. Sharing similar purposes of quantifying indoor air environments, indirect approaches attempt to model such complexity computationally, which can be guided by iterative data-based calibration and hypotheses testing.

Accuracy and robustness of computational IAQ models can be evaluated and improved through field measurements. However, depending on the methods employed, computational modelling may oversimplify the spatial-temporal dynamics in which the physical-chemical processes of air particles or gases (e.g., particulate matters $PM_{2.5}$, PM_{10}, O_3, NO_x, CO, SO_2) take place. Nevertheless, one of the key benefits of the indirect approaches is the applicability of computational IAQ models to evaluate the likely effects of interventions proposed for improving IAQ at scale. Below, the fundamental principles and equations commonly used in quantifying the physical processes of indoor air environment are summarised.

2.1.1. Mass balance models

Derived from understanding the underlying physical factors and processes that govern the transfers and transformations of pollutants in indoor environments, mass balance models provide a relatively simple means of estimating changes in average concentrations of indoor pollutants both spatially (e.g., in a room or group of rooms) and temporally. They are often applied with the assumption that a room or a building can be characterised by well-mixed air volumes. In its basic form, a mass balance model that describes the indoor concentration of air pollutants under specified emissions or removal processes can be expressed by an ordinary differential equation (see **Nomenclature** for the unit of each symbol in equations):

$$\frac{dC_i}{dt} = C_{Sources} - C_{Sinks} \tag{1}$$

Because the air in a controlled indoor environment is intrinsically complex, no single mass balance model is well suited for addressing all pollutants or issues under investigation, even though all mass balance models are based on the same fundamental principle of mass conservation. Following [34], with the assumption that the indoor particles attributes are uniform throughout the interior space, Equation (1) can be expanded to represent a range of factors that determine the indoor concentrations of the particle attribute in a single well-mixed zone:

$$\frac{d(C_iV)}{dt} = E + C_o[Q_S(1-\eta_S) + Q_N + Q_LP] - C_i\,[Q_F\eta_F + \beta V + (Q_s + Q_N + Q_L)] \tag{2}$$

Nazaroff [34] provided an illustration of the mass balance approach which systematically depicts the processes represented in Equation (2) (**Figure 1**). It is worth noting that Equation (2) may be extended to include further processes when the indoor environment under study is represented as multiple well-mixed zones. This includes terms that account for the supply and loss of particle attributes by interzone and infiltration airflows [35].

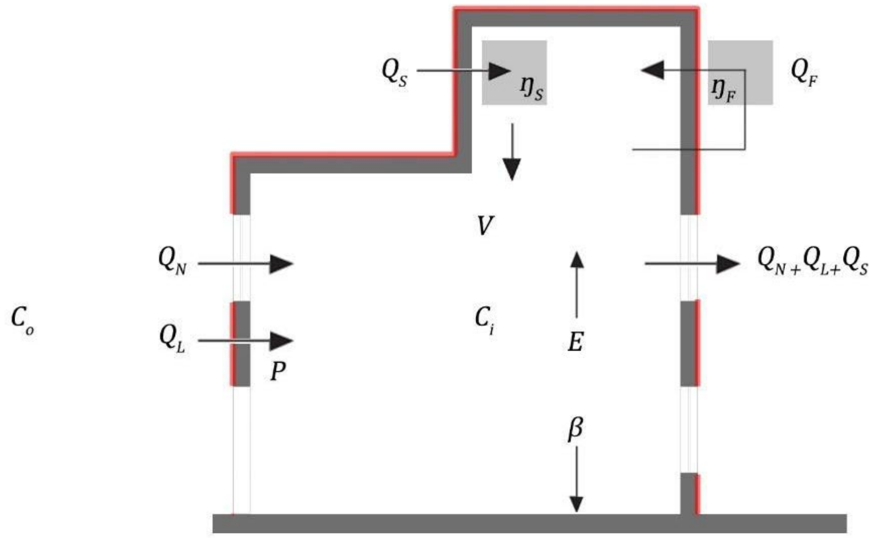

Figure 1. Schematic representation of the physical processes affecting indoor particle concentration levels. (A building section view based on Nazaroff [34]).

In addition to the differences in air pollutants properties, there are myriad variations in how different indoor environments are operated, which renders it difficult nor practical for a single mass balance model to cover all circumstances [36]. Previous studies have tried to illustrate the processes involved in different formulas of the mass balance models [37–40]. However, these studies have primarily been small in scale, and applied only for short periods of time over a small number of locations [41]. The factors and terms in these models are subject to variability and uncertainty in the relationships between the physical environmental phenomena, building characteristics, and dynamic composition of pollutants (see **Figure 2** for a summary). To achieve reliable predictions of IAQ at multiple spatial and temporal resolutions, simulation tools need to be built with mathematical models of indoor particle dynamics that capture the complex physical and environmental phenomena as accurately as possible.

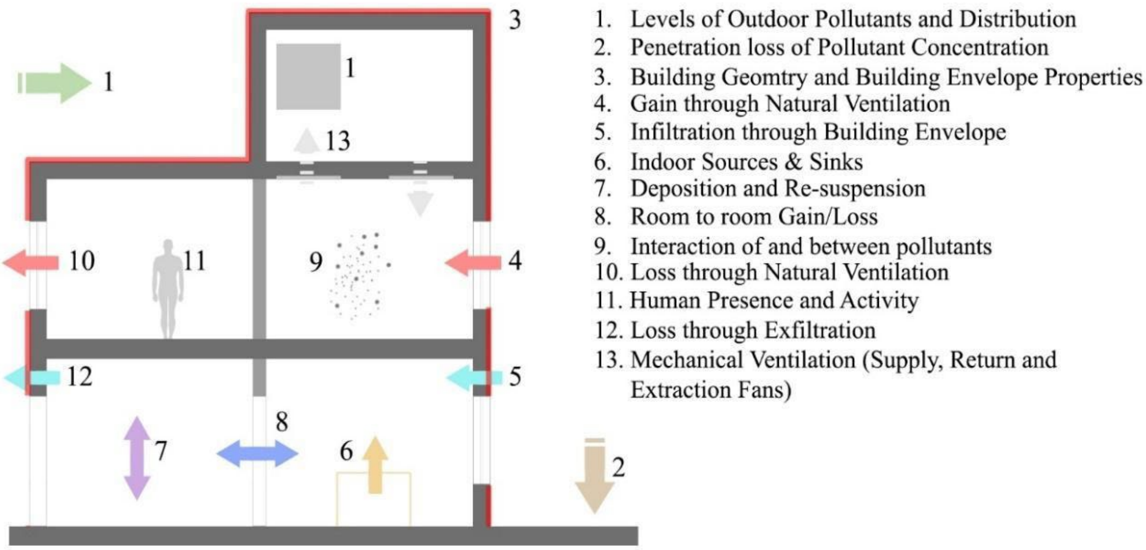

Figure 2. Summary of the main factors and processes affecting indoor concentrations of pollutants (red lines indicate the boundary of the building envelope), based on IEHIAS [42].

2.1.2. Single zone, multi-zone and computational fluid dynamics IAQ models

As summarised in **Figure 3**, single-zone models and multi-zone models adopt different principles, strategies and solvers that generate different outputs. It is difficult to determine which model is the best because of the different requirements of modelling and simulation such as the complexity of the building case, the parameters investigated, the results expected, and the degree of accuracy required [43]. In fact, a wide range of input parameters are required to perform IAQ simulations, including climate data, building fabric and geometry, building systems, and occupancy schedules [44]. As buildings have become more complex, conceptual understanding of the fundamental principles of ventilation and building systems including HVAC must now be coupled with computational modelling to predict contaminant behaviour and the impact on human health accurately.

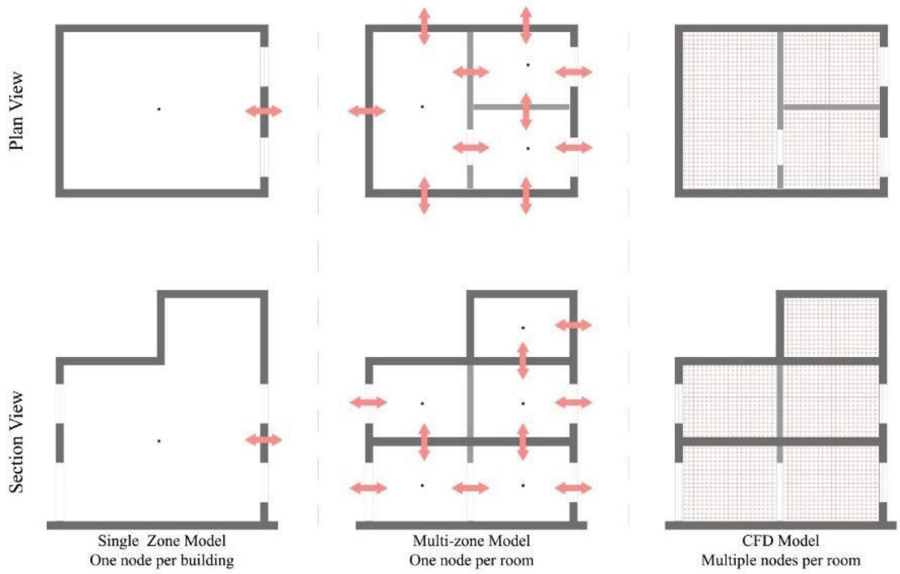

Figure 3. Approaches to modelling IAQ in a building. Left: Single zone models; Middle: Multi-zone models; Right: CFD Models. Each node represents a well-mixed volume. (Based on Axley [45]).

Assuming homogeneous physical properties of air (i.e., uniform temperature, air pressure, and contaminant concentrations), well-mixed single-zone models normally take a *macroscopic* view of air within one volume represented by a node. Meanwhile, multi-zone models define multiple nodes (or zones), with each node representing a room, or a group of rooms connected by several airflow paths. In both models, the airflows between each zone and the outdoor air are calculated iteratively using mass balance equations until the pressure relationships are solved at each time step. Hence, in face of multiple challenges such as the stochastic nature of weather, occupant's behaviours, building components, and uncertainties in simulation input parameters, model choice could have significant implications for estimating indoor contaminant concentrations.

Computational fluid dynamics (CFD) modelling takes a *microscopic* view of airflow in a zone or a group of zones within a building [42]. CFD models are particularly relevant where uniform mixing within a zone or zones cannot be assumed reasonably to represent the airflow conditions under investigation [46]. CFD-based models can compute fine-grained indoor contaminants concentrations and personal exposures, and they have been widely used to simulate contaminants infiltration from outdoor generated sources and contaminants transport between zones within a building [47,48].

2.2. Scope of the housing stock IAQ models review

Previously, there have been a number of reviews covering the field of indoor air science from various angles: (1) the mathematical models used for modelling air infiltration and IAQ [49], (2) IAQ multi-zone models verification methods [50], (3) interaction between IAQ parameters [22], (4) applications of machine learning and statistical IAQ models [51], (5) indoor air pollution exposures across different socio-economic status [52], and (6) the impacts of portable air purification systems on IAQ and health [53]. These past reviews mainly focused on how various statistical methods have been implemented in IAQ modelling concerning single buildings. In this review, a systematic survey of IAQ models developed for housing stocks is presented. The review aims to achieve the following objectives:

1. To identify and summarize the theoretical IAQ models, methods and simulation tools that have been taken as the basis for housing stock IAQ modelling;
2. To survey the sources and types of data used in developing housing stock IAQ models;
3. To identify the scope and components of existing housing stock IAQ models;
4. To propose a performative framework for evaluating housing stock IAQ models; and
5. To identify the gaps in housing stock IAQ modelling for further research.

Our literature search was limited to the IAQ field. Peer-reviewed journal articles and conference papers were searched using the Google Scholar, Science Direct and Scopus search engines. The keywords used for the search in either title, abstract, or keywords were "IAQ" AND "prediction" AND "stock modelling" AND ("building stock" OR "housing" OR "domestic" OR "deterministic" OR "probabilistic" OR "metamodelling" OR "sensitivity analysis" OR "building simulation" OR "multi-zone model" OR "machine learning" OR "neural network"). Following the searches, all articles identified were analytically reviewed by: (1) looking into the compositions and dynamics of housing stocks, including sources of data and housing stock modelling and sampling methods, (2) analysing the computation processes and model assumptions, and (3) applying a set of evaluation criteria on each of the selected housing stock IAQ models to identify the gaps for further research.

