

To appear in the *Journal of Building Performance Simulation*
Vol. 00, No. 00, Month 20XX, 1–18

Development of an England-wide indoor overheating and air pollution model using artificial neural networks

Phil Symonds^{a*}, Jonathon Taylor^a, Zaid Chalabi^{a,b}, Anna Mavrogianni^a, Michael Davies^a, Ian Hamilton^a,
Sotiris Vardoulakis^c, Clare Heaviside^c, Helen Macintyre^c

^a*Institute of Environmental Design and Engineering, University College London, Central House, 14 Woburn Place, London, WC1H 0NN, UK;* ^b*London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, UK;* ^c*Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Didcot, UK*

(v2 released February 2016)

With the UK climate projected to warm in future decades, there is an increased research focus on the risks of indoor overheating. Energy-efficient building adaptations may modify a buildings risk of overheating and the infiltration air pollution from outdoor sources. This paper presents the development of a national model of indoor overheating and air pollution, capable of modelling the existing and future building stocks, along with changes to the climate, outdoor air pollution levels, and occupant behaviour. The model presented is based on a large number of EnergyPlus simulations run in parallel. A metamodelling approach is used to create a model that estimates the indoor overheating and air pollution risks for the English housing stock. The performance of Neural Networks (NNs) is compared to a Support Vector Regression (SVR) algorithm when forming the metamodel. NNs are shown to give almost a 50% better overall performance than SVR.

Keywords: metamodelling; machine learning; neural networks; stock modelling; overheating; indoor air pollution

1. Introduction

The role of housing in population exposure to environmental hazards is an important area of research in the UK due to the fact that the population spends around 90% of its time indoors (ONS 2005; Kornartit et al. 2010). Two areas of particular interest are indoor overheating and air pollution exposure which are predicted to change due to the warming climate and the need to improve housing energy efficiency, and therefore airtightness, to meet greenhouse gas reduction targets.

A warming climate and an increase in heatwave events is predicted to lead to greater heat-related mortality in England (Hajat et al. 2014). Recent history shows the health impacts of hot periods, with the 2003 heatwave in Europe estimated to have contributed to over 2,000 excess deaths across England and Wales (Johnson et al. 2005). Heatwave events similar to this are projected to become commonplace by the 2040s (Murphy et al. 2009). Recent work has examined the issue of overheating using both indoor monitoring (e.g. in Beizaee, Lomas, and Firth (2013); Wright, Young, and Natarajan (2005); Lomas and Kane (2012); Mavrogianni et al. (2010)) and modelling approaches (e.g. in Mavrogianni et al. (2012); Oikonomou et al. (2012); Hamdy and Hensen ((2015); Taylor et al. (2015a); Porritt and Cropper (2010); Peacock, Jenkins, and Kane (2010); Gupta and Gregg (2013); Taylor et al. (2015b, 2016)), with results indicating a wide range in overheating risks

*Corresponding author. Email: p.symonds@ucl.ac.uk

with top-floor flats, bungalows, and those with poor ventilation at increased risk. Overheating in UK homes has been comprehensively reviewed by the Zero Carbon Hub (ZCH 2015b).

Indoor air pollution is another health risk that may be influenced by environmental and dwelling characteristics. Indoor air quality is inherently coupled to indoor temperatures due to the need for ventilation during hot weather (Taylor et al. 2015a) which is typically achieved through increased window opening in the predominantly naturally-ventilated UK housing stock. Health risks due to air pollution are significant: a study by Gowers, Miller, and Stedman (2014) showed that the mortality burden attributable to human-made particulate air pollution was equivalent to approximately 25,000 deaths or around 5.6% of all-cause adult mortality in England in 2010. Particles smaller than 2.5 micrometres ($PM_{2.5}$) are thought to cause around 3,500 premature deaths a year in London (Walton et al. 2010), and is acknowledged to play an important role in air quality and health. Modelling (Shrubsole et al. 2015; Taylor et al. 2014b; Shrubsole et al. 2012) and monitoring studies (Nasir and Colbeck 2013; Jones et al. 2000; Lai et al. 2006; Hoek et al. 2008) suggest that the levels of indoor air pollution in UK dwellings are affected by various building, environmental, and occupancy characteristics. There is, therefore, a growing need to understand how climate, housing stock, and population changes may alter population exposure to both overheating and indoor air pollution in England under a range of different scenarios (Vardoulakis et al. 2015).

This paper describes the development of an overheating, outdoor pollution infiltration, and energy use model for the English housing stock, to be applied in epidemiological research. The model must be capable of modelling the existing stock, energy-efficiency improvements to the stock, the construction of new buildings, changes to climate and outdoor pollution levels, and occupant adaptations to hot weather. There are a large number of parameters which may influence indoor overheating risks, and which need to be accounted for in the model. Previous research shows that indoor temperatures in England and the UK may be influenced by; building geometry and orientation; fabric characteristics such as the thermal mass and conductivity of walls, windows, floors, and roofs (Mavrogianni et al. 2012); natural or purpose-provided ventilation (Taylor et al. 2015a); occupant behaviour (Mavrogianni et al. 2013); adaptation measures (Porritt and Cropper 2010; Gupta and Gregg 2013); and climate, including region (Taylor et al. 2014a) or locations within an Urban Heat Island (UHI) (Oikonomou et al. 2012). In order to model all of these influences, a large set of building physics simulations are required. EnergyPlus (US-DoE. 2013) is a building physics software capable of modelling all of these variables. EnergyPlus was initially intended for use as an application to be run on a desktop computer. The computational challenges faced when using EnergyPlus as a tool for the simulation of thousands of scenarios in parallel have been highlighted by Sanyal and New (2013). We harness the power of parallel computing in order to streamline the model development process.

Health-relevant performance metrics and the energy-use of domestic buildings are calculated using the outputs produced by EnergyPlus. A metamodelling framework is created to enable the rapid estimate of these performance metrics for the current English housing stock and potential future stocks under a range of scenarios. Replacing a run time intensive physics model with a quick approximate model allows sensitivity analyses and multi-criteria optimisation to be performed. Metamodelling techniques have been shown to give reliable estimates of complex models (Barton and Meckesheimer 2006), and have been used in previous studies to enable the rapid calculation of building performance estimates from EnergyPlus simulations (Van Gelder et al. 2014; Eisenhower et al. 2012; De Wilde and Tian 2010; New et al. 2012). We compare the performance of Neural Networks (NNs) and Support Vector Regression (SVR) machine learning algorithms for several domestic building and occupancy types.

2. Modelling Methods

The development of the model involves several steps as shown in Figure 1. These steps include the random sampling of input parameters, EnergyPlus input file generation, building physics sim-