3. Housing stock dynamics and computational methods

Policymakers in many countries have actively engaged in establishing regulations and guidelines for improving and maintaining urban air quality [54–56]. In order to improve the IAQ for both new and existing neighbourhoods, policymakers and other stakeholders need to understand which factors contribute to IAQ deterioration and likely effects of proposed retrofitting strategies. To meet such needs, housing stock IAQ models have been developed as one of the primary resources for reviewing and formulating IAQ and associated public health policies. Below, the key questions and tasks in developing housing stock IAQ models are discussed under two subsections: (3.1) Housing stock composition and dynamics, and (3.2) Computational methods.

3.1. Housing stock composition and dynamics

Housing stock IAQ modelling can be defined as an attempt at quantifying and predicting the IAQ of dwelling types that are statistically representative of housing stock at a city, regional or national scale. A housing stock located in a geographical domain is the total account of dwellings that are subject to planned or organic changes over time. Changes in the composition and characteristics of a housing stock can be attributed to multiple factors such as climate and environmental changes, socioeconomic and demographic changes in households, or retrofitting measures applied to improve energy efficiency.

3.1.1. Sources of data for developing housing stock IAQ models

Previous studies in housing stock energy modelling have shown that the obligatory data requirements are of two types: *data demand* and *data robustness*. *Data demand* specifies the scope, amount and type of input data required to achieve a satisfactory level of prediction accuracy and consistency [28]. Modellers have used different sources of data to calculate or simulate the energy consumptions attributable to the constituents of a housing stock. As datasets at higher resolutions are increasingly available, the level of detail in energy or IAQ modelling will also increase dramatically. This may help better accounting for *heterogeneity* in the physical and socioeconomic characteristics of a housing stock, if increased complexity and cost in data processing is not a concern. To achieve an appropriate level of disaggregation in cases where only limited data demand can be met, selecting appropriate modelling and sampling techniques to make best use of such data and information is essential [30].

However, applying specific sampling techniques may require assumptions to be made and result in oversimplifying the data analyses, consequently affecting *data robustness* that conveys the sensitivity to data anomalies [57]. Such anomalies are attributed to sampling methods used in the formation of representative datasets of the entire housing stock. Therefore, utmost attention must be paid to appropriate sampling criteria and procedures to reduce bias and errors [28]. Moreover, the average quality of the IAQ modelling process can be affected by instrumental, translational and data entry errors or gaps [58]. Therefore, it is important to track error propagation through each layer and reduce data gaps either by retaining missing values, modifying incorrect measures, or applying *imputations* (assumptions).

For housing stock research, national population and housing censuses can provide essential statistical information on household details ranging from the demographic and socioeconomic characteristics (*e.g.,* income, education, employment status, age, gender, *etc*.), and building characteristics (*e.g.,* built-up area, number of rooms, number of storeys, fuel sources, heating/cooling systems, *et*c.). As an example, the English Housing Survey (EHS) [59] is a periodical national survey of the English housing stock. The EHS is a particularly valuable source of statistical information representing more than 14K English dwelling variants (83.3% of the national housing stock) with weights associated to each variant depending on its occurrence in the stock. It includes a wide range of physical dwelling characteristics, region and local terrain, and household socioeconomics. Similarly, the U.S. Census Bureau's American

Housing Survey (AHS) [60] provides detailed information on a representative sample of approximately 56K dwellings selected using a classification sampling method.

In cases where building data is not available from census surveys, records of *building permits* may contain information about building floor areas and building ages. In fact, building *floor area* and *floor height* are two key parameters most relevant to both heat and mass transfer models [61]. Moreover, building age can infer likely levels of insulation, construction practices, envelope air permeability, energy demand, and indoor air pollution sources. As demonstrated in the Irish dwelling archetypes study [62], and urban energy modelling of the housing archetypes in Kuwait City [63], characterisation of archetypes, an important step in housing stock modelling, can be based on such information.

Previously, Vardoulakis showed that household behaviour within a dwelling can be more significant in determining the energy and IAQ than either the dwelling or household size [15]. The studies by Gunn et al. [64] and Huebner et al. [65] showed that socioeconomic factors can have substantial implications on the household behaviours and energy uses. This is attributed to occupants' varying needs according to age and health, domestic habits and consumption patterns, as well as indoor thermal sensations and preferences. Hence, the domain of household behaviour is a known source of uncertainty which can significantly affect a model's prediction accuracy [64]. To address this uncertainty, household behaviours have been represented by average time-activity profiles which can be collected from specially designed diaries and questionnaires [15], and then combined with population census data to determine the level of population exposure to indoor air pollution. Adopting a time-activity (or micro-environmental) modelling approach, Dimitroulopoulou et al. showed how individual and group exposures to indoor air pollution could be reconstructed by summing time-weighted contaminant concentrations in various dwelling microenvironments where people spent most of their time [39].

Furthermore, internal thermostat set points and infiltration rates can have significant effects on the predicted energy use and mass concentrations. The Energy Follow Up Survey (EFUS) [65] conducted by the British Research Establishment provides data on the internal heating set-points (HSP's) temperatures to reduce modelling assumptions on how energy is used in the UK dwellings. On the other hand, infiltration rates are harder to collect due to the complexity of *in-situ* methods [66]. In this case, reference data may be used as typical values based on built form and construction period [67,68]. In most cases, historical energy use data or energy benchmark data (e.g., energy performance certificates EPC) are available at an individual building level and can be used to calibrate building stock energy models with actual energy data [30]. For instance, the US Department of Energy (DOE) Residential Energy Consumption Survey (RECS) is conducted every three to four years and provides statistical information on housing units, households and energy consumption.

According to Molina et al. [61], the data required for modelling the IAQ of a housing stock can be classified into three categories: (1) Demographic Characteristics, (2) Housing Stock Morphology, and (3) Environmental Data. Where data does not exist or exists only on an aggregated level, the assumptions made may be more robust, if choices of methods used in the development of housing stock IAQ models are informed by reliable sources of data [61]. More generally, Amasyali et al. [69] suggested three types of data sources: (1) *Real Data*: data collected by meters, censuses, surveys or monitoring stations (e.g. energy meter readings, population censuses, and socioeconomic household information), (2) *Simulated Data*: data generated by simulation software tools (e.g. air change rates and HVAC runtimes) where real data is limited or not available, and (3) *Public Benchmark Data*: reference datasets taken as assumed values for some input parameters (e.g. ASHRAE 62.1 [70] and 62.2 [71], CIBSE Guide A [72]). **Table 1** presents a summary of the data source, type and attribution in housing stock IAQ modelling.

Table 1
Housing stock IAQ modelling: a summary of data source, type and attribution.

Data Category	Data Source	Details	Type of Data	Data Collection Method	Attribution Level
Demographic Characteristics	Household Population	Number of Dwelling Units, Households, and Total Number of People including Omission (%)	Real Data	Censuses / Surveys	Stock Characterisation and Predication of Indoor Personal Exposure to Pollutants
	Socioeconomic & Demographic Data	Age, Gender, Level of Education, Employment Status, Profession, Marital Status.	Real Data	Censuses / Surveys	Stock Characterisation and Predication of Indoor Personal Exposure to Pollutants
Housing Stock Morphology	Housing Characteristics	Housing Type, Age, Geometry, Location, Construction Materials, Energy Sources, Time of Use (Activity)	Real Data	Surveys / Building Registers	Stock Characterisation, Pollutants Sources and Sinks, Prediction of Indoor Personal Exposure and Pollutants Concentration
	Building Systems	Thermostat (Heating and Cooling) Set Points, Filter Efficiency	Real Data	Censuses / Surveys	Prediction of Energy Consumption Associated with Pollutants Removal
		HVAC Runtimes & Air Change Rates	Simulated Data	Computational Modelling	Energy Consumption and Indoor Pollutants Concentration
	Building Physics	Envelope Thermal Properties, Windows Opening Areas and Schedules, Occupancy Schedules	Real Data	Surveys / Building Permits	Modelling of Indoor Temperatures and Indoor Humidity for Pressure Differences
		Envelope Airtightness	Real Data / *Benchmark Data	Field Studies / Reference Data	Stock Characterisation, Modelling Infiltration and Ventilation Rate for Indoor Temperatures and Pollutant Indoor/Outdoor Ratio
		Wind Pressure Coefficients	Simulated Data / *Benchmark Data	Computational Modelling / Reference Data	Infiltration Rates, Indoor Temperatures, and Pollutants Concentrations
Environmental Data	Climate and Micro-environment	Spatial and Temporal Weather and Ambient Pollutant Data	Real Data	Monitoring Stations / Field Studies	Stock Characterisation in terms of Climate, and Hourly Weather Data and Pollutant Concentrations for Running Simulations

* Public Benchmark Data: Utilising publicly available datasets when required data cannot be obtained through real datasets or simulated datasets [69]

3.1.2. Modelling techniques and sampling methods

According to Ugursal & Swan [30] and Sousa et al. [28], bottom-up building stock modelling follows an inductive path of consolidating microscopic measures such as building properties, internal conditions, usage schedules, and building services systems. Bottom-up models thus require extensive empirical data from surveys, field measurements, and assumptions (in the absence of data) to describe each component required of an engineering (physical) approach [71]. Based on building physics, several researchers have applied bottom-up modelling techniques to develop representative buildings (archetypes) and used them to calibrate and predict building stock energy performance (e.g. Persily et al. [73], Sokol et al. [74] and Ghiassi et al. [75]). Data entries sharing similar or equal categories were grouped or clustered to classify the dwelling types. After the classification, each archetype was characterised with a set of attributes to represent a proportion of the housing stock. So the larger the number of archetypes developed, the more representative of the stock they become and the more widespread are the conclusions derived from the modelling results [61].

To improve quality of predictions in housing stock IAQ modelling, the issue of uncertainty needs to be addressed. Some uncertainties are related to the mathematical models used to represent the physical phenomena, some to the heterogeneity of the housing stock under investigation, and some to unknown or random variations of the input's values (*epistemic* or *aleatoric,* see section 3.2.2 for more detailed discussion). The methods used to quantify these uncertainties include the use of clustering techniques, which reflect the variability between groups, and Monte-Carlo sampling methods that account for variability in the descriptive parameters within groups [76]. More recent methods include Gaussian process emulators for uncertainty quantification and sensitivity analyses to perform complex stochastic building performance modelling [77,78].

Similar to energy modelling, bottom-up housing stock IAQ modelling typically involves (a) classification and characterisation of the dwelling types (archetypes) representative of the housing stock under modelling, and (b) utilisation of modelling tools to evaluate the IAQ performance of the archetypes [29]. The outputs for all archetypes are then extrapolated to a whole stock of dwellings using weighting factors. However, deterministic bottom-up models produce only one output for one building with given inputs. Hence, the deterministic bottom-up engineering methods can be less applicable to many buildings with different sizes, types, ages, functions, and operating conditions.

As housing stocks are complex dynamic entities that undergo constant evolution, the scope of targeted performance indicators and potential interventions (e.g., likely parameters of dwelling retrofitting) should be considered before stock model implementation [61]. Finally, update and calibration processes should be carried out regularly to minimise the errors between the predicted and observed values. **Table 2** summarises the existent housing stock IAQ modelling approaches including their sampling methods and parameter types.

Table 2
Bottom-up housing stock IAQ models: stock modelling approach, stock formulation, and sampling methods.

Modelling Approach	Housing Stock Model Formulation Approach	Sampling Method	Parameter Types	Variability
Deterministic	Archetype Approach A	Classification Characterisation	Deterministic Deterministic Parameters from Literature or Building Data	No
Hybrid	Archetype Approach B	Classification Clustering	Deterministic Key Descriptive Factors Aggregated into clusters, or *cells* Utilising Factorial Design	Reflects Variability Between Groups
Probabilistic	*Metamodel (Utilising Machine Learning)	Latin Hypercube Sampling / Monte-Carlo Sampling	Variable Probability Distribution Functions to represent Uncertainty / Variability	Reflects Variability in the Descriptive Parameters within Groups

* Simplified algebraic or statistical model as a surrogate of the more detailed engineering model which allows for lower computational requirements [74]

3.2. ***Computational simulation assumptions***

As mentioned previously, the ideal method for IAQ assessment of existing dwellings is through large-scale data collection campaigns. However, due to both cost and time constraints, computational IAQ modelling has become preferable, especially when evaluating intervention proposals. State-of-the-art IAQ models include multi-zone or, *airflow network* and CFD models. These models can calculate the indoor air properties such as indoor air temperatures, airflow rates, and indoor contaminant concentrations. In predicting a building's IAQ, airflow network and CFD models perform differently in complexity, reliability and accuracy. CFD-based models are computationally expensive as they often resolve airflow dynamics at high spatial and temporal resolutions. Hensen and Lamberts pointed out that there appeared a widespread misconception that uses of CFD will reduce uncertainties and increase accuracy of IAQ predictions [79]. In fact, deviation from the ideal case to either higher or lower complexity can induce risks of simulation errors. Therefore, the selection of appropriate computation methods should be guided by the purpose of the simulation (e.g., airflow network methods for bulk airflow analysis, or CFD to study trends (sensitivity of flow patterns to small changes).

Robinson [80] stated that all simulation models are simplifications of reality, and they are based on abstract representations of real-world phenomena. In this regard, it is necessary to make the assumptions explicit about the computational methods employed in housing stock IAQ modelling. The assumptions made in IAQ simulation tools are summarised below including: computational unit, abstraction of building components and systems, and input variables and parameters.

3.2.1. Computational unit (single-zone and multi-zone models)

As previously mentioned, (section *2.1.2*), both single-zone models and multi-zone models are based on the assumption of perfectly homogeneous or, *well-mixed* conditions (i.e., each zone has an average air pollutant concentration value). In single-zone models, a building is simplified to be represented by a single zone or, *node*, without considering its interior partitions [81]. Consequently, the physical details of heat and mass transfer between rooms within a building caused by temperature and pressure variations are ignored [43]. **Figure 4** illustrates the assumptions, showing the air temperature in a single-zone model represented by an average value of T_{in} (°C) [81]. Basically, a steady state model (see Eq. 4

& Eq. 5.) stipulates that the mass flow rate $\dot{m}_{in}$ (kg/s) should be equal to the outlet mass flow rate $\dot{m}_{out}$ (kg/s) when infiltration is neglected, and the energy is conserved between $\dot{q}_{in}$ (rate of heat energy supplied into room/building (Watts)), $\dot{q}_{out}$ (rate of heat energy removed from room/building (Watts)), and $\dot{q}_{loss/gain}$ (rate of heat energy transferred through room/building structures (Watts)).

$$\frac{dM_{space}}{dt} = \dot{m}_{in} - \dot{m}_{out} = 0 \tag{4}$$

$$\frac{dQ_{space}}{dt} = \dot{q}_{in} + \dot{q}_{out} + \dot{q}_{loss/gain} = 0 \tag{5}$$

Additionally, a single node represents the outdoor climate, and the physical parameters of this node are assigned from weather conditions. Notwithstanding, single-zone well-mixed models are relatively easy to implement and fast to compute, and they are suitable for estimating bulk airflow properties when the domain of interest can be treated as a single zone or node.

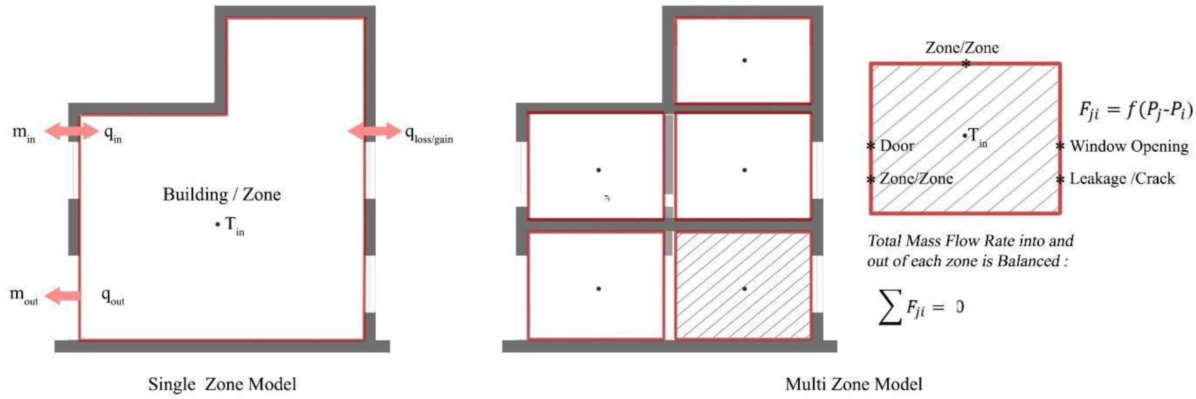

Figure 4. A summary of IAQ simulation assumptions of single-zone steady state models and multi-zone models. (Red lines delineate the inner volume of a zone, based on Yu et al. [43]).

Multi-zone models use *rooms* as the minimum computational unit. They calculate the airflow and contaminant transport inside a building within minutes or seconds. However, shorter computing times can be achieved by assuming homogeneity in each zone, that is, the distributions of air pressure, air temperature and contaminant concentration in each room are assumed uniform and leaving out the air momentum effect from an inflow opening [45]. Clearly, this is not always the case because vertical temperature gradient exists in rooms filled with stratified flows driven by displacement ventilation or water heating systems [82]. The well-mixing assumptions could be problematic for simulations of poorly mixed air and contaminants. In an earlier review of airflow and infiltration models, Haghighat [49] stated that a multi-zone airflow model should be able to fully account for the driving forces that cause air to flow from outdoor to indoor and between indoor zones, including the stack effect, the wind pressure effect on building envelope, and the effect of HVAC systems on airflow.

In general, multi-zone airflow models are based on constructing a matrix of equations that represent all airflow paths connecting zones (nodes) within a building. A mathematical equation describing each airflow path (i.e. door, window, crack, etc.) is used to numerically solve the resulting matrix, typically by the Newton-Raphson method [83]. All equations are solved simultaneously to ensure reaching the convergence state when the sum of all mass flow rates through all flow paths approaches zero as illustrated in Eq. 6.

$$\sum F_{ji} = 0 \tag{6}$$

In a multi-zone model, the mass airflow rate at each airflow path is some function of the flow pressure drop along the flow path, P_j-P_i [84], and is expressed as:

$$F_{ji} = f(P_j\text{-}P_i) \tag{7}$$

The mass of air, m_i (kg), in zone *i* is given by the ideal gas law:

$$m_i = \rho V_i = \frac{P_i V_i}{R T_i} \tag{8}$$

Although multi-zone models of individual buildings can provide spatial average estimates of pollutant concentrations with a reasonable simplification of indoor physical phenomena, it is possible to describe the building's attributes (e.g., contaminant sources, airflow paths, occupancy schedules, and building service systems) with a high level of resolution. However, to achieve prediction accuracy in housing stock IAQ modelling at a reasonable computation cost, consideration of variability and uncertainty in input parameters is required in selecting appropriate computational modelling methods and tools without risking oversimplification.

3.2.2. Deterministic and stochastic input parameters

As conventional methods of building engineering calculation tend to be deterministic, predetermined values (defaults) are often used without tackling uncertainty [47]. In general, uncertainties in housing stock modelling are of three sources: (1) the heterogeneity within a building stock (e.g., a large range of building characteristics), (2) the first-order or *aleatoric* uncertainties where different simulation outputs are probable given the same building, and (3) the second-order or *epistemic* uncertainties where input parameters can take different values in light of new data or knowledge [52]. Increasingly, uncertainty quantification has been introduced to housing stock energy and IAQ modelling. Based on generating distributions of predictions followed by sensitivity analyses, Das, Shrubsole, Jones et al. developed a probabilistic framework for quantifying uncertain parameters in housing stock IAQ modelling [85]. More recently, Molina et al. applied a similar framework to the Chilean housing stock to quantify the uncertainties in indoor pollutant concentration levels, ventilation, and infiltration [86]. A deterministic model does not consider probable fluctuations of some input parameters of any initial conditions and the solution is one and only [87]. In contrast, stochastic models attempt to quantify some or all the parameters by probabilistic distributions rather than single assumed definitive values.

There are other input parameters in IAQ modelling that can vary according to stock variability and/or measurement uncertainty, such as wind pressure coefficients on building envelope, discharge coefficient of flow paths, temperature stratification, household behaviour, building envelope airtightness, and air exchange rates [88–90]. Das et al. [85] showed that uncertain input data can be contaminants related, such as ambient concentrations, generation rates, and deposition rates. Booth et al. [91] stated that any housing stock model should provide information about the potential risks associated with proposed interventions by displaying a distribution of confidence levels due to the diverse sources of uncertainty. To do so, there are mathematical and statistical methods available for evaluating uncertainties in model inputs and outputs [61]. **Table 3** summarises the sources of uncertain input parameters found in the literature on housing stock IAQ modelling.

Table 3
Sources of uncertain input parameters in housing stock IAQ modelling.

Sources	Descriptive Parameters	Key References
Environment and Climate	Spatial-Temporal Variations of Ambient Contaminant Concentrations, Wind Speed and Direction, Local Outdoor Temperature and Terrain Properties	[66,92,93]
Physical Characteristics	Dwellings Geometry and Layout (e.g. Block Aspect Ratio), Space/Zone Volume, Material Properties, Orientation, Dwelling/Flat Height, and Number of Exposed Facades	[28,85,94]
Building Physics	Zone Pressure, Local Zone Temperature, Air Temperature Stratification, Wind Pressure Coefficients, Building Envelope Airtightness, Ventilation and Infiltration Rates, Flow Path Discharge Coefficient, Airflow Exponent n, and Flow Path Area	[66,86,88,89,91,92,95]
Building Components / Systems	HVAC Runtimes, HVAC Supply and Return Flow Rates, Air Exchange Rates (AER), Filter Efficiency and Removal Rate, Combustion Sources and Emission Rates	[39,94,96,97]
Household and Activity	Household Population, Time-Activity-Location Factor, Occupancy Schedules (HVAC Runtime, Window Opening Area and Time, and Thermostat Set Points), and CO_2 Generation Rate.	[64,90,98,99]
Contaminant Properties	Contaminant Generation Rates (e.g. cooking emission rate), Source Strength, Contaminant Sinks, Contaminant Penetration Factor, and Deposition Rates	[39,85]

3.2.3. Simulation tools used in housing stock IAQ models

Over the past three decades, a number of IAQ simulation tools have been developed such as CONTAM and COMIS [84,100]. These tools have been used primarily in modelling IAQ of individual buildings. More recently, CONTAM and EnergyPlus were used to model IAQ of archetypes in housing stock studies. CONTAM is a multi-zone airflow and contaminant transport simulation tool developed and maintained by NIST [101], which has been validated in many studies in various building types and locations [50,102]. CONTAM has been built with an updated version of the AIRNET model [103] and provides a graphical user interface for intuitive inputs of building zones and construction, airflow paths and other building elements [104].

More specifically, CONTAM provides users with the ability to model airflow rates including infiltration, exfiltration, zone-to-zone airflows driven by mechanical ventilation systems, wind pressures on building envelope and buoyancy effects. CONTAM's contaminant dispersal model is an implementation of Axley methods [105,106] and has been widely used in many studies to predict contaminant concentrations in buildings under multiple design and retrofitting scenarios [107,108].

As a standalone package, CONTAM does not modify zonal air density in response to environmental changes due to building interactions and occupant behaviours. Therefore, CONTAM does not have the capability of performing thermal dynamic simulations on its own. On the other hand, as one of widely used whole building energy simulation engines, EnergyPlus [109] has the ability to simulate airflows in buildings using the multi-zone Airflow Network Tool, an airflow model based on the early versions of COMIS [100] and AIRNET [103]. The Airflow Network Tool is capable of simulating both infiltration and exfiltration rates driven by indoor/outdoor pressure differences, ventilation mechanisms, building envelop permeability in addition to zone-to-zone airflows.

From the perspective of IAQ modelling, CONTAM and EnergyPlus have advantages in respect of each other. CONTAM simulates complex airflow networks in a building and enable users to model an unlimited amount of airflow paths and multiple contaminant species. On the other hand, EnergyPlus performs thermal dynamic simulations and accounts for pressure differences between multiple zones in a building. However, interzone airflows and infiltrations in EnergyPlus are user specified and not pressure dependent as in CONTAM. Moreover, EnergyPlus does not require interzone airflows to be in balance with system airflow rates [110]. Lately, attempts have been made to couple multi-zone airflow models with dynamic multi-zone thermal models to perform dynamic IAQ-Energy co-simulation. However, to date none of the existing housing stock IAQ models adopted the co-simulation approach which will be discussed further in section 5.2.