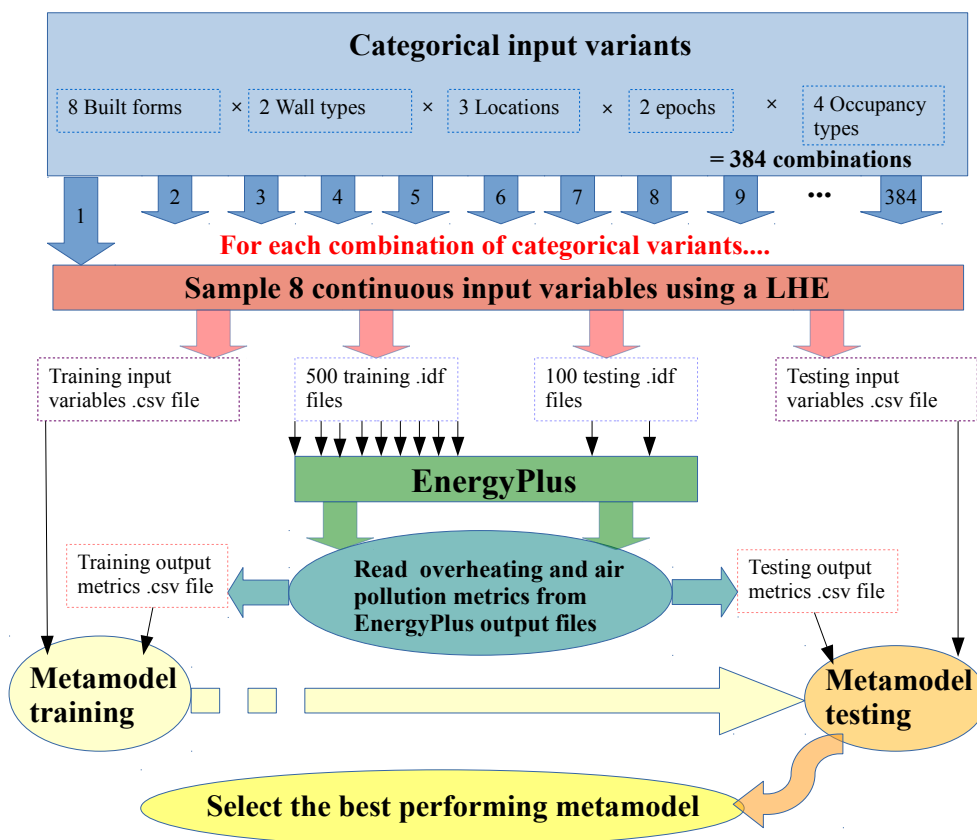


Figure 1. Workflow for the development of a national indoor overheating, air quality and energy use model using High Performance Computing (HPC). The first step involves the generation of EnergyPlus input definition files (.idfs). For each combination of categorical variant, independent sets of 500 and 100 idfs are produced for training and testing, respectively. Eight continuous building and occupancy relevant variables are sampled using a latin hypercube experimental (LHE) design for each idf. The idfs are then simulated with EnergyPlus in parallel sets of 25 simulations. The next step is to calculate the relevant metric from the EnergyPlus output files. Finally, the training and then the testing of candidate metamodels is performed. The best performing metamodel is selected for the inclusion in the metamodeling framework. All model development steps are performed in parallel for each combination of the categorical variants using HPC.

ulation, the calculation of health-relevant and energy use output metrics, and the training and testing of candidate metamodels. Due to the large scale of the model, it is built entirely on the University College London (UCL) high performance computing (HPC) resource, Legion which has 7,500 CPU cores. This allows many of the steps shown in Figure 1 to be run in parallel saving vast amounts of time. A set of metamodels (metamodel framework) is created rather than a large individual metamodel encompassing all categorical variants. An advantage of this, is we are able to exploit the use of HPC to build individual metamodels for each combination of the categorical variants (built form, wall type, occupancy type and location). A set of 384 (8 building archetypes \times 4 occupancy types \times 2 wall types \times 3 locations \times 2 epochs) metamodels are therefore produced. Producing metamodels individually also makes the model scalable such that it can be expanded to include additional locations, occupancies or building types without having to rebuild the entire model from scratch.

At the base of the metamodeling framework are 230,400 simulations run using EnergyPlus (US-DoE. 2013). A set of 500 simulations are used for the training of each individual metamodel based on convergence results shown in section 3. An independent sample of 100 simulations is used to test the goodness fit of each trained metamodel. EnergyPlus is a dynamic thermal physics simulation tool with a generic contaminant model that is able to model the movement of harmful particulates

within the building. EnergyPlus has been used in previous studies to model both hourly indoor temperatures (Mavrogianni et al. 2012; Oikonomou et al. 2012; Taylor et al. 2015a,b) and indoor air pollutant levels (Taylor et al. 2014b, 2015a; Shrubsole et al. 2015) for dwellings representative of the English housing stock. A Red Hat Enterprise Linux 5 (RHEL5) version of EnergyPlus 8.1 provided by the United States - Department of Energy (US-DoE) enables the simulations to be performed on Legion. Each EnergyPlus simulation takes between five and 25 minutes of CPU time to complete depending on the complexity of the archetype with high rise flats taking the longest. HPC enabled 19,200 building physics simulations to be run within about 5 hours by parallel processing 768 sets of 25 simulations.

EnergyPlus simulation input definition files (idf) are generated using an in-house tool, EnergyPlus Generator 2 (EPG2); written using the Python programming language (Van Rossum and eds). EPG2 is designed for the batch generation of building input files based on user-defined variables. The tool is able to generate a few hundred idf files in several minutes CPU time. Time can again be saved by producing idf files in parallel for each building archetype. The EPG2 tool is supplemented with the pyDoE (Baudin 2015) package (an experimental design package for Python) to enable the selection of input variable values based on a Latin Hypercube Experimental (LHE) design. LHE is a space-filling sampling method that allows for the random generation of various parametric values, according to underlying distributions and ensures that similar runs are never repeated (Tang 2012). A LHE design was previously used for building simulation modelling, for example, to examine indoor thermal comfort with EnergyPlus (De Wilde and Tian 2010) and indoor air quality in CONTAM (Das et al. 2014). An advantage of the LHE sampling technique is that different distributions can be used for different input variables. If a certain input parameter's range and shape of probability distribution is well known, a specific distribution can be chosen for sampling. Otherwise, a uniform distribution or a normal distribution with a large standard deviation is chosen. The variable ranges are designed to be large such that the entire existing English housing stock database and future retrofit scenarios can be represented. Correlations between input parameters using the current building stock are not included since the model is designed to be adaptable to include various future building stock scenarios.

2.1. Model Inputs

In this section, the input variables and the assumptions used in the EnergyPlus simulations are described. The model's input parameters can be categorical, continuous, or discrete in nature. Individual metamodels are created for each combination of the categorical variants, whilst the continuous and discrete variants are randomly sampled within the creation of each metamodel. The input parameters cover a range of building, environmental, and occupancy variants described in sections 2.1.1 and 2.1.2.

2.1.1. Categorical Variants

The categorical variants were chosen to cover a large range of housing types, external environments, and occupancy behaviours such that the majority of cases in England and future England are covered. The full range of categorical variants is shown in Table 1. The number of combinations of these categorical variants yields the creation of 384 individual metamodels.

Building archetypes representative of the English housing stock are used as inputs to the model. Eight building types are used as defined in the English Housing Survey (EHS) (DCLG 2011) as shown in Table 1. These built forms were developed by Oikonomou et al. (2016) using information on floor areas, window sizes, and ceiling heights from the EHS, and typical floor plans from architectural drawings. Whole buildings are modelled, with flats containing ground, middle, and top-floor levels within each building. Geometric and layout variations within built forms are not considered due to the complexity of producing EnergyPlus input files with different geometries for all zones within the building. More information on the archetypes is given in the supplementary

Table 1. Categorical input variables with the corresponding categories modelled.

Parameter	Number of Variants	Categories
Built Form	8	End Terrace/Mid Terrace/ Semi-Detached/Detached/Bungalow/ Converted Flat/Low-rise Flat/High-rise Flat
Wall Type	2	Solid/Cavity
Location	3	London/Birmingham/Newcastle
Epoch ^a	2	Present/2050
Occupancy Type	2	Pensioners/Family
Retrofit	2	With/Without Shutters

^aWeather scenarios are modelled using data produced by the PROMETHEUS project (Eames, Kershaw, and Coley 2010).

material (section 1).

The construction of the wall type and placement of the insulation can influence the overheating risk (Mavrogianni et al. 2012). We model wall type as a binary categorical variant, selecting from either solid or cavity walls in order to account for the differences in thermal mass between the two types. In the case of solid walls, those with a U-value below $0.6 \text{ W/m}^2\text{K}$ are assumed to have internal solid wall insulation. This represents a worst case scenario for retrofit, as internal solid-wall insulation is linked to an increase in potential overheating risk in some studies (Peacock, Jenkins, and Kane 2010).

Trickle vents in windows have been required in UK Building regulations since 2002 in order to allow for compensatory ventilation (HM-Government 2010); their presence may influence building ventilation rate and consequently indoor air pollution and overheating risk (Taylor et al. 2015a). All windows where U-values are representative of those post-2002 (window U-value $\leq 2.0 \text{ W/m}^2\text{K}$), are therefore modelled with trickle vents installed.

Input weather data for England has been selected based on previous research which indicates that London, South and Central England, and Northern England have climates which result in statistically significant differences in relative overheating risk amongst dwellings (Taylor et al. 2014a). These climates can therefore be modelled using hourly weather data from a recent hot summer year, produced by the UKCP09 data-based PROMETHEUS project (Eames, Kershaw, and Coley 2010) for London, Birmingham, and Newcastle, respectively. Simulations will also be performed under future high CO_2 emission scenarios which will lead to a warmer climate. Metamodels can be trained on various locations and climates such that the whole of England is covered.

A buildings' exposure to the sun is specified by the presence of adjoining and neighbouring dwellings. Houses are modelled with adjoining surfaces according to the built form when it is implied (e.g. mid-terrace, semi-detached, end-terrace and flats). Bungalows are modelled as being detached. Flats are modelled with adjoining dwellings on either side and above and/or below depending on the level.