Using EnergyPlus to model contaminant transport, Taylor et al. have developed the Generic Contaminant Model (GCM) tool, allowing users to model the behaviour of one specific pollutant within a building. GCM enables the modelling of both dynamic thermal behaviour and single pollutant transport within one simulation package [111]. Additionally, Polluto, another in-house tool developed at the University College London (UCL), offers multiple contaminants transport modelling when run with EnergyPlus. **Table 4** presents a comparison between CONTAM and the UCL in-house IAQ tools.

Table 4
Comparison of IAQ simulation tools used in housing stock IAQ modelling.

	Simulation Tools		
	CONTAM [84]	EnergyPlus GCM [111]	EnergyPlus Polluto [111]
Main Usage	Airflow rates, contaminant transport through airflow, and building occupant exposure.	Energy analysis, thermal load simulation, airflow, and contaminant transport.	Energy analysis, thermal load simulation, airflow, and contaminant transport.
User Interface	Simple	Complex	Complex
Thermal Behaviour	Static [Dynamic if coupled with a thermal engine]	Dynamic	Dynamic
Contaminant Behaviour	Yes (A rich set of sources and sinks including deposition and re-suspension)	No	No
Changes in Occupant Behaviour Consideration	Yes	Yes	Yes
Modelling of Pollutants	Multiple Pollutants	Single Pollutant	Multiple Pollutants
Air Leakage Points	Multiple Airflow Leakage Points	A one-to-one correspondence between heat transfer and air leakage	A one-to-one correspondence between heat transfer and air leakage
Mechanical Systems Modelling	Complex & Multiple Systems	One System	One System
Warm-up Days	No	Yes, to ensure any thermal capacitance values are representative of the zone.	Yes, to ensure any thermal capacitance values are representative of the zone.

Capability of building control operations	Yes	Yes, indoor concentrations as flags for ventilation system operation	No
Non-trace contaminants	Yes, already included in air density calculations.	Yes, if coupled with the Heat and Moisture Transport (HAMT) model.	Yes, if coupled with the Heat and Moisture Transport (HAMT) model.

4. Housing stock IAQ models (2012-2020)

Since early 2010s, a number of housing stock IAQ models have been developed to assess IAQ of housing stock at a city or national scale. Based on various datasets and computational IAQ simulation tools as described in the previous section, these models were built to perform mainly simulations of mass transfer processes in sampled representative dwellings. In this section, a detailed review of eight housing stock IAQ models published during 2012-2020 is presented.

4.1. REIAQM

The Residential Energy-IAQ Model (REIAQM) was developed to model and predict the annual energy use for space conditioning and indoor concentrations of various pollutants across the residential building sector in the US [89]. This model takes into account the interactions between energy use and IAQ in different building properties and climates. The main modelling framework was coded in Python and was based on an integration of EnergyPlus, and the mass balance models (see section 2.1.1). REIAQM utilises the geometries and housing characteristics of 209 housing archetypes developed previously by Persily [73], representing 80% of the US residential stock.

Based on the building geometries, foundation types (crawl space, concrete slab, or basement), and construction types (attached/detached garage, and attached/ detached home construction), BeOpt.xml files were generated for the 209 dwellings. Further detailed characteristics were attributed according to the climate zones and years of construction across 19 US cities, including vintage (represented by normalized leakage NL), type of heating and cooling system, building envelope insulation, and thermostat setting. This gives 3,971 home models (209 base models * 19 cities = 3,971) coded in BEopt.xml for producing EnergyPlus outputs including cooling energy use, HVAC runtimes and air change rate (ACR) through natural ventilation, infiltration and mechanical ventilation.

The EnergyPlus simulation results along with data regarding outdoor air pollution concertation levels, indoor emission rates and loss mechanisms were used as inputs to a discrete time-varying single zone mass balance model. The Python script then outputs the annual mean indoor concentrations for various pollutants. Finally, population and dwelling weighting factors were applied to estimate the chronic health impact using a disability-adjusted life-years (DALYs) approach following the method developed by Logue et al. [112]. REIAQM concluded that modelled population weighted annual mean $PM_{2.5}$ concentrations from indoor sources were higher in newer homes in comparison to the modelled population weighted annual mean $PM_{2.5}$ concentrations from outdoor sources because of lower air change rates. Additionally, lower air change rates due to tighter building envelopes in newer homes resulted in lower population weighted annual mean ultrafine particulates (UFPs) and Ozone (O_3).

Finally, the estimate of the total DALY burden of pollutant exposure in U.S. residences was approximately 192 DALYs loss per 100K persons per year.

4.2. LNDN-A

The Domestic Stock $PM_{2.5}$ Model for London **(**LNDN-A**)** was based on a deterministic physical approach carried out to model and predict the indoor exposure to $PM_{2.5}$ in London's domestic building stock [29]. CONTAM was used to simulate two scenarios: (1) the base case scenario for London's domestic stock in 2010, and (2) a hypothetical scenario of the stock undertaking energy efficient refurbishments to meet the greenhouse gas emissions reduction targets set for 2050. The data used for CONTAM simulations were divided into two categories: (a) data common to all simulations of current stock (2010), and (b) data for future (2050) stock. Candidate interventions such as reduction in dwellings envelope permeability and the introduction of mechanical ventilation and heat recovery (MHVR) systems were modelled as energy efficient refurbishment strategies for the 2050 stock.

Input data were divided to common input data for both scenarios and selected input data were allowed to vary within known limits or defined scenarios to ensure that these characteristics were broadly representative of London's domestic stock. Three categories of occupancy schedules formed the basis for personal exposure models: (1) a 'household average' concentration of $PM_{2.5}$ in the living room, bedroom, and kitchen, (2) a 'cook' who occupies all zones during periods of cooking, and (3) a 'non-cook' who never enters the kitchen. The annual mean indoor exposure to $PM_{2.5}$ was calculated from the simulation results for each category of occupants using permeability distributions with the UK domestic stock and assuming that 50% of current London's domestic stock is homes and 50% are apartments. Following that, a sensitivity analysis showed that indoor $PM_{2.5}$ emission and deposition rates and window opening behaviour influenced by indoor temperature had the largest influence on the overall $PM_{2.5}$ concentrations with higher uncertainties associated with these parameters.

The LNDN-A model concluded that cooking-related sources were the main contributor to indoor $PM_{2.5}$ in non-smoking dwellings under present-day weather conditions (2010). Conversely, in the 2050 refurbishment scenario, the annual average indoor $PM_{2.5}$ reduced from 28.4 $\mu g/m^3$ to 8.2 $\mu g/m^3$ because of envelope permeability reductions and introduction of correctly installed and perfectly functioned MVHR systems. However, separate scenarios of 2050, with the reduction of air permeability to 3 $m^3/h/m^2$ at 50 Pa but without introducing the MVHR system, result in an increase of annual average indoor exposure to $PM_{2.5}$ both indoor and outdoor for all occupancy schedules.

4.3. LNDN-B

London Housing Stock $PM_{2.5}$ Model (LNDN-B) **was** funded by the UK Natural Environment Research Council (NERC) as part of the AWESOME Project [13]. This project aimed to determine the indoor $PM_{2.5}$ concentrations from outdoor sources for different housing typologies across London. The study was based on a deterministic physical approach, utilising the Airflow Network Model and the EnergyPlus GCM Model (see Table 4) to simulate the infiltration of $PM_{2.5}$ through the envelope of 15 dwelling archetypes developed previously by Oikonomou et al. [113]. These archetypes represented approximately 76% of the dwelling stock in the Greater London Authority according to the Building Class Geodatabase maintained by the Geoinformation Group [114]. Among the data inputs required for the dwelling archetypes included the U values derived from the Standard Assessment Procedure (SAP) for Energy Ratings of Dwellings [115], and the Permeability values ($m^3/h/m^2$ at 50 Pa).

EnergyPlus GCM simulations were performed with four different orientations and assuming two different scenarios representing occupants' interaction with building components (infiltration only and natural ventilation). Based on the resultant hourly indoor $PM_{2.5}$ concentrations in the living room and bedrooms of each dwelling archetype, the project reported a range of I/O ratios of $PM_{2.5}$: hourly I/O ratio, hourly-monthly I/O ratio, seasonal and yearly average I/O ratios. Besides, annual average ACH (h^{-1}) values were calculated for the occupied rooms. Additionally, a differential sensitivity analysis was carried out to assess the sensitivity of the model to variations in input parameters.

The LNDN-B model demonstrated a range of I/O ratios of $PM_{2.5}$ for different dwelling types, with detached and semi-detached dwellings most vulnerable to high levels of $PM_{2.5}$ ingresses. When the results were mapped in GIS to indicate areas where London housing stock is most vulnerable to high outdoor pollutant levels, central London showed lower I/O ratios of $PM_{2.5}$ compared with outer London, which was most likely caused by the prevalence of flats rather than detached or semi-detached dwellings.

4.4. ENG-A

English Housing Stock IAQ Model (ENG-A) was primarily developed for modelling IAQ in the English housing stock [85]. This model was based on a single-storey dwelling unit to model the winter indoor concentrations of $PM_{2.5}$ linked to internal and external sources. A simplified two-zonal model of the single-storey dwelling was constructed in CONTAM with one zone representing the kitchen and the other represents the remainder of the dwelling. To account for stock variability and uncertainty in input parameters, distributions in dwelling characteristics (ground floor area, height of dwelling, floor level, number of exposed façades, and permeability of building envelope) and the external weather conditions were informed by the EHS [59], giving 2,585 variants of the single-storey dwelling stock. In addition, other input data were assumed to follow either uniform distribution (orientation, indoor temperature, and kitchen window opening time and area) or normal distribution (ambient $PM_{2.5}$ concentrations, $PM_{2.5}$ generation rate, and the indoor deposition rate of internal and external $PM_{2.5}$).

To obtain a *metamodel*, Latin hypercube sampling was used to apply a range of sensitivity analyses to identify and select most relevant IAQ determinants. Results of the sensitivity analysis showed that the opening area of the kitchen window, the generation rate of internal $PM_{2.5}$, and the indoor temperature are the most important variables for indoor concentration of $PM_{2.5}$ from internal sources. Conversely, the volumetric weather corrected infiltration rate, the indoor deposition rate of external $PM_{2.5}$, and the ambient concentration of $PM_{2.5}$ are most important variables for the indoor concentration of $PM_{2.5}$ from external sources.

Two types of artificial neural networks (cascade-forward and feed-forward) were used to predict the winter indoor concentrations of $PM_{2.5}$ from both external and internal sources. The ENG-A model showed that winter indoor concentrations of external, internal, and all sources of $PM_{2.5}$ are 3.4, 12.7, and 16.1 mg/m^3 respectively, but with lower median values, indicating a positively skewed distribution. Additionally, winter I/O ratios were 0.3, 1.0, and 1.3 from external, internal, and all sources respectively. Finally, a linear regression test between the two metamodels and CONTAM predictions revealed good agreements (R^2=0.9 for the internal sources of $PM_{2.5}$ and R^2=0.842 for the external sources).

4.5. ENG-B

England-wide Indoor Overheating and Air Pollution Model (ENG-B) as primarily developed to model and predict the indoor overheating and air pollution risk in England's domestic stock [96]. This model is capable of modelling both current and future domestic stock, along with changes to the climate, outdoor air pollution levels, and occupant behaviour. A set of categorical, continuous and discrete variants were identified to cover a large range of housing types, envelope details, occupancy behaviour, and external environment such that the majority of cases in England and future (2050) were covered. A total of 384 metamodels were produced from the combination of each categorical variable (built form, wall type, location, epoch, occupancy type, and retrofit strategy). Eight continuous and discrete occupancy and building relevant variables were randomly sampled using Latin hypercube sampling for formulating each metamodel as an EnergyPlus input data file. A large number of EnergyPlus simulations were run to calculate health-relevant and energy use output metrics such as overheating metric, $PM_{2.5}$ I/O ratio, relative humidity, and annual heating energy use. Finally, two metamodeling methods (neural networks and radial basis function) were used to reproduce nonlinear non-monotonic relations between model inputs and simulation outputs. The ENG-B modelling concluded that the performance of a metamodel improves as the number of training runs increases. Additionally, the model performance improves with the number of neurons in the range 5–12 and then levels off.