We model the infiltration of outdoor $\text{PM}_{2.5}$ pollution into the indoor environment using the approach of Taylor et al. (2014b). $\text{PM}_{2.5}$ is modelled with an external concentration of $9.8 \mu\text{g}/\text{m}^3$, reflecting the UK average. Infiltration of outdoor $\text{PM}_{2.5}$ is modelled with a penetration factor of 0.8 when windows are closed and 1.0 when windows are open (Long et al. 2001), representing the deposition of the particulates inside cracks as it enters the dwellings. Once inside, $\text{PM}_{2.5}$ is modelled with a deposition rate of 0.18 hour^{-1} (Long et al. 2001). Pollution from indoor sources is not modelled due to the inability of the current EnergyPlus Generic Contaminant Model to simulate multiple pollutants.

Occupant behaviour has a significant impact on the risk of overheating within dwellings through ventilation and shading behaviour, occupied hours, and internal gains (Mavrogianni et al. 2013; Porritt and Shao 2010). Previous studies, for example (Taylor et al. 2014a), used occupancy scenarios representing a family out during the day or two pensioners home all day. In addition, there are a number of preventative measures occupants may take to reduce risk, with a number of studies suggesting that installing external shutters may be the most effective means of reducing indoor temperatures (Porritt and Cropper 2010; Gupta and Gregg 2013; Porritt et al. 2010). Therefore,

four occupant-behaviour scenarios are modelled:

- (1) Two pensioners, spending all day at home, who open windows during the summer months (May-September);
- (2) Same as occupancy type-1 but with external shutters closed from 9am to 6pm during summer months, with lights turned on when shutters are closed;
- (3) A family of five, out all day (8am–6pm), who open windows when at home, and with no adaptive shading;
- (4) Same as occupancy type-3 but with external shutters closed during work hours (8am–6pm) in the summer.

In order to model the impact of changes in the building fabric on energy consumption, space heating is modelled year-round. Simple electric radiators, sized for rooms based on (CIBSE 2003) guidelines (see section 2 of the supplementary material), heat rooms to the temperature setpoint for 9 hours on weekdays (6am-8am, 5pm-12pm) and 16 hours on weekends (7am-11pm) (BRE 2009). Internal heat gains from individual occupants, lighting, cooking, and other electrical equipment are prescribed according to building occupancy schedules. More information on internal gains and the occupancy schedules used is given in the supplementary material (section 3).

2.1.2. Continuous and Discrete Variants

The shapes and ranges of the continuous distributions are chosen in order to cover the full English housing stock and future stocks. Sufficient simulation data was produced in order for the meta-models to be able to make meaningful estimations across the full range of input variants shown in Table 2.

Table 2. Continuous and discrete input variables and the corresponding value ranges sampled from during metamodel training. The continuous variables are randomly sampled from; Truncated Normal (TN) with mean (μ) and standard deviation (σ), and Uniform (U) distributions.

Parameter	Range	Distribution
Wall U-value	0.15-2.55 W/m ² K	U
Roof U-value	0.10-2.25 W/m ² K	U
Window U-value	0.85-4.80 W/m ² K	U
Floor U-value	0.15-1.30 W/m ² K	U
Fabric air permeability	0-∞ m ³ /h/m ² @ 50 Pa	TN($\mu = 20, \sigma = 10$)
Orientation	0-360°	U
Terrain Type	City/Urban/Rural	Discrete
Occupant window opening temperature threshold	10-∞ °C	TN($\mu = 24, \sigma = 5$)
Occupant thermostat setting	15-26 °C	TN($\mu = 22, \sigma = 3$)

Built forms are simulated with building fabrics (windows, walls, roofs, and floors) sampled from a uniform range of U-values representing the thermal performance of the element shown in Table 2. We use the Governments Standard Assessment Procedure (SAP) (BRE 2015) for Energy Rating of Dwellings which uses the age, fabric type, and geometry of buildings to estimate the range of U-values expected for fabric elements in the existing English housing stock. This range has been extended to represent minimum U-values required to meet Passivhaus standards (PassivHaus 2015). This will allow us to explore the potential of future, strict building regulations. U-values are implemented in EnergyPlus by adjusting the thermal resistance and/or thickness of the composite construction materials. Window solar transmittance is adjusted according to window U-value, assuming a positive linear relationship (BRE 2015).

Building air change rate is modelled by assigning a permeability (m³/h/m² @ 50 Pa) to external building fabric elements. Modelling air infiltration due to envelope permeability, enables us to account for the role of wind exposure on air change rate and therefore pollutant infiltration. The range of building permeabilities is chosen based on the distribution of values found in the English housing stock, as described by Stephen (2000). The buildings orientation is randomly selected from a uniform distribution in the range 0-360 degrees East of North.

Wind exposures are attributed to buildings in terms of shelter and shading from nearby buildings. EnergyPlus allows for the terrain of the buildings surroundings to be defined as city, urban, or rural. This impacts on the exposure of the building to wind, and therefore the background ventilation rate and pollution infiltration. Although terrain is a categorical variant, it is trained within individual metamodels as a discrete variant with city, urban, and rural being assigned values of minus one, zero, and plus one, respectively. This enabled the number of EnergyPlus simulations required to be reduced by a factor of three.

The thermal comfort of occupants is modelled in terms of the window opening threshold during the summer (May-September) and the thermostat setpoint during the winter (October-April). These values are both randomly sampled according to a truncated normal distribution within the ranges specified in Table 2.

2.2. Building Performance Output Metrics

EnergyPlus output files are analysed using a Python script which calculates the overheating, indoor air pollution, and energy-use metrics. Several health-relevant indoor environment variables and an energy use metric are modelled as shown in Table 3.

Table 3. Health-relevant and energy use output metrics with corresponding symbols, units and the zones for which they are calculated. I/O and RH are an average over the living room and bedroom for the whole year, whilst E_{heat} is a sum over all rooms with heaters installed.

Parameter	Symbol	Unit	Zones
Overheating metric ^a	T_{OH}	$^{\circ}\text{C}$	Liv, Bed
PM _{2.5} Indoor/Outdoor ratio	I/O	NA	Mean(Liv, Bed)
Relative Humidity	RH	%	Mean(Liv, Bed)
Annual Heating Energy Use	E_{heat}	kWh	Sum(All Heated Zones)

^a The overheating metric is calculated during the day (8am-10pm) and at night(10pm-8am) for both the living room and the master bedroom.

2.2.1. Overheating metric (T_{OH})

There are a number of different indoor overheating metrics available based on research from different fields (Carlucci and Pagliano 2012; Hamdy and Hensen (2015; ZCH 2015a). We have elected to use a health-relevant metric developed by Armstrong et al. (2011), and implemented in the indoor environment by Taylor et al. (2015b), where it was used to assess the overheating risk on excess heat-related mortality for London. This metric can be used to quantify the indoor overheating risk of a home when outdoor temperatures exceed regional mortality thresholds ($T_{\text{region}}^{\text{thresh.}}$). This threshold is defined as the outdoor temperature above which excess overheating related deaths are observed. In Birmingham, for example, this threshold is 23°C (Armstrong et al. 2011). The metric is calculated as the mean of the maximum indoor temperature ($T_{\text{in},d}^{\text{max}}$), on days, d , where the two day mean of the maximum outdoor temperature ($T_{\text{ext}}^{\text{max}}$) exceeds $T_{\text{region}}^{\text{thresh.}}$:

$$T_{\text{OH}} = \frac{\sum_{d=1 \dots N_d} T_{\text{in},d}^{\text{max}}}{N_d} \quad (1)$$

Here, $N_d = \sum I(T_{\text{ext}}^{\text{max}} > T_{\text{region}}^{\text{thresh.}})$ where I is the indicator function, with $I(A) = 1$ if A is true, and $I(A) = 0$ if A is false. $T_{\text{ext}}^{\text{max}}$ is calculated using days, d and $d - 1$. This metric is calculated for the living room and the bedroom during both the day time (8am-10pm) and the night time (10pm-8am).

2.2.2. $PM_{2.5}$ indoor/outdoor ratio (I/O)

The estimated absolute concentration of indoor $PM_{2.5}$ from outdoor sources is modelled by EnergyPlus and converted to an hourly indoor/outdoor (I/O) ratio during post-processing. The average of the living room and master bedroom is calculated. I/O ratios are used, as they allow indoor air pollution levels to be estimated based on spatially or temporally-varying levels of outdoor pollution.

2.2.3. Relative Humidity (RH)

Relative humidity is included as an additional health-relevant metric as it can be used to estimate the mould severity index (MSI). MSI is a metric which can be used to assess the damage of moisture on buildings and also the health impact caused to asthma sufferers (Fisk, Lei-Gomez, and Mendell 2007). The mean RH is calculated for the living room and master bedroom.