4.6. GBM

Housing Stock Indoor Overheating and Air Pollution across Great Britain (GBM) was developed to produce baseline estimates of indoor heat and indoor air pollution exposure level that would enable comparison with postcode-level mortality data [116]. A database representative of the national housing stock was developed from various sources, and the unique building variants were identified. Based on the modelling framework previously developed for outdoor air pollution, overheating, and coupled overheating and indoor-outdoor air pollution [117,118], these variants were simulated for indoor overheating risk and indoor air pollution levels from both outdoor and indoor sources using EnergyPlus. The simulation results were compiled and mapped from postcode blocks to Lower Layer Super Output

Area, a UK statistical boundary area typically containing a population of around 1500. The GBM output aggregates show that urban areas had higher numbers of dwellings prone to overheating, reduced levels of indoor air pollution of outdoor origins, and higher indoor air pollution from indoor sources relative to rural areas, attributed to the variations in building types.

4.7. ENGW

English and Walsh Housing Stock IAQ Meta-model (ENGW) [117] is an updated version of the model ENG-B as described previously [96]. Developed from the Energy Performance Certificate (EPC) dataset which contains information on approximately 11.5 million dwellings, this metamodel was applied to the housing stock models of England and Wales. The progress made by this meta-modelling includes (1) the ability to vary ceiling height, floor area, and glazing ratio of dwellings; (2) the capacity to vary indoor emission rates of pollutants using a power law distribution; (3) the ability to predict multiple pollutants one at a time by adding deposition velocity of common indoor pollutants; and (4) additional input data on gas connectivity and heating system type to provide insights on potential indoor sources of air pollution. EnergyPlus simulations were run assuming the static occupant behaviour and window-opening schedules according to CIBSE Guide A [118]. The metamodel was then used to estimate the indoor concentrations of $PM_{2.5}$ and NO_2 from outdoor sources in London, and the national estimates of indoor CO levels. Finally, the results were then mapped to show the spatial variation in indoor concentrations corresponding to variations in outdoor concentrations levels for $PM_{2.5}$, NO_2, and CO. Clusters of increased indoor concentration were found in urban areas with higher outdoor concentrations and smaller dwellings due to reduced ventilation potential.

4.8. CHAARM

Chilean Housing Archetype AiR quality Model (CHAARM) was developed to predict uncertainties in indoor pollutant concentrations, ventilation, infiltration rates and associated energy demand in the heating season, including the sensitivity of the model outputs to the inputs [86]**.** The model was based on the previously identified set of archetypes to represent the national Chilean housing stock [61]. CHAARM is a hybrid model that follows a physical engineering approach associated with a stochastic framework for uncertainty quantification and sensitivity analysis. The framework was developed using CONTAM as the simulation tool to model eight archetypes representing 35% of the Chilean Housing Stock. Archetypes and associated inputs were manipulated using bespoke R code [119]. The archetypes were simulated for the astronomical winter period between June 21st and September 21st. Two main extreme scenarios representing occupants window behaviour opening where considered: (1) windows all opened scenario and (2) windows all closed scenario.

CHAARM follows the sampling method described in the ENG-A model [85], and in [66,120], in which deterministic input variables and probabilistic distributions of uncertain input variables were systematically varied and multiple simulations were performed to generate distributions of output variables. Latin hypercube sampling (LHS) was used to obtain the values of each probabilistic input variable to generate multiple sets of input variates. LHS was chosen to improve the stratification of the sample and reduce the amount of simulations required to reach convergences. Simulations were performed for 240 sets of converged data (8 archetypes x 15 geographical locations x 2 window scenarios). Output variables computed from the simulations included: median ventilation rates, total $PM_{2.5}$ exposure levels and total airflow heat losses. A global sensitivity analysis was performed to test for linear, monotonic and non-monotonic relationships between inputs and outputs. CHAARM predictions concluded that 66 % of the Chilean dwellings have daily mean $PM_{2.5}$ concentrations below 25 µg/m^3 (the WHO 24h guideline value), even when the windows are closed at all times.

A list of the eight models is presented in **Table 5**, which are further evaluated in Section 5.

Table 5

The housing stock IAQ models developed and published during 2012-2020.

Nation	Model	Date	Stock Scale	Pollutant	IAQ Performance Measure	Simulation Engine	Modelling Approach	
US	REIAQM	2018	National	$PM_{2.5}$	Indoor Concentration and HVAC Runtimes	EnergyPlus	Physical Deterministic Approach	[97]
UK	LNDN-A	2012	City	$PM_{2.5}$	Indoor Concentration and Personal Exposure	CONTAM	Physical Deterministic Approach	[29]
	LNDN-B	2014	City	$PM_{2.5}$	Indoor Concentration and Mapped I/O	EnergyPlus	Physical Deterministic Approach	[13]
	ENG-A	2014	National	$PM_{2.5}$	Indoor Concentration	CONTAM	Meta-modelling Probabilistic Approach	[85]
	ENG-B	2016	National	$PM_{2.5}$	I/O Ratio, RH, EUI, & Overheating Metric	EnergyPlus	Meta-modelling Deterministic Approach	[96]
	GBM	2016	Regional	$PM_{2.5}$	Mapping I/O and Overheating	EnergyPlus	Physical Deterministic Approach	[116]
	ENGW	2019	Regional	$PM_{2.5}$ and NO_2	I/O and Indoor Concentrations	EnergyPlus	Meta-modelling Deterministic Approach	[117]
Chile	CHAARM	2020	National	$PM_{2.5}$	Indoor Concentration, Ventilation and Infiltration Rates	CONTAM	Probabilistic Approach	[86]

5. Discussion

5.1. Evaluation of the housing stock IAQ models (2012-2020)

Combining the discussions in Section 3 and 4, a descriptive matrix for mapping existent housing stock IAQ model developments is presented along with data and computation (**Table 6).** The table shows that all models/studies adopted a bottom-up engineering approach to estimating/forecasting IAQ at certain disaggregated levels. However, they varied in certain aspects of stock IAQ modelling including the disaggregation level, resolution of output, performance measure, pollutant species, simulation engine, and uncertainty quantification. Here a number of empty rows indicate the data or modelling areas not addressed by the 8-housing stock IAQ models reviewed.

Table 6

Evaluation of housing stock IAQ models: a descriptive matrix.

Location	US	UK						CHILE	General Information
Year of Publication	**2018**	**2012**	**2014**	**2014**	**2016**	**2016**	**2019**	**2020**	

Housing Stock IAQ Models (See Section 4 for detailed discussion of each model)	REIAQM	LNDN-A	LNDN-B	ENG-A	ENG-B	GBM	ENGW	CHAARM		
Composition and Dynamics of Building Stock (Section 3.1)										
Neighbourhood Scale									Stock Spatial Resolution	Stock Modelling Assumptions
City Scale		✓	✓							
National Scale	✓			✓	✓			✓		
Regional Scale						✓	✓			
Top-Down Approach									Modelling Approach	
Bottom-Up Approach	✓	✓	✓	✓	✓	✓	✓	✓		
Deterministic	✓	✓	✓		✓	✓	✓			
Stochastic / Probabilistic				✓				✓		
White-Box	✓	✓	✓			✓				
Grey-Box				✓	✓		✓	✓		
Archetypes	✓	✓	✓			✓		✓	Sample Representation	
Metamodels				✓	✓		✓			
Single Zone Modelling	✓								Airflow Modelling Technique	
Multi-Zone Modelling		✓	✓	✓	✓	✓	✓	✓		
(A: Annual, M: Monthly, D: Daily, H: Hourly)	A	A	A	A	A	A	A	D	Temporal Resolution	Stock Model End-Use
Indoor Pollutants Concentration	✓	✓	✓	✓	✓	✓	✓	✓	IAQ	
Indoor Humidity Ratio										
Indoor/Outdoor Ratio	✓		✓	✓		✓	✓			
Population Exposure	✓	✓	✓					✓		
Annual Space Conditioning	✓								Building Energy	
HVAC Systems Run Time	✓									
Energy Demand								✓		
Air Change Rate	✓									
Overheating Risks					✓	✓				
Ventilation / Infiltration Flow Rates	✓							✓		
Computation & Validation (Section 3.2)										
Geometry	✓	✓	✓	✓	✓	✓	✓	✓	Building Fabric	Simulation Assumptions
Material Properties	✓	✓	✓	✓	✓	✓	✓	✓		
Envelope Leakage (Infiltration)	✓	✓	✓	✓	✓	✓	✓	✓		
Wind Pressure Coefficients		✓	✓	✓	✓	✓	✓	✓		
Zone/Unit Height within Building			✓	✓	✓	✓	✓	✓	Building Layout	
Zonal Representation*	N	D	D	P	D	D	D	D		
Indoor Temperatures (S: Static, V: Variable)	V	V	V	V	V	V	V	V	Indoor Environment Quality	
Variable Indoor Air Flow Rate		✓	✓		✓	✓	✓	✓		
Natural Ventilation Air Change Rates	✓	✓		✓	✓	✓	✓	✓		
Indoor Pollution Sources	✓	✓		✓	✓	✓	✓	✓		
Season (A: Annual, H: Heating)	A	A	A	H	A	H	A	H	Ambient Environment	
Ambient Pollutants (S: Static – averaged, V: Variable)	V	V	S	S	S	S	V			
Weather Data	✓	✓	✓	✓	✓	✓	✓	✓		
Cooling / Heating Systems (Set-points)	✓				✓	✓	✓	✓	Mechanical Systems	
Presence/Operability of Exhaust Fans	✓	✓				✓	✓	✓		
Air Change Rate	✓	✓			✓	✓	✓	✓		
Filter Efficiency	✓									
CO_2 Generation Rate					✓	✓	✓		Occupancy	
Ventilation Schedules	✓	✓	✓	✓	✓	✓	✓	✓		
Behaviour Schedules (Cooking, Smoking, Showering, *etc...*)	✓	✓		✓		✓	✓	✓		
Deposition/Decay Rate	✓	✓	✓	✓	✓	✓	✓	✓	Pollutants Behaviour	
Emission Rate	✓	✓		✓		✓	✓	✓		
IAQ Field Measurements (H: Historical, X: Not Included)	H	X	X	X	X	X	X	X	IAQ Measurement Data	Validation
Calibrated against Survey or Sensing Data	✓								Calibration	

* (D: Detailed Airflow Network, P: Partial Consideration, N: Neglects Detailed Airflow Network), are all internal zones/rooms included in the simulation model/archetype

Following the descriptive matrix above, a performative matrix comprised of 11 criteria for evaluating housing stock IAQ models is proposed. **Table 7** presents an evaluation of the eight housing stock IAQ models in terms of Data Requirements and Model Performance.

Table 7

Evaluation of housing stock IAQ models: a performative matrix.

Location	US	UK						CHILE	General Information	
Year	2018	2012	2014	2014	2016	2016	2019	2020		
Housing Stock IAQ Models	REIAQM	LNDN-A	LNDN-B	ENG-A	ENG-B	GBM	ENGW	CHAARM		
Data Transparency	□	□	□	▣	▣	□	▣	□	Data Requirements	Evaluation Criteria
Data Representativeness	□	□	□	□	□	□	□	□		
Data Granularity	□	□	□	□	□	□	□	□		
Regularity of Updates	□	□	□	□	□	□	□	▣		
Model Transparency	□	□	□	▣	▣	□	▣	▣	Model Performance	
Reproducibility	□	□	□	▣	▣	□	▣	□		
Replicability	■	■	■	■	■	■	■	■		
Sensitivity & Uncertainty Quant.	■	■	■	□	□	■	□	□		
Versality (Design Decision Support)	□	□	□	□	□	□	□	□		
Results Comprehensibility	□	□	□	□	□	□	□	□		
Computationally Efficient	□	▣	▣	■	■	■	■	□		

*No/Low (■), Somewhere in-between (▣), Yes/High (□)

5.1.1. Data requirements

As previously mentioned in **Section 3.1**, housing stock IAQ models rely on available empirical data of the stock under investigation. The following four questions should be addressed when evaluating how the data requirements were met:

(1) *Data Transparency*: Was it clear what types of data have been used? How easy was it for the model users to access and work with the data? Was the accuracy of the data used in the model assessed and explained?

(2) *Data Representativeness*: Have the data collected and used in the development of the model been subject to rigorous statistical significance tests?

(3) *Data Granularity*: Among available data, was the granularity (or level of detail) sufficient to derive insights from the model inputs and outputs?