2.2.4. Energy Use (E_{heat})

We calculate annual energy use for space heating as a sum over all rooms with 100% efficiency electric heaters installed. It will be used in order to demonstrate the potential for energy savings in a retrofitted stock, under the assumption that an increase in building fabric energy efficiency will result in a reduction in energy use. However, we recognise this may not be the case in reality, as it has been reported that an increase in energy-efficient retrofit levels may not always result in a decrease in space heating energy consumption (BRE 2015), as occupant heating behaviours may change in more energy-efficient housing (Hamilton et al. 2011).

2.3. Metamodel Development

There are a number of different metamodeling methods that can be used to approximate the compiled output metrics described in section 2.2. Feed-forward Neural Networks and radial basis function (RBF) machine learning algorithms can both be used. These methods are examples of supervised machine learning, and are able to reproduce non-linear and non-monotonic relations between model input and output variables.

2.3.1. Neural networks (NNs)

Neural networks comprise of a set of neurons that connect input and output model variants (McCulloch and Pitts 1988). The inputs, outputs and hidden neurons are connected to one another by synapses that carry weights and biases, which determine the strength of the connections. Each neuron is associated with a transfer or activation function. Linear transfer functions are used at the input and output layers, whilst connections to hidden neurons are characterised by a sigmoid transfer function. The value given by an output neuron, y , is given by:

$$y = \frac{1}{1 + e^{-\eta}} \quad (2)$$

where $\eta = \sum_{i=1..n} w_i x_i + \beta$, with the inputs to a hidden neuron denoted by $\{x_1, x_2, \dots, x_n\}$. The regression weights and the bias value are represented by w_i and β , respectively. The weights and the bias values are updated in the fitting of the output metrics over a pre-defined number of training epochs.

The Python-Based Reinforcement Learning, Artificial Intelligence and Neural Network (PyBrain) Library (Schaul 2015) is used to implement the NN algorithm. This software is flexible and allows the user to construct the network architecture in terms of the number of layers and neurons, the connections and also the transfer functions.

Networks were set up with various architectures having either one or two hidden layers with each layer containing 4-20 neurons. When two hidden layers were used, the same number of neurons are present in both layers. A backwards propagation training algorithm is performed where the errors on the input variables are minimised with respect to the weights and the biases (Rumelhart, Hinton, and Williams 1988). Training for each NN is performed until the training error has converged. This is typically after about 100 training epochs.

2.3.2. Radial basis function (RBF)

This method uses a kernel for parameter estimation and is a type of support vector machine (SVM) (Broomhead and Lowe 1988). RBFs are real valued functions whose values depend on the radial distance, $r = \sqrt{\sum_{i=1..n} (x_0 - x_i)^2}$, relative to a point of origin, x_0 . Each output metric value is then given by:

$$y = \sum_{i=1..n} v_i \phi(r) \tag{3}$$

A Gaussian function is commonly used to represent $\phi(r)$, e.g. $\phi(r) = e^{-\epsilon r^2}$. Training is performed in order to determine the parameters v_i , ϵ and x_0 . A least squares method is used to do this such that metamodel output values are close to that of the true (original model) output values. The scikit-learn, SVM Python module (Pedregosa et al. 2011) was used to implement the RBF algorithm.

2.3.3. Metamodel validation

Testing of the metamodeling techniques involves passing an independent set of input values through each trained metamodel. The outputs produced by a metamodel are then compared to the outputs produced by the true EnergyPlus model.

Several goodness of fit metrics used in previous studies (Van Gelder et al. 2014) allow the accuracy of the different algorithms to be compared. These metrics include the Root Mean Square Error (*RMSE*), the coefficient of determination (R^2), and the Maximal Absolute Error (*MAE*) and are defined as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{j=1..m} (\hat{y}_j - y_j)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{j=1..m} (\hat{y}_j - y_j)^2}{\sum_{j=1..m} (\bar{y}_j - y_j)^2} \tag{5}$$

$$MAE = \max(|\hat{y}_1 - y_1|, \dots, |\hat{y}_m - y_m|) \tag{6}$$

where y_j represents the original model outputs (EnergyPlus), \bar{y} the mean original model output value, \hat{y}_j the metamodel output, and m the number of testing samples. These performance metrics are calculated for each output variable. Since metamodels are fitted using several input and output variables, the values for all of the continuous variables shown in Tables 2 and 3 are normalised to have a mean of zero, and a standard deviation of one prior to training. This ensures that the algorithms are not biased towards a particular variable and allows the performance of individual

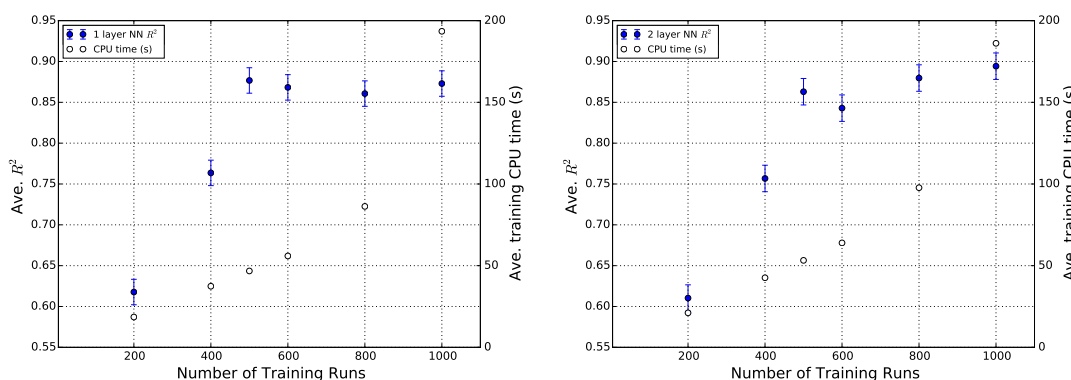


Figure 2. Number of training simulation runs used vs. the performance (R^2 averaged over output metrics and NN architectures with between 12 and 20 neurons) and the training CPU time (seconds) of the neural networks. The error bar represents the standard deviation of the nine networks. The study was performed for networks with one layer (left) and two layers (right) for a semi-detached dwelling with pensioner occupants.

output metrics to be compared. The sum of each normalised error gives the overall performance for a particular metamodel for both the $RMSE$ and MAE , whilst an average is calculated for the R^2 .

3. Results

We present results detailing the performance of the metamodels for several built form and occupancy types. The number of training samples required to give a reasonable metamodel performance is first needed to be determined. This was implemented by calculating the performance of the metamodel using varying numbers of training samples in the range 200 to 1000 EnergyPlus simulations. An independent sample of 100 simulations was used to test the performance of each trained NN. In this study, a semi-detached house was used (the most common built form in the EHS) with two pensioner occupants (type-1 as described in section 2.1.1). The NN performance (R^2) averaged over all output metrics and NN architectures is shown as a function of the number of training samples used in Figure 2. The average CPU time taken to train the networks is also displayed.

The results indicate that the performance of the metamodel improves as the number of training runs increases from 200 to 500. A significant improvement in performance is not observed when beyond 500 simulations are used. Due to computational limits on time and storage space (each EnergyPlus output file is about 2.5MB), we have therefore elected to use 500 EnergyPlus simulations to train each metamodel. There is not a significant difference in the overall performance between NNs with one or two layers.

The range of neural network architectures to use has also been investigated. This was again established using a semi-detached dwelling with pensioner occupants. Figure 3 graphically represents the performance of the NNs as a function of the number of neurons for one and two layered networks. The performance improves with the number of neurons in the range 5 to 12 and then levels off. To save CPU time due to training, we therefore only train networks for one or two layers with the number of neurons in the range of 12 to 20. The best overall performing network will then be selected. Tables 4-6 display the performance of NNs with various architectures for a semi-detached dwelling with pensioners using 500 training samples. The performance of the RBF algorithm is also shown for comparison. Results are shown for individual output variables as well as the overall performance of each metamodel.