(4) *Regularity of Updates*: Could the data used in the development of the model be updated on a regular basis in sync with evolution of housing stocks?

5.1.2. Model performance

With reference to Sousa's et al. evaluation of the UK housing stock energy models [28], the following questions are asked when evaluating the performance of the housing stock IAQ models:

(1) *Model Transparency*: Were the model's underlying algorithms, scientific techniques and codes clearly described so that users and researchers can comprehend the model results?

(2) *Model Reproducibility*: Is it likely that different users aiming to solve the same scientific question will make the same input choices, computational steps, and conditions of analysis, achieving comparable results?
(3) *Model Replicability*: Can the model answer similar scientific questions and obtain consistent results when new data are obtained (e.g. different cities and populations)?
(4) *Sensitivity and Uncertainty Quantification*: Does the model accounts for uncertainties in model inputs and outputs? Are dominant contributors (if any) identified?
(5) *Versatility*: Does the modelling methodology allow for applicability to a wide range of building retrofit options or new builds under consideration?
(6) *Results Comprehensibility*: Can non-expert users easily interpret and comprehend model outcomes correctly to reduce the possibility of false inferences and misinformed decisions.
(7) *Computational Efficiency*: Were there appropriate simplifications introduced to optimise computational time? Was the model execution amenable to High Performance Computing?

Previous researches have suggested classifying model transparency into *White-Box Models* (Physics Based Models), *Black-Box Models* (Data-Driven Models), and *Grey-Box Models* (Hybrid Models) (see [121] for example). Although data-driven *(Black-Box)* IAQ modelling for individual buildings has gained increased attention in the last decade [51], current housing stock *IAQ* models rely on building physics models in their simplest form for data collection and validation. In comparison, historical IAQ data measured at the stock level are scarce. The past and current housing stock IAQ models as reviewed here are either *White-Box* or *Grey Box* models.

Accessibility or transparency is a pre-condition to achieve model reproducibility which represents the minimum attainable standard when compared to replicability [122]. It has been suggested previously that black-box data-driven models are potentially not reproducible, and as a result, they are constrained by limited applicability specific to the range of datasets used in developing the models [123]. For example, a model that was trained to predict the IAQ by learning from limited datasets (e.g. data collected from a small group of dwellings) may not perform well outside of the training data (e.g. different physical properties, occupant behaviour, climate context, future interventions, chaotic events, etc.) [70]. Thus, for non-expert users, the purpose of prediction should be made clear and guidance on whether the models are applicable in a new context or not should be provided.

White-Box models offer a higher level of transparency by releasing and maintaining the core calculation algorithms as open-source programs (e.g. CONTAM and EnergyPlus). With the high transparency offered, white-box models can be highly reproducible and versatile. However, there are foreseen issues surrounding the deployment of such models [120]: (1) these models can be oversimplified when the spatial resolution is increased i.e., specific level of abstraction or spatial resolution, therefore, outputs could be erroneous; (2) expert knowledge is required when model assumptions are made or when prediction outputs need interpretation; and (3) assumptions pertaining to the input variables of these models are prone to all kinds of uncertainties.

Open-source IAQ simulation engines allow for user interaction or integration of scripting tools such as the EnergyPlus Generator 2 Tool (EPG2) developed in Python [124]. EPG2 was used in the ENG-B model [96] for batch processing of building input files configured with user-defined variables. The REIAQM project [97] used multiple Python scripts to solve mass balance equations and automate the majority of the simulation processes. This can allow for flexibility, input variability, automation, and increased computation versatility [121]. However, *White-Box* housing stock IAQ models tend to be static and deterministic in nature and often assume linear relationships exist between multiple variables in an ideal system without uncertainty [125].

Alternatively, the development of *Grey-Box (hybrid)* models that integrate physics-based models and multiple statistical analyses can account for uncertainty assessment and quantification. This has been achieved by deploying sampling methods such as Latin Hypercube for Monte Carlo integration. Although the near-random samples generated could be quite large, applying statistical significance test can reduce the number of samples (*e.g. archetypes*) required to represent the entire housing stock with reduced model resolution (e.g. the CHAARM [86]). Furthermore, hybrid models can account for both linear and nonlinear systems by constructing *metamodels* (e.g. Artificial Neural Networks) and use them in predictions. This is of particular interest as air pollutant concentrations outdoor and indoor are best characterised by nonlinear and irregular behaviours due to behavioural, social, and chaotic events [126]. In contrast to *White-Box* models, some hybrid models (for instance, the ENG-A [85] and ENG-B [96] models) suffer from reduced transparency and accessibility, particularly in the metamodel construction phase whereby multiple hidden layers and neurons generate outputs that are extremely difficult to replicate.

In contrast to black-box models, grey-box models can be *scalable* and *versatile*. For instance, the ENG-A model [85] and ENG-B model [96] are based on multiple metamodels constructed individually for each combination of housing stock physical properties, locations, epochs and occupancy profiles. This makes the models *scalable* to include additional information without having to reconstruct the entire model from scratch, and *versatile* to compare the results of different what-if scenarios (e.g. seasonal variation, future technological interventions, chaotic events, etc.).

On the other hand, in spite of the advantages aforementioned, grey-box models may incur higher computation costs. In fact, both white- and grey-box models can be computationally expensive when a large number of *archetypes* or *metamodels* are involved. However, some of these models managed computing efficiency by: (1) using only single-zone models to represent the entire dwelling stock such as REIAQM [97] (but with reduced prediction accuracy for both airflow and contaminant concentrations, see section 4.1); (2) running the simulations on a high performance parallel computing platform as in ENG-A [85] and ENG-B [96]; and (3) reducing the number of archetypes to a statistically acceptable level while acknowledging the loss in model resolution as in the CHAARM [86].

5.2. Gaps in housing stock IAQ modelling for future research

First, the complex and dynamic nature of modelling IAQ, indoor thermal performance and energy efficiency should be evaluated simultaneously. It is important to capture the interdependencies between airflow and heat transfer in zones within a building and between indoor and outdoor. The advantages and disadvantages of EnergyPlus and CONTAM in terms of performing comprehensive and dynamic thermal-airflow simulation are summarised in section 3.1.2. Separately, each engine is limited in its ability to account for detailed thermal processes upon which building airflow may depend critically and vice versa [110]. Future research should tap into the new capabilities offered by *direct co-simulation* [104,110,127] rather than indirect co-simulation (e.g. EnergyPlus in-house tools GMC). By utilising multi-zone direct CONTAM-EnergyPlus co-simulation, airflow and contaminant modelling can better capture the effects from building thermodynamic domains on airflow and consequently contaminant transport [128]. The direct co-simulation approach will increase the accuracy of airflow and IAQ modelling in housing stock models by taking into account the variations in building configurations, occupant activities, interventions (e.g. realistic energy performance technologies as part of energy efficiency retrofit strategies), and weather conditions [129,130].

Secondly, the effect of the dynamic interaction of occupants within dwellings should be extended to reflect more realistic representations of activity, location and time. It is essential to capture the full range of occupant behaviours and their potential influences on energy uses, indoor temperatures and air quality. For instance, the Energy Follow Up Survey (EFUS) [65] dataset provides indoor air temperature measurements for dwellings surveyed as part of the 2010-11 English Housing Survey. Another way is to deploy dwelling time-activity surveys in a form of personal diaries. Such data can be useful to estimate personal exposures in a high spatial-temporal resolution by capturing the average and range of occupant behaviours (e.g. window opening) more accurately.

Thirdly, improvements in current housing stock IAQ models should consider incorporating model calibration. For instance, the Bayesian calibration method has been used to address multiple sources of uncertainties by characterising some model parameters as probabilistic distributions and examining the discrepancies between model predictions and filed measurement data under multiple temporal and spatial scales [78,131]. Comparing with deterministic models and non-calibrated probabilistic models, the Bayesian calibration has proved to achieve a better simulation fit against measured data with reduced errors [131]. However, it remains untested if calibration methods can be applied to housing stock IAQ model development where measured IAQ data may not be widely available at a scale as metred energy use data.

6. Conclusion

A total of eight housing stock IAQ models were identified and reviewed. These models were published during 2012-2020 and provide valuable information for summarising of a common set of descriptive and performative attributes. Firstly, this review shows that there have been statistical, physical and social science frameworks developed for quantifying indoor air pollutant concentrations and exposure levels in housing stocks of various scales. Broadly they are of two types: white-box and grey-box models. White-box models built on deterministic engineering methods, while grey-box models combined engineering and statistical methods to account for stock composition dynamics and uncertainty qualification. The existing housing stock IAQ models vary considerably in levels of disaggregation, complexity, resolution of output, scenario analysis performed, and computational methods.

Secondly, a descriptive framework for summarising the key constituents of the eight housing stock IAQ models surveyed. A performative matrix is proposed for assessing the models in terms of data requirements and model performance. There is useful information gathered here suggesting that: (1) White-box housing stock IAQ models achieve better transparency to modelling algorithms compared to grey-box models; (2) Grey-box models are more versatile and scalable by developing updateable metamodels covering different simulation scenarios and ranges; (3) Both grey-box and white-box models have not yet captured occupant behaviours accurately in terms of activity and location; (4) All models have not accounted for the thermodynamic effects on airflow and contaminant behaviour which should be improved through dynamic IAQ-Energy co-simulation; and (5) all models were not calibrated due to absence of field IAQ measurement data, and therefore the simulation result fits remain unknown.

Thirdly, although these housing stock models were geospatial and demographic specific, the underlying fundamental principles, modelling methods, statistical techniques and computation assumptions can be shared by the research community. By identifying and articulating the data requirements and model performance criteria, this review paper presents an up-to-date reference to further research and development in housing stock IAQ models which will be essential to inform policy-making and implementation of interventions.

References

[1] Holgate ST. 'Every breath we take: the lifelong impact of air pollution' – a call for action. Clin Med (Northfield Il) 2017; 17:8 LP – 12. https://doi.org/10.7861/clinmedicine.17-1-8.

[2] OECD. The Economic Consequences of Outdoor Air Pollution. vol. 15. Paris: OECD; 2016. https://doi.org/10.1787/9789264257474-en.

[3] RCP. Every Breath We Take: The Lifelong Impact of Air Pollution. London, UK: 2016. https://doi.org/10.1088/2058-7058/29/9/40.

[4] GLOBALABC. Global Alliance for Buildings and Construction, 2018 Global Status Report 2018:325. https://doi.org/https://doi.org/10.1038/s41370-017-0014-9.

[5] Wilkinson S, Dixon T, Miller N, Sayce S, editors. Routledge Handbook of Sustainable Real Estate. Routledge; 2018. https://doi.org/10.1201/9781315622750.

[6] Berardi U. A cross-country comparison of the building energy consumptions and their trends. Resour Conserv Recycl 2017; 123:230–41. https://doi.org/10.1016/j.resconrec.2016.03.014.

[7] Reinhart CF, Cerezo Davila C. Urban building energy modeling - A review of a nascent field. Build Environ 2016; 97:196–202. https://doi.org/10.1016/j.buildenv.2015.12.001.

[8] Wang Z, Bai Z, Yu H, Zhang J, Zhu T. Regulatory standards related to building energy conservation and indoor-air-quality during rapid urbanization in China. Energy Build 2004; 36:1299–308. https://doi.org/10.1016/j.enbuild.2003.09.013.

[9] Feijó-Muñoz J, González-Lezcano RA, Poza-Casado I, Padilla-Marcos MÁ, Meiss A. Airtightness of residential buildings in the Continental area of Spain. Build Environ 2019; 148:299–308. https://doi.org/10.1016/j.buildenv.2018.11.010.

[10] Zheng X, Mazzon J, Wallis I, Wood CJ. Airtightness measurement of an outdoor chamber using the Pulse and blower door methods under various wind and leakage scenarios. Build Environ 2020; 179:106950. https://doi.org/10.1016/j.buildenv.2020.106950.

[11] Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Sci Environ Epidemiol 2001; 11:231–52. https://doi.org/10.1038/sj.jea.7500165.

[12] Leprince V, Carrié FR, Kapsalaki M. Building and ductwork airtightness requirements in Europe – Comparison of 10 European countries. 38th AIVC Conf Vent Heal Low-Energy Build 2017:192–201.

[13] Taylor J, Shrubsole C, Davies M, Biddulph P, Das P, Hamilton I, et al. The modifying effect of the building envelope on population exposure to PM2.5 from outdoor sources. Indoor Air 2014; 24:639–51. https://doi.org/10.1111/ina.12116.