In support of previous findings (Van Gelder et al. 2014), the overall performance of the NNs appear to be better than for the RBF method. The RBF does however have a superior performance for the heating energy use variable. The relationships between input parameters such as the

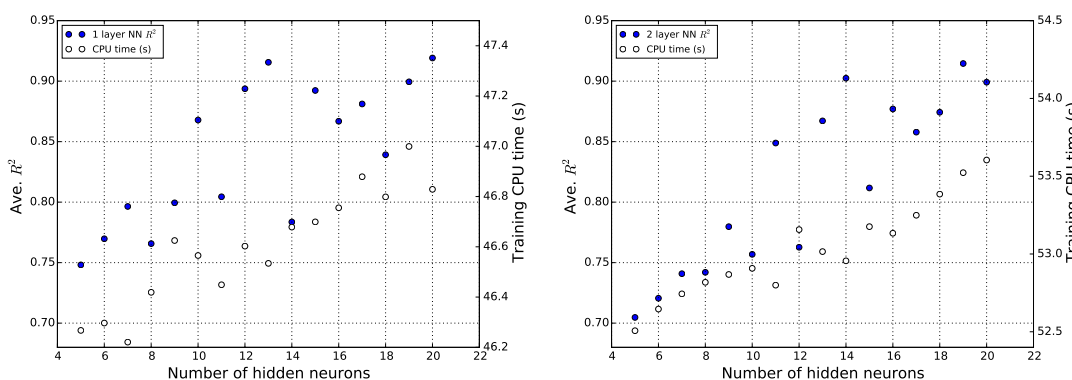


Figure 3. Number of hidden NN neurons vs. the performance of the network (R^2 averaged over all output metrics) and the training CPU time (seconds). The study has been performed for NNs with one layer (left) and two layers (right) for a semi-detached dwelling with pensioner occupants using 500 training simulations. The training CPU time axis have different scale ranges due to the training CPU time taking slightly longer for NNs with two layers.

Table 4. *RMSE* values for each output variable using neural networks (NNs) with various architectures (layers, neurons) and a radial basis function (RBF). Results are shown for a semi-detached dwelling with pensioner occupants.

Model	I/O	E_{heat}	$T_{OH}^{\text{bed-day}}$	$T_{OH}^{\text{bed-night}}$	$T_{OH}^{\text{liv-day}}$	$T_{OH}^{\text{liv-night}}$	RH	Tot.
NN(1,12)	0.29	0.22	0.21	0.35	0.34	0.35	0.45	2.22
NN(1,13)	0.36	0.22	0.15	0.28	0.25	0.25	0.43	1.94
NN(1,14)	0.33	0.21	0.24	0.34	0.43	0.87	0.48	2.92
NN(1,15)	0.34	0.20	0.17	0.31	0.29	0.37	0.50	2.19
NN(1,16)	0.30	0.23	0.17	0.30	0.33	0.61	0.43	2.37
NN(1,17)	0.34	0.22	0.18	0.29	0.31	0.48	0.46	2.30
NN(1,18)	0.34	0.21	0.23	0.36	0.43	0.64	0.43	2.64
NN(1,19)	0.33	0.22	0.18	0.30	0.29	0.34	0.48	2.13
NN(1,20)	0.33	0.20	0.15	0.28	0.23	0.27	0.44	1.90
<hr/>								
NN(2,12)	0.34	0.28	0.28	0.36	0.47	0.86	0.54	3.13
NN(2,13)	0.36	0.29	0.32	0.43	0.37	0.34	0.42	2.53
NN(2,14)	0.31	0.24	0.23	0.34	0.32	0.29	0.42	2.15
NN(2,15)	0.44	0.29	0.28	0.36	0.41	0.61	0.55	2.93
NN(2,16)	0.34	0.31	0.24	0.36	0.34	0.43	0.41	2.42
NN(2,17)	0.40	0.25	0.19	0.32	0.34	0.53	0.50	2.52
NN(2,18)	0.38	0.23	0.25	0.38	0.35	0.36	0.47	2.42
NN(2,19)	0.31	0.22	0.18	0.30	0.29	0.32	0.38	2.00
NN(2,20)	0.36	0.28	0.21	0.35	0.26	0.29	0.42	2.17
<hr/>								
RBF	0.29	0.15	0.69	0.70	0.86	0.91	0.46	4.06

thermostat setting and E_{heat} are linear in nature. This suggests that RBFs perform better when modelling simpler types of relationship. The RBF, however, struggles to model more complicated performance metrics such as day-time overheating which depends sinusoidally on the orientation of the building due to the location of the windows. Figure 4 shows the output values simulated using EnergyPlus plotted against the values output from the best performing NN (with one layer and 20 neurons) for a semi-detached house with pensioner occupants. It is clear that the metamodel performs less well at the extremes, for example, for buildings with high relative humidity. This is due to the lack of training simulation data for these buildings.

The performance of the metamodel has also been tested for all other built forms with the two pensioner occupancy types. This is to ensure that our methods are robust and can be applied in various scenarios. The performances of the selected metamodels for all built forms are shown for pensioners with and without shutters in Tables 7 and 8, respectively. A good performance is achieved for all built form and occupancy variants studied. The results suggest that the performance is slightly better for pensioners without shutters than those with shutters. This could be a result of the solar gains due to the building orientation variable having more of an influence on metamodel training.

Table 5. R^2 values for each output variable using neural networks (NNs) with various architectures (layers, neurons) and a radial basis function (RBF). Results are shown for a semi-detached dwelling with pensioner occupants.

Model	I/O	E_{heat}	$T_{OH}^{\text{bed-day}}$	$T_{OH}^{\text{bed-night}}$	$T_{OH}^{\text{liv-day}}$	$T_{OH}^{\text{liv-night}}$	RH	Ave.
NN(1,12)	0.91	0.95	0.95	0.88	0.88	0.88	0.79	0.89
NN(1,13)	0.87	0.95	0.98	0.92	0.94	0.94	0.81	0.92
NN(1,14)	0.89	0.95	0.94	0.89	0.81	0.24	0.77	0.78
NN(1,15)	0.88	0.96	0.97	0.91	0.92	0.86	0.75	0.89
NN(1,16)	0.91	0.95	0.97	0.91	0.89	0.62	0.81	0.87
NN(1,17)	0.88	0.95	0.97	0.91	0.91	0.77	0.79	0.88
NN(1,18)	0.88	0.96	0.95	0.87	0.81	0.59	0.81	0.84
NN(1,19)	0.89	0.95	0.97	0.91	0.92	0.89	0.77	0.90
NN(1,20)	0.89	0.96	0.98	0.92	0.95	0.93	0.81	0.92
<hr/>								
NN(2,12)	0.89	0.92	0.92	0.87	0.78	0.26	0.71	0.76
NN(2,13)	0.87	0.91	0.90	0.81	0.87	0.89	0.82	0.87
NN(2,14)	0.91	0.94	0.95	0.89	0.90	0.92	0.82	0.90
NN(2,15)	0.81	0.92	0.92	0.87	0.83	0.63	0.70	0.81
NN(2,16)	0.88	0.91	0.94	0.87	0.89	0.81	0.83	0.88
NN(2,17)	0.84	0.94	0.96	0.90	0.88	0.72	0.75	0.86
NN(2,18)	0.86	0.95	0.94	0.85	0.88	0.87	0.78	0.87
NN(2,19)	0.90	0.95	0.97	0.91	0.91	0.90	0.86	0.91
NN(2,20)	0.87	0.92	0.96	0.88	0.93	0.91	0.82	0.90
<hr/>								
RBF	0.89	0.97	0.56	0.25	-1.44	-5.88	0.63	-4.02

Table 6. MAE values for each output variable using neural networks (NNs) with various architectures (layers, neurons) and a radial basis function (RBF). Results are shown for a semi-detached dwelling with pensioner occupants.

Model	I/O	E_{heat}	$T_{OH}^{\text{bed-day}}$	$T_{OH}^{\text{bed-night}}$	$T_{OH}^{\text{liv-day}}$	$T_{OH}^{\text{liv-night}}$	RH	Tot.
NN(1,12)	1.30	0.70	0.81	1.15	0.95	0.99	3.00	8.90
NN(1,13)	1.62	0.73	0.42	0.88	0.84	0.69	3.22	8.40
NN(1,14)	1.87	0.65	0.89	1.28	1.36	2.37	2.81	11.22
NN(1,15)	2.24	0.56	0.42	0.99	1.07	1.22	4.18	10.68
NN(1,16)	1.59	0.65	0.44	1.05	1.13	1.73	3.15	9.75
NN(1,17)	1.69	0.74	0.45	0.88	0.87	1.40	3.01	9.04
NN(1,18)	1.77	0.53	0.62	1.19	1.18	1.57	3.05	9.92
NN(1,19)	1.60	0.63	0.56	1.02	0.99	1.01	3.13	8.94
NN(1,20)	1.94	0.49	0.38	0.99	0.96	0.71	3.36	8.83
<hr/>								
NN(2,12)	1.09	1.11	0.92	1.47	1.30	2.46	2.55	10.91
NN(2,13)	1.44	1.08	0.94	1.37	1.01	1.05	2.42	9.30
NN(2,14)	1.39	1.10	0.68	0.93	0.91	0.97	2.73	8.70
NN(2,15)	2.14	0.97	0.95	1.12	1.08	1.87	3.78	11.91
NN(2,16)	1.93	0.87	0.71	1.29	1.05	1.68	1.93	9.47
NN(2,17)	1.86	0.71	0.52	1.35	0.99	1.70	2.96	10.09
NN(2,18)	1.75	0.78	1.02	1.93	0.95	1.15	2.75	10.32
NN(2,19)	1.29	1.00	0.37	0.94	0.94	1.21	1.49	7.25
NN(2,20)	1.78	0.89	0.43	1.08	0.64	0.74	1.89	7.45
<hr/>								
RBF	1.45	0.38	2.44	2.11	2.06	2.33	3.62	14.40

Table 7. R^2 values for each output variable for the best performing metamodel for the various built forms with cavity walls and pensioner occupants with shutters closed during the summer months. Results for the flats are shown for the middle floor.