[14] Smith JD, Mitsakou C, Kitwiroon N, Barratt BM, Walton HA, Taylor JG, et al. London Hybrid Exposure Model: Improving Human Exposure Estimates to NO 2 and PM 2.5 in an Urban Setting. Environ Sci Technol 2016; 50:11760–8. https://doi.org/10.1021/acs.est.6b01817.

[15] Vardoulakis S. Human exposure: Indoor and outdoor. Issues Environ Sci Technol 2009; 28:85–107.

[16] Tham KW. Indoor air quality and its effects on humans—A review of challenges and developments in the last 30 years. Energy Build 2016; 130:637–50. https://doi.org/10.1016/j.enbuild.2016.08.071.

[17] Liu Z, Ye W, Little JC. Predicting emissions of volatile and semivolatile organic compounds from building materials: A review. Build Environ 2013; 64:7–25. https://doi.org/10.1016/j.buildenv.2013.02.012.

[18] Tong Z, Chen Y, Malkawi A, Adamkiewicz G, Spengler JD. Quantifying the impact of traffic-related air pollution on the indoor air quality of a naturally ventilated building. Environ Int 2016;89–90:138–46. https://doi.org/10.1016/j.envint.2016.01.016.

[19] Höllbacher E, Ters T, Rieder-Gradinger C, Srebotnik E. Emissions of indoor air pollutants from six user scenarios in a model room. Atmos Environ 2017; 150:389–94. https://doi.org/10.1016/j.atmosenv.2016.11.033.

[20] Amouei Torkmahalleh M, Gorjinezhad S, Unluevcek HS, Hopke PK. Review of factors impacting emission/concentration of cooking generated particulate matter. Sci Total Environ 2017; 586:1046–56. https://doi.org/10.1016/j.scitotenv.2017.02.088.

[21] Dai X, Liu J, Yin Y, Song X, Jia S. Modeling and controlling indoor formaldehyde concentrations in apartments: On-site investigation in all climate zones of China. Build Environ 2018; 127:98–106. https://doi.org/10.1016/j.buildenv.2017.10.036.

[22] Chen C, Zhao B. Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor. Atmos Environ 2011; 45:275–88. https://doi.org/10.1016/j.atmosenv.2010.09.048.

[23] Clements-Croome DJ, Awbi HB, Bako-Biro Z, Kochar N, Williams M. Ventilation Rates in

Schools. Build Environ 2008; 43:362–7.

[24] Waring MS, Wells JR. Volatile organic compound conversion by ozone, hydroxyl radicals, and nitrate radicals in residential indoor air: Magnitudes and impacts of oxidant sources. Atmos Environ 2015; 106:382–91. https://doi.org/10.1016/j.atmosenv.2014.06.062.

[25] Weschler CJ, Shields HC. Potential reactions among indoor pollutants. Atmos Environ 1997; 31:3487–95. https://doi.org/10.1016/S1352-2310(97)00219-7.

[26] Wainman T, Weschler CJ, Lioy PJ, Zhang J. Effects of surface type and relative humidity on the production and concentration of nitrous acid in a model indoor environment. Environ Sci Technol 2001; 35:2200–6. https://doi.org/10.1021/es000879i.

[27] Tichenor BA. Adsorption and Desorption of Pollutants to and from Indoor Surfaces. In: Pluschke P, editor. Air Pollut. Indoor Air Pollut., Berlin, Heidelberg: Springer Berlin Heidelberg; 2004, p. 73–87. https://doi.org/10.1007/b94831.

[28] Sousa G, Jones BM, Mirzaei PA, Robinson D. A review and critique of UK housing stock energy models, modelling approaches and data sources. Energy Build 2017; 151:66–80. https://doi.org/10.1016/j.enbuild.2017.06.043.

[29] Shrubsole C, Ridley I, Biddulph P, Milner J, Vardoulakis S, Ucci M, et al. Indoor PM2.5exposure in London’s domestic stock: Modelling current and future exposures following energy efficient refurbishment. Atmos Environ 2012; 62:336–43. https://doi.org/10.1016/j.atmosenv.2012.08.047.

[30] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew Sustain Energy Rev 2009; 13:1819–35. https://doi.org/10.1016/j.rser.2008.09.033.

[31] Jones B, Phillips G, O’Leary C, Molina C, Hall I. Diagnostic barriers to using PM2. 5 concentrations as metrics of indoor air quality. Smart Vent Build 2018.

[32] Coleman JR, Meggers F. Sensing of indoor air quality—characterization of spatial and temporal pollutant evolution through distributed sensing. Front Built Environ 2018; 4:1–12. https://doi.org/10.3389/fbuil.2018.00028.

[33] O’Leary C, de 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; 29:423–38. https://doi.org/10.1111/ina.12542.

[34] Nazaroff WW. Indoor particle dynamics. Indoor Air, Suppl 2004;14:175–83. https://doi.org/10.1111/j.1600-0668.2004.00286.x.

[35] Miller SL, Nazaroff WW. Environmental tobacco smoke particles in multizone indoor environments. Atmos Environ 2001; 35:2053–67. https://doi.org/10.1016/S1352-2310(00)00506-9.

[36] Nazaroff WW. Indoor particle dynamics. Indoor Air, Suppl 2004;14:175–83. https://doi.org/10.1111/j.1600-0668.2004.00286.x.

[37] Riley WJ, McKone TE, Lai ACK, Nazaroff WW. Indoor particulate matter of outdoor origin: Importance of size-dependent removal mechanisms. Environ Sci Technol 2002; 36:200–7. https://doi.org/10.1021/es010723y.

[38] Schneider T, Alstrup Jensen K, Clausen PA, Afshari A, Gunnarsen L, Wåhlin P, et al. Prediction of indoor concentration of 0.5-4 µm particles of outdoor origin in an uninhabited apartment. Atmos Environ 2004; 38:6349–59. https://doi.org/10.1016/j.atmosenv.2004.08.002.

[39] Dimitroulopoulou C, Ashmore MR, Hill MTR, Byrne MA, Kinnersley R. INDAIR: A probabilistic model of indoor air pollution in UK homes. Atmos Environ 2006; 40:6362–79. https://doi.org/10.1016/j.atmosenv.2006.05.047.

[40] Hussein T, Kulmala M. Indoor aerosol modeling: Basic principles and practical applications. Water, Air, Soil Pollut Focus 2008;8:23–34. https://doi.org/10.1007/s11267-007-9134-x.

[41] Jamieson S. A Partial Review of Mass Balance Models. Dep Epidemiol Biostat Imp Coll London 2008:1–9.

[42] IEHIAS. Indoor air pollution models | Integrated Environmental Health Impact Assessment System n.d. http://www.integrated-assessment.eu/eu/guidebook/indoor_air_pollution_models.html (accessed October 5, 2020).

[43] Yu Y, Megri AC, Jiang S. A review of the development of airflow models used in building load calculation and energy simulation. Build Simul 2019; 12:347–63. https://doi.org/10.1007/s12273-018-0494-0.

[44] Persily AK, Ivy EM. Input Data for Multizone Airflow and IAQ Analysis. 2001:42.

[45] Axley J. Multizone airflow modeling in buildings: History and theory. HVAC R Res 2007; 13:907–28. https://doi.org/10.1080/10789669.2007.10391462.

[46] Shimada M, Okuyama K, Okazaki S, Asai T, Matsukura M, Ishizu Y. Numerical Simulation and Experiment on the Transport of Fine Particles in a Ventilated Room. Aerosol Sci Technol 1996; 25:242–55. https://doi.org/10.1080/02786829608965394.

[47] Panagopoulos IK, Karayannis AN, Kassomenos P, Aravossis K. A CFD simulation study of VOC and formaldehyde indoor air pollution dispersion in an apartment as part of an indoor pollution management plan. Aerosol Air Qual Res 2011; 11:758–62. https://doi.org/10.4209/aaqr.2010.11.0092.

[48] Yang L, Ye M, he BJ. CFD simulation research on residential indoor air quality. Sci Total Environ 2014; 472:1137–44. https://doi.org/10.1016/j.scitotenv.2013.11.118.

[49] Haghighat F. Air infiltration and indoor air quality models — a review. Int J Ambient Energy 1989; 10:115–22. https://doi.org/10.1080/01430750.1989.9675130.

[50] Emmerich SJ, Hirnikel D. Validation of multizone IAQ modeling of residential-scale buildings: A review. ASHRAE Trans 2001;107 PART 2:619–28.

[51] Wei W, Ramalho O, Malingre L, Sivanantham S, Little JC, Mandin C. Machine learning and statistical models for predicting indoor air quality. Indoor Air 2019; 29:704–26. https://doi.org/10.1111/ina.12580.

[52] Ferguson L, Taylor J, Davies M, Shrubsole C, Symonds P, Dimitroulopoulou S. Exposure to indoor air pollution across socio-economic groups in high-income countries: A scoping review of the literature and a modelling methodology. Environ Int 2020; 143:105748. https://doi.org/10.1016/j.envint.2020.105748.

[53] Cheek E, Guercio V, Shrubsole C, Dimitroulopoulou S. Portable air purification: review of impacts on indoor air quality and health. Sci Total Environ 2020:142585. https://doi.org/10.1016/j.scitotenv.2020.142585.

[54] DEFRA (Deparment for Environment Food and Rural Affairs). Air Pollution in the UK. 2019.

[55] Birmingham City Council. 2019 Air Quality Annual Status Report (ASR) 2019.

[56] DEFRA (Deparment for Environment Food and Rural Affairs) & DfT (Department for Transport) UK. UK Plan for Tackling Roadside Nitrogen Dioxide Concentrations: Detailed Plan. 2017.

[57] Pasichnyi O, Wallin J, Kordas O. Data-driven building archetypes for urban building energy modelling. Energy 2019; 181:360–77. https://doi.org/10.1016/j.energy.2019.04.197.

[58] Chapman J. Data accuracy and model reliability. Bepac 1991:10–9.

[59] Depatment for Communities and Local Government. English Housing Survey: Technical

Report, 2017-2018. London, UK: 2017.

[60] HUD 1999b. American Housing Survey for the United States. H150/99. U.S. Department of Housing and Urban Development, U.S. Department of Commerce 2000.

[61] Molina C, Kent M, Hall I, Jones B. A data analysis of the Chilean housing stock and the development of modelling archetypes. Energy Build 2020; 206:109568. https://doi.org/10.1016/j.enbuild.2019.109568.

[62] Famuyibo AA, Duffy A, Strachan P. Developing archetypes for domestic dwellings - An Irish case study. Energy Build 2012; 50:150–7. https://doi.org/10.1016/j.enbuild.2012.03.033.

[63] Cerezo C, Sokol J, Reinhart C, Al-Mumin A. Three methods for characterizing building archetypes in urban energy simulation. A case study in Kuwait City. 14th Conf Int Build Perform Simul Assoc 2015:2873–80. https://doi.org/10.1523/JNEUROSCI.4572-10.2011.

[64] Ben H, Steemers K. Household archetypes and behavioural patterns in UK domestic energy use. Energy Effic 2018; 11:761–71. https://doi.org/10.1007/s12053-017-9609-1.

[65] BRE. Energy Follow-Up Survey 2011, Report 2: Mean Household Tempeatures 2013.

[66] Jones B, Das P, Chalabi Z, Davies M, Hamilton I, Lowe R, et al. Assessing uncertainty in housing stock infiltration rates andassociated heat loss: English and UK case studies. Build Environ 2015; 92:644–56. https://doi.org/10.1016/j.buildenv.2015.05.033.

[67] Stephens R. Airtighness in UK Dwellings. Watford, UK: IHS BRE Press; 2000.

[68] Chan WR, Nazaroff WW, Price PN, Sohn MD, Gadgil AJ. Analyzing a database of residential air leakage in the United States. Atmos Environ 2005; 39:3445–55. https://doi.org/10.1016/j.atmosenv.2005.01.062.

[69] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018; 81:1192–205. https://doi.org/10.1016/j.rser.2017.04.095.

[70] ASHRAE. ANSI/ASHRAE Standard 62.1-2019 Ventilation for Acceptable Indoor Air Quality. Atlanta, GA: 2019.

[71] ASHRAE. ANSI/ASHRAE Standard 62.2-2019 Ventilation and Acceptable Indoor Air Quality in Residential Buildings. American Society of Heating, Refrigerating, and Air-Conditioning Engineers; 2019.

[72] CIBSE. Environmental design CIBSE Guide A. 2018.

[73] Persily A, Musser A, Leber D. A collection of homes to represent the U.S. housing stock. Gaithersburg, MD: 2006. https://doi.org/10.6028/NIST.IR.7330.

[74] Sokol J, Cerezo Davila C, Reinhart CF. Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. Energy Build 2017; 134:11–24. https://doi.org/10.1016/j.enbuild.2016.10.050.

[75] Ghiassi N, Mahdavi A. Reductive bottom-up urban energy computing supported by multivariate cluster analysis. Energy Build 2017; 144:372–86. https://doi.org/10.1016/j.enbuild.2017.03.004.

[76] Wate P, Coors V, Iglesias M, Robinson D. Uncertainty assessment of building performance simulation. Urban Energy Syst. Low-Carbon Cities, Elsevier; 2019, p. 257–87. https://doi.org/10.1016/B978-0-12-811553-4.00007-X.

[77] Wate P, Iglesias M, Coors V, Robinson D. Framework for emulation and uncertainty quantification of a stochastic building performance simulator. Appl Energy 2020; 258:113759. https://doi.org/10.1016/j.apenergy.2019.113759.

[78] Lim H, Zhai ZJ. Review on stochastic modeling methods for building stock energy prediction.

Build Simul 2017; 10:607–24. https://doi.org/10.1007/s12273-017-0383-y.

[79] Hensen J, Lamberts R. Building Performance Simulation for Design and Operation. New York: Spon Press, 2011; 2011.

[80] Robinson S. Conceptual modelling for simulation Part I: Definition and requirements. J Oper Res Soc 2008; 59:278–90. https://doi.org/10.1057/palgrave.jors.2602368.

[81] Megri AC, Haghighat F. Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications. HVAC R Res 2007; 13:887–905. https://doi.org/10.1080/10789669.2007.10391461.

[82] Wang L, Chen Q. Analysis on the well-mixing assumptions used in multizone airflow network models. IBPSA 2007 - Int Build Perform Simul Assoc 2007 2007:1485–90.

[83] Conte SD, de Boor C. Elementary Numerical Analysis: An Algorithmic Approach. New York: McGraw-Hill: 1972.

[84] W.S. Dols BJP. NIST Technical Note 1887 CONTAM User Guide and Program Documentation 2007. https://doi.org/10.1071/PP98146.

[85] 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. Build Environ 2014; 78:171–82. https://doi.org/10.1016/j.buildenv.2014.04.017.

[86] Molina C, Jones B, Hall IP, Sherman MH. CHAARM: A model to predict uncertainties in indoor pollutant concentrations, ventilation and infiltration rates, and associated energy demand in Chilean houses. Energy Build 2020:110539. https://doi.org/10.1016/j.enbuild.2020.110539.

[87] Renard P, Alcolea A, Ginsbourger D. Stochastic versus Deterministic Approaches. Environ. Model., Chichester, UK: John Wiley & Sons, Ltd; 2013, p. 133–49. https://doi.org/10.1002/9781118351475.ch8.

[88] Cóstola D, Blocken B, Hensen JLM. Uncertainties due to the use of surface averaged wind pressure coefficients. Proc 29th AIVC Conf Kyoto, Japan 2008:14–6.

[89] Herring SJ, Batchelor S, Bieringer PE, Lingard B, Lorenzetti DM, Parker ST, et al. Providing pressure inputs to multizone building models. Build Environ 2016; 101:32–44. https://doi.org/10.1016/j.buildenv.2016.02.012.

[90] Yan D, O'Brien W, Hong T, Feng X, Burak Gunay H, Tahmasebi F, et al. Occupant behavior modeling for building performance simulation: Current state and future challenges. Energy Build 2015; 107:264–78. https://doi.org/10.1016/j.enbuild.2015.08.032.

[91] Booth AT, Choudhary R, Spiegelhalter DJ. Handling uncertainty in housing stock models. Build Environ 2012; 48:35–47. https://doi.org/10.1016/j.buildenv.2011.08.016.

[92] Malkawi A, Augenbroe G. Advanced building simulation. 1st editio. Spon Press; 2004.

[93] Dias D, Tchepel O. Spatial and temporal dynamics in air pollution exposure assessment. Int J Environ Res Public Health 2018;15. https://doi.org/10.3390/ijerph15030558.

[94] Jones B, Das P, Chalabi Z, Davies M, Hamilton I, Lowe R, et al. Assessing uncertainty in housing stock infiltration rates andassociated heat loss: English and UK case studies. Build Environ 2015; 92:644–56. https://doi.org/10.1016/j.buildenv.2015.05.033.

[95] Hyun SH, Park CS, Augenbroe GLM. Analysis of uncertainty in natural ventilation predictions of high-rise apartment buildings. Build Serv Eng Res Technol 2008; 29:311–26. https://doi.org/10.1177/0143624408092424.

[96] Symonds P, Taylor J, Chalabi Z, Mavrogianni A, Davies M, Hamilton I, et al. Development of an England-wide indoor overheating and air pollution model using artificial neural networks. J

Build Perform Simul 2016; 9:606–19. https://doi.org/10.1080/19401493.2016.1166265.

[97] Fazli T, Stephens B. Development of a nationally representative set of combined building energy and indoor air quality models for U.S. residences. Build Environ 2018; 136:198–212. https://doi.org/10.1016/J.BUILDENV.2018.03.047.

[98] Mavrogianni A, Davies M, Taylor J, Chalabi Z, Biddulph P, Oikonomou E, et al. The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments. Build Environ 2014; 78:183–98. https://doi.org/10.1016/j.buildenv.2014.04.008.

[99] Bouvier JL, Bontemps S, Mora L. Uncertainty and sensitivity analyses applied to a dynamic simulation of the carbon dioxide concentration in a detached house. Int J Energy Environ Eng 2019; 10:47–65. https://doi.org/10.1007/s40095-018-0291-7.

[100] Feustel HE. COMIS-an international multizone air-flow and contaminant transport model. Energy Build 1999; 30:3–18. https://doi.org/10.1016/S0378-7788(98)00043-7.

[101] Dols WS, Polidoro BJ. CONTAM User Guide and Program Documentation Version 3.2. Tech. rept. NIST; 2015.

[102] Haghighat F. A comprehensive validation of two airflow models - COMIS and CONTAM. Indoor Air 1996; 6:278–88. https://doi.org/10.1111/j.1600-0668.1996.00007.x.

[103] Walton GN. AIRNET - A Computer Program for Building Airflow Network Modelling. Nistir 89-4072 1989:77.

[104] McDowell TP, Emmerich S, Thornton JW, Walton G. Integration of airflow and energy simulation using CONTAM and TRNSYS. ASHRAE Trans 2003;109 PART 2:757–70. https://doi.org/10.1107/S1600536812032953.

[105] Axley JW. Indoor Air Quality Modeling Phase II Report. National Bureau of Standards (U.S.); 1987.

[106] Axley J. Progress Toward a General Analytical Method for Predicting Indoor Air Pollution in Buildings, Indoor Air Quality Modelling Phase III Report. National Bureau of Standards (U.S.); 1988.

[107] Underhill LJ, Fabian MP, Vermeer K, Sandel M, Adamkiewicz G, Leibler JH, et al. Modeling the resiliency of energy-efficient retrofits in low-income multifamily housing. Indoor Air 2018; 28:459–68. https://doi.org/10.1111/ina.12446.

[108] García-Tobar J. Weather-dependent modelling of the indoor radon concentration in two dwellings using CONTAM. Indoor Built Environ 2019; 28:1341–9. https://doi.org/10.1177/1420326X19841119.

[109] Office of Energy Efficient and Renewable Energy. EnergyPlus | Department of Energy. ENERGYGOV n.d. https://www.energy.gov/eere/buildings/downloads/energyplus-0 (accessed February 26, 2020).

[110] Dols WS, Emmerich SJ, Polidoro BJ. Coupling the multizone airflow and contaminant transport software CONTAM with EnergyPlus using co-simulation. Build Simul 2016; 9:469–79. https://doi.org/10.1007/s12273-016-0279-2.

[111] Taylor J, Shrubsole C, Biddulph P, Jones B, Das P, Davies M. Simulation of pollution transport in buildings: The importance of taking into account dynamic thermal effects. Build Serv Eng Res Technol 2014; 35:682–90. https://doi.org/10.1177/0143624414528722.

[112] Logue JM, Price PN, Sherman MH, Singer BC. A method to estimate the chronic health impact of air pollutants in U.S. residences. Environ Health Perspect 2012; 120:216–22. https://doi.org/10.1289/ehp.1104035.

[113] Oikonomou E, Davies M, Mavrogianni A, Biddulph P, Wilkinson P, Kolokotroni M.

Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London. Build Environ 2012; 57:223–38. https://doi.org/10.1016/j.buildenv.2012.04.002.

[114] The Geoinformation Group. National Building Class Database. Cambridge, UK, The Geoinformation Group: 2013.

[115] BRE. SAP: The Government's Standard Assessment Procedure for Energy Rating of Dwellings. Watford, UK, Building Research Establishment.: 2009.

[116] Taylor J, Davies M, Mavrogianni A, Shrubsole C, Hamilton I, Das P, et al. Mapping indoor overheating and air pollution risk modification across Great Britain: A modelling study. Build Environ 2016; 99:1–12. https://doi.org/10.1016/j.buildenv.2016.01.010.

[117] Taylor J, Shrubsole C, Symonds P, Mackenzie I, Davies M. Application of an indoor air pollution metamodel to a spatially-distributed housing stock. Sci Total Environ 2019; 667:390–9. https://doi.org/10.1016/j.scitotenv.2019.02.341.

[118] CIBSE. Environmental Design. London, UK, Chartered Institution of Building Services Engineers.: 2006. https://doi.org/10.1016/b978-0-323-02948-3.00016-x.

[119] Molina C, Jones B, Hall IP. R Code 2020:1–40.

[120] O'Leary C, Jones B, Dimitroulopoulou S, Hall IP. Setting the standard: The acceptability of kitchen ventilation for the English housing stock. Build Environ 2019; 166:106417. https://doi.org/10.1016/j.buildenv.2019.106417.

[121] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014; 37:517–37. https://doi.org/10.1016/j.rser.2014.05.056.

[122] Morrison R. Energy system modeling: Public transparency, scientific reproducibility, and open development. Energy Strateg Rev 2018; 20:49–63. https://doi.org/10.1016/j.esr.2017.12.010.

[123] Li J, Liu L, Le TD, Liu J. Accurate data-driven prediction does not mean high reproducibility. Nat Mach Intell 2020; 2:13–5. https://doi.org/10.1038/s42256-019-0140-2.

[124] Van Rossum G. Python Reference Manual 2001. https://docs.python.org/3/reference/.

[125] Oladokun MG, Odesola IA. Household energy consumption and carbon emissions for sustainable cities – A critical review of modelling approaches. Int J Sustain Built Environ 2015; 4:231–47. https://doi.org/10.1016/j.ijsbe.2015.07.005.

[126] Chelani AB, Devotta S. Nonlinear Analysis and Prediction of Coarse Particulate Matter Concentration in Ambient Air. J Air Waste Manag Assoc 2006; 56:78–84. https://doi.org/10.1080/10473289.2006.10464432.

[127] Chen Y, Gu L, Zhang J. EnergyPlus and CHAMPS-Multizone co-simulation for energy and indoor air quality analysis. Build Simul 2015; 8:371–80. https://doi.org/10.1007/s12273-015-0211-1.

[128] Heibati S, Maref W, Saber HH. Assessing the energy and indoor air quality performance for a three-story building using an integrated model, part one: The need for integration. Energies 2019;12. https://doi.org/10.3390/en12244775.

[129] Ng LC, Musser A, Persily AK, Emmerich SJ. Multizone airflow models for calculating infiltration rates in commercial reference buildings. Energy Build 2013; 58:11–8. https://doi.org/10.1016/j.enbuild.2012.11.035.

[130] Underhill LJ, Dols WS, Lee SK, Fabian MP, Levy JI. Quantifying the impact of housing interventions on indoor air quality and energy consumption using coupled simulation models. J Expo Sci Environ Epidemiol 2020; 30:436–47. https://doi.org/10.1038/s41370-019-0197-3.

[131] 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 Build 2017; 154:321–34. https://doi.org/10.1016/j.enbuild.2017.08.029.