Built form	Model	I/O	E_{heat}	$T_{OH}^{\text{bed-day}}$	$T_{OH}^{\text{bed-night}}$	$T_{OH}^{\text{liv-day}}$	$T_{OH}^{\text{liv-night}}$	RH	Ave.
End terrace	NN(2,17)	0.89	0.97	0.90	0.81	0.88	0.79	0.83	0.87
Mid terrace	NN(1,18)	0.93	0.94	0.97	0.96	0.83	0.82	0.80	0.89
Semi-detached	NN(1,13)	0.88	0.97	0.94	0.92	0.84	0.83	0.81	0.89
Detached	NN(1,16)	0.91	0.94	0.96	0.94	0.86	0.93	0.83	0.91
Bungalow	NN(1,17)	0.75	0.98	0.90	0.93	0.94	0.94	0.63	0.87
Converted flat	NN(2,14)	0.95	0.92	0.87	0.87	0.94	0.90	0.92	0.91
Low-rise flat	NN(1,12)	0.90	0.91	0.72	0.78	0.72	0.80	0.82	0.81
High-rise flat	NN(1,12)	0.91	0.94	0.83	0.94	0.93	0.94	0.72	0.89

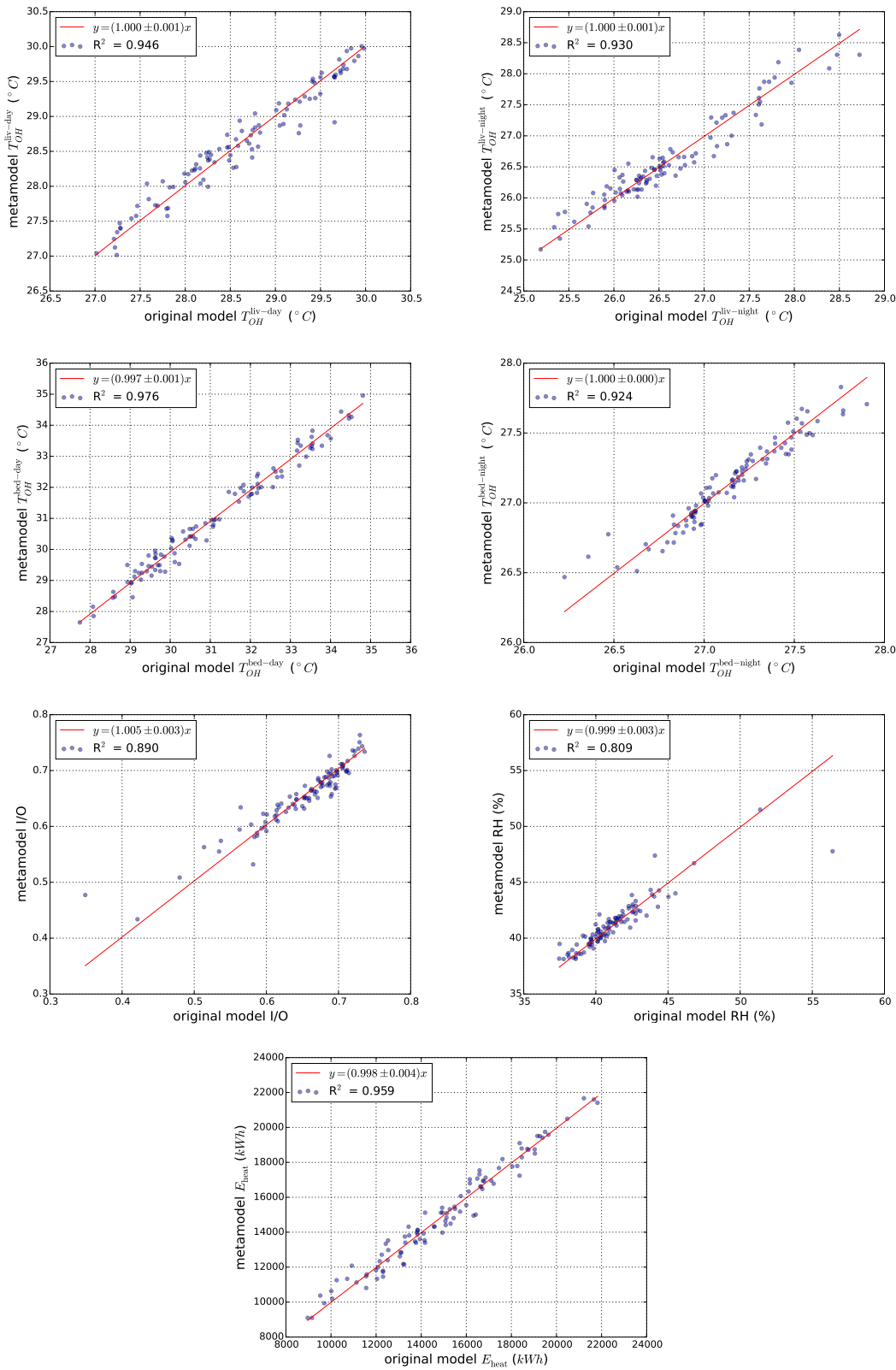


Figure 4. Original EnergyPlus simulation values plotted against the outputs of the best performing metamodel (NN with one layer containing 20 neurons) using 100 testing samples for a semi-detached house with pensioner occupants. Distributions are shown for $T_{OH}^{liv-day}$ (top left), $T_{OH}^{liv-night}$ (top right), $T_{OH}^{bed-day}$ (upper middle left), $T_{OH}^{bed-night}$ (upper middle right), I/O (lower middle left), RH (lower middle right) and E_{heat} (bottom). The linear fits on the graphs indicate that the metamodels do an excellent job of replicating the original model.

Table 8. R^2 values for each output variable for the best performing metamodel for the various built forms with cavity walls and pensioner occupants. Results for the flats are shown for the middle floor.

Built form	Model	I/O	E_{heat}	$T_{OH}^{\text{bed-day}}$	$T_{OH}^{\text{bed-night}}$	$T_{OH}^{\text{liv-day}}$	$T_{OH}^{\text{liv-night}}$	RH	Ave.
End terrace	NN(2,17)	0.92	0.95	0.96	0.93	0.93	0.95	0.90	0.93
Mid terrace	NN(2,18)	0.91	0.96	0.95	0.92	0.91	0.88	0.85	0.91
Semi-detached	NN(1,20)	0.89	0.96	0.98	0.92	0.95	0.93	0.81	0.92
Detached	NN(1,20)	0.89	0.98	0.97	0.96	0.94	0.90	0.92	0.94
Bungalow	NN(2,15)	0.94	0.97	0.92	0.90	0.91	0.91	0.91	0.92
Converted flat	NN(2,18)	0.90	0.89	0.95	0.94	0.91	0.85	0.79	0.89
Low-rise flat	NN(2,13)	0.94	0.90	0.87	0.89	0.90	0.88	0.91	0.90
High-rise flat	NN(2,15)	0.90	0.95	0.84	0.92	0.96	0.77	0.87	0.89

4. Discussion, Limitations and Future Work

The model described is able to rapidly estimate the indoor overheating risk, air pollution levels, and space heating energy use in a number of dwellings under current and future climate scenarios. Due to the large number and range of categorical and continuous input variables, the model is built entirely using parallel processing on HPC facilities. The model is able to account for changes in the stock due to new builds and the retrofit of existing buildings. The metamodeling framework also allows for optimisation and sensitivity analyses to be performed. Metamodel performance results for the various built forms shown in Table 8 indicate that a high level of accuracy is achieved. There are, however, several limitations to the model. Here, we acknowledge these and present some potential solutions:

- (1) While we enable the energy-efficient retrofit of building fabrics to passivhaus standards, we do not model the addition of any supplementary ventilation systems such as Mechanical Heat and Ventilation Recovery (MVHR). These systems may reduce overheating risks by increasing ventilation above passive levels, while also reducing pollutant infiltration through the use of filters. Further work will consider such an implementation.
- (2) While four different occupancy scenarios are modelled, these do not capture the full range of occupant behaviours and their potential influences on indoor temperatures and air pollution. A handle on window opening and heating occupancy behaviour can be achieved by using the Energy Follow Up Survey (EFUS) (BRE 2013) data set, which has indoor temperature measurements for dwellings surveyed as part of the 2010-2011 EHS. This data can be indirectly used to estimate the average and range of window opening and heating behaviour for England, whilst validating the modelling of indoor temperatures;
- (3) Uncertainties in the building fabric may be assessed in the metamodel due to the random sampling of model inputs, but certain parameters such as building geometry and layout are held static, and their uncertainty unquantifiable. Building geometry and layout is likely to have a significant role in overheating and outdoor pollution infiltration risks. Accounting for this additional level of complexity has not yet been implemented;
- (4) We assume a constant outdoor air pollution level, and do not account for spatial or temporal variations in pollution levels. Calculating I/O retains the option to estimate absolute indoor concentrations of outdoor pollution should high-resolution outdoor pollution levels be made available.
- (5) Our calculations of energy use assume that there is no “take-back”, and that changes to the building envelope result in a corresponding change to energy consumption. This is not necessarily the case, as homeowners often prefer to heat their homes for longer and/or at higher temperatures (Hamilton et al. 2011).

In future work, the metamodel will be used to calculate metrics for each entry within the EHS based on the specified building type and fabric building characteristics within the database. U-values of buildings within the EHS are inferred from fabric types, insulation levels, and ages based on SAP. The Homes Energy Efficiency Database (HEED), a database of dwellings held by the Energy Saving Trust (EST) that has details of energy-efficiency retrofits. This has physical dwelling

information as well as postcode information. Postcode information in HEED allows for the results to be mapped and adjusted by local weather data, where available. Furthermore, Office for National Statistics (ONS 2011) data on small-area urban/rural classifications for Large Super Output Areas may be used to estimate the wind and sun exposure of the individual buildings within the building stock. Examining the relationship between indoor overheating and outdoor temperatures enables the model to account for highly localised temperatures, such as UHI effects.

The rapid estimation of overheating and indoor air pollution exposure risks and energy use for an entire housing stock, enables a number of different potential scenarios to be easily tested. Modelling housing stocks with geographical references, allows results to be easily visualised across a spatial area, as has been done for static housing stocks under current conditions for outdoor air pollution and overheating (Taylor et al. 2015b). As a result, the model provides a useful tool for policy makers, urban planners, and public health experts when estimating the influences of the housing stock on population exposure to overheating and outdoor air pollution inside dwellings across England.

5. Summary

A national overheating and indoor air pollution metamodelling framework is presented, developed with the use of HPC. A statistical sampling technique and metamodelling approach allows the tool to rapidly estimate a number of dwelling indoor overheating, pollution indoor/outdoor (I/O) ratios, relative humidity, and energy use based on the characteristics of buildings. The results indicate that using a neural network with between 12 and 20 hidden neurons with one or two hidden layers produces metamodels with the best performance. We elected to use 500 training simulations due to constraints on the storage of EnergyPlus output files and the computer time taken to run these simulations. This metamodelling approach is able to build on existing national models of overheating and indoor air pollution risk by being adaptable to changes in the building stock in the coming years. Furthermore, adaptations to reduce overheating risks, such as shading using external shutters, fabric and ventilation changes, and local environmental changes to the UHI, can be accounted for. This model acts as a powerful tool for assessing changes to indoor risks caused by policy, climate change, and a transforming housing stock.

Acknowledgments

This research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) on the topic of Environmental Change and Health. The project is lead by the London School of Hygiene and Tropical Medicine in partnership with Public Health England (PHE), and in collaboration with the University of Exeter, University College London, and the Met Office. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

References

- Armstrong, B. G., Z. Chalabi, B. Fenn, S. Hajat, S. Kovats, and A. Milojevic. 2011. "Association of mortality with high temperatures in a temperate climate: England and Wales." *Journal of Epidemiology and Community Health* (65): 340–5.
- Barton, R. R., and M. Meckesheimer. 2006. "Chapter 18 Metamodel-Based Simulation Optimization." In *Simulation*, Vol. 13 of *Handbooks in Operations Research and Man-*

- agement Science edited by S. G. Henderson and B. L. Nelson. 535 – 574. Elsevier. <http://www.sciencedirect.com/science/article/pii/S0927050706130182>.
- Baudin, M. et al. 2015. “pyDOE.” <http://pythonhosted.org/pyDOE/>.
- Beizaee, A., K. J. Lomas, and S. K. Firth. 2013. “National survey of summertime temperatures and overheating risk in English homes.” *Building and Environment* (65): 1–17.
- BRE. 2009. “The Governments Standard Assessment Procedure for Energy Rating of Dwellings.” *Building Research Establishment* .
- BRE. 2013. “Energy Follow-Up Survey 2011: Report 9: Domestic appliances, cooking & cooling equipment.” *EFUS* .
- BRE. 2015. “Energy Use in Homes: A series of reports on domestic energy use in England.” *Building Research Establishment* .
- Broomhead, D. S., and D. Lowe. 1988. “Multivariable Functional Interpolation and Adaptive Networks.” *Complex Systems* 2 321–355.
- Carlucci, S., and L. Pagliano. 2012. “A review of indices for the long-term evaluation of the general thermal comfort conditions in buildings.” *Energy and Buildings* 53: 194–205.
- CIBSE. 2003. “British Standards: Heating systems in buildings. Method for calculation of the design heat load.” *BSi* (12831).
- Das, P., C. Shrubsole, B. Jones, I. Hamilton, Z. Chalabi, and M. Davies. 2014. “Using probabilistic sampling-based sensitivity analyses for indoor air quality modelling.” *Building and Environment* (78): 171–82.
- DCLG. 2011. “English Housing Survey.” *Department for Communities and Local Government* (76).
- De Wilde, P., and W. Tian. 2010. “The role of adaptive thermal comfort in the prediction of the thermal performance of a modern mixed-mode office building in the UK under climate change.” *Journal of Building Performance Simulation* (3): 87–101.
- Eames, M., T. Kershaw, and D. Coley. 2010. “On the creation of future probabilistic design weather years from UKCP09.” *Building Services Engineering Research and Technology* (32): 127–42.
- Eisenhower, B., Z. O'Neill, S. Narayanan, V. A. Fonoberov, and I. Mezi. 2012. “A methodology for meta-model based optimization in building energy models.” *Energy and Buildings* (47): 292–301.
- Fisk, W. J., Q. Lei-Gomez, and M. J. Mendell. 2007. “Meta-analyses of the associations of respiratory health effects with dampness and mold in homes.” *Indoor Air* 17 (4): 284–296. <http://dx.doi.org/10.1111/j.1600-0668.2007.00475.x>.
- Gowers, A., B. Miller, and J. Stedman. 2014. “Estimating Local Mortality Burdens associated with Particulate Air Pollution.” *Public Health England. Centre for Radiation, Chemical and Environmental Hazards, UK* <https://www.gov.uk/government/publications/estimating-local-mortality-burdens-associated-with-particulate-air-pollution>.
- Gupta, R., and M. Gregg. 2013. “Preventing the overheating of English suburban homes in a warming climate.” *Building Research & Information* (41): 281–300.
- Hajat, S., S. Vardoulakis, C. Heaviside, and B. Eggen. 2014. “Climate change effects on human health: projections of temperature-related mortality for the UK during the 2020s, 2050s and 2080s.” *Journal of Epidemiology and Community Health* <http://jech.bmj.com/content/early/2014/01/08/jech-2013-202449.abstract>.
- Hamdy, M., and J. Hensen. (2015). “Ranking of Dwelling types in terms of Overheating Risk and Sensitivity to Climate Change.” *International Conference of the International Building Performance Simulation Association (BS2015), 7-9 December 2015, Hyderabad, India* .
- Hamilton, I. G., M. Davies, I. Ridley, T. Oreszczyn, M. Barrett, R. Lowe, S. Hong, P. Wilkinson, and Z. Chalabi. 2011. “The impact of housing energy efficiency improvements on reduced exposure to cold—the “temperature take back factor”.” *Building Services Engineering Research and Technology* 32 (1): 85–98.
- HM-Government. 2010. “Approved Document F.” *Government Report* .
- Hoek, G., G. Kos, R. Harrison, J. de Hartog, K. Meliefste, and H. ten Brink. 2008. “Indoor–outdoor relationships of particle number and mass in four European cities.” *Atmospheric Environment* (42): 156–69.
- Johnson, H., R. S. Kovats, G. McGregor, J. Stedman, M. Gibbs, and H. Walton. 2005. “The impact of the 2003 heat wave on daily mortality in England and Wales and the use of rapid weekly mortality estimates.” *Euro Surveill.* 10 (7): 168–171.
- Jones, N., C. Thornton, D. Mark, and R. Harrison. 2000. “Indoor/outdoor relationships of particulate matter in domestic homes with roadside, urban and rural locations.” *Atmospheric Environment* (34): 2603–12.
- Kornartit, C., R. S. Sokhi, M. A. Burton, and K. Ravindra. 2010. “Activity pattern and personal exposure to nitrogen dioxide in indoor and outdoor microenvironments.” *Environ. Int* (36): 36–45.

- Lai, H. K., L. Bayer-Oglesby, R. Colville, T. Gtschi, M. J. Jantunen, and N. Knzli. 2006. “Determinants of indoor air concentrations of PM_{2.5}, black smoke and NO₂ in six European cities (EXPOLIS study).” *Atmospheric Environment* (40): 1299–313.
- Lomas, K. J., and T. Kane. 2012. “Summertime temperatures in 282 UK homes: thermal comfort and overheating risk.” In *Proceedings of 7th Windsor Conference: The changing context of comfort in an unpredictable world*, .
- Long, C. M., H. H. Suh, P.J. Catalano, and P. Koutrakis. 2001. “Using Time- and Size-Resolved Particulate Data To Quantify Indoor Penetration and Deposition Behavior.” *Environmental Science & Technology* (35): 2089–99.
- Mavrogianni, A., M. Davies, J. Taylor, Z. Chalabi, P. Biddulph, and E. Oikonomou. 2013. “The impact of occupancy patterns, occupant-controlled ventilation and shading on indoor overheating risk in domestic environments.” *Building and Environment* (78): 183198.
- Mavrogianni, A., M. Davies, P. Wilkinson, and A. Pathan. 2010. “London housing and climate change: impact on comfort and health-preliminary results of a summer overheating study.” *Open House International* 35 (2): 49–59.
- Mavrogianni, A., P. Wilkinson, M. Davies, P. Biddulph, and E. Oikonomou. 2012. “Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings.” *Building and Environment* (55): 117–30.
- McCulloch, W. S., and W. Pitts. 1988. “Neurocomputing: Foundations of Research.” chap. A Logical Calculus of the Ideas Immanent in Nervous Activity, 15–27. Cambridge, MA, USA: MIT Press. <http://dl.acm.org/citation.cfm?id=65669.104377>.
- Murphy, J. M., D. Sexton, G. Jenkins, P. Boorman, B. Booth, K. Brown, R. Clark, M. Collins, G. Harris, and E. Kendon. 2009. “UKCP09 Climate change projections.” *Met Office Hadley Centre, Exeter* .
- Nasir, Z. A., and I. Colbeck. 2013. “Particulate pollution in different housing types in a UK suburban location.” *The Science of the Total Environment* (445): 165–76.
- New, J., J. Sanyal, M. Bhandari, and S. Shrestha. 2012. “Autotune E+ Building Energy Models.” *5th Sim-Build Conference of IBPSA-USA, International Building Performance Simulation Association (IBPSA), August 1-3* .
- Oikonomou, E., M. Davies, A. Mavrogianni, P. Biddulph, P. Wilkinson, and M. Kolokotroni. 2012. “Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London.” *Building and Environment* (57): 223–38.
- Oikonomou, E., A. Mavrogianni, R. Raslan, J. Taylor, T. Oreszczyn, and M. Davies. 2016. “English Archetypes: Developing a domestic model for building performance calculations.” *under preparation* .
- ONS. 2005. “United Kingdom Time Use Survey.” *Office for National Statistics* .
- ONS. 2011. “2011 rural/urban classification for small-area geographies.” *Office for National Statistics* .
- PassivHaus. 2015. “Passive House in Darmstadt [online].” <http://www.passivehouse.com>.
- Peacock, A. D., D. P. Jenkins, and D. Kane. 2010. “Investigating the potential of overheating in UK dwellings as a consequence of extant climate change.” *Energy Policy* (38): 3277–88.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. “Scikit-learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12: 2825–2830.
- Porritt, S., and P. Cropper. 2010. “Heat wave adaptations for UK dwellings and introducing a retrofit toolkit.” *International Journal of Disaster Resilience in the Built Environment* (4:3): 269–286.
- Porritt, S., and L. Shao. 2010. “Occupancy patterns and their effect on interventions to reduce overheating in dwellings during heat waves.” *Proceedings of Conference: Adapting to Change: New Thinking on Comfort* .
- Porritt, S., L. Shao, P. Cropper, and C. Goodier. 2010. “Ranking of interventions to reduce overheating in dwellings during heat waves.” *Energy and Buildings* (55): 16–27.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1988. “Neurocomputing: Foundations of Research.” chap. Learning Representations by Back-propagating Errors, 696–699. Cambridge, MA, USA: MIT Press. <http://dl.acm.org/citation.cfm?id=65669.104451>.
- Sanyal, J., and J. New. 2013. “Simulation and big data challenges in tuning building energy models.” In *Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), 2013 Workshop on*, 1–6. May.
- Schaul, T. et al. 2015. “pyBrain.” <http://pybrain.org/>.
- Shrubsole, C., I. Ridley, P. Biddulph, J. Milner, S. Vardoulakis, and M. Ucci. 2012. “Indoor PM_{2.5} exposure in Londons domestic stock: Modelling current and future exposures following energy efficient refurbishment.” *Atmospheric Environment* (62): 33643.

- Shrubsole, C., J. Taylor, P. Das, I. Hamilton, and M. Davies. 2015. "Impacts of energy efficiency retrofitting measures on indoor PM2.5 concentrations across different income groups in England: a modelling study." *Advances in Building Energy Research* .
- Stephen, R. 2000. "Airtightness in UK Dwellings." *BRE* .
- Tang, B. 2012. "Orthogonal Array-Based Latin Hypercubes." *Journal of the American Statistical Association* .
- Taylor, J., M. Davies, A. Mavrogianni, Z. Chalabi, P. Biddulph, and E. Oikonomou. 2014a. "The relative importance of input weather data for indoor overheating risk assessment in dwellings." *Building and Environment* (76): 81–91.
- Taylor, J., M. Davies, A. Mavrogianni, C. Shrubsole, I. Hamilton, P. Das, B. Jones, E. Oikonomou, and P. Biddulph. 2016. "Mapping indoor overheating and air pollution risk modification across Great Britain: A modelling study." *Building and Environment* 99: 1–12. <http://www.sciencedirect.com/science/article/pii/S0360132316300105>.
- Taylor, J., A. Mavrogianni, M. Davies, P. Das, C. Shrubsole, and P. Biddulph. 2015a. "Understanding and mitigating overheating and indoor PM2.5 risks using coupled temperature and indoor air quality models." *Building Services Engineering Research and Technology* (0143624414566474).
- Taylor, J., C. Shrubsole, M. Davies, P. Biddulph, P. Das, and I. Hamilton. 2014b. "The modifying effect of the building envelope on population exposure to PM2.5." *Indoor Air* .
- Taylor, T., P. Wilkinson, M. Davies, B. Armstrong, Z. Chalabi, A. Mavrogianni, P. Symonds, E. Oikonomou, and S. I. Bohnenstengel. 2015b. "Mapping the effects of urban heat island, housing, and age on excess heat-related mortality in London." *Urban Climate* <http://www.sciencedirect.com/science/article/pii/S2212095515300146>.
- US-DoE. 2013. "EnergyPlus V8." <http://apps1.eere.energy.gov/buildings/energyplus/>.
- Van Gelder, L., P. Das, H. Janssen, and S. Roels. 2014. "Comparative study of metamodelling techniques in building energy simulation: Guidelines for practitioners." *Simulation Modelling Practice and Theory* (49): 245–57.
- Van Rossum, G., and F.L. Drake (eds). 2001. "Python Reference Manual." <http://www.python.org>.
- Vardoulakis, S., S. Dimitroulopoulou, J.E. Thornes, K-M. Lai, J. Taylor, I. Myers, C. Heaviside, et al. 2015. "Impact of climate change on the domestic indoor environment and associated health risks in the UK." *Environment International* 85: 299–313.
- Walton, H., D. Dajnak, S. Beevers, M. Williams, P. Watkiss, and A. Hunt. 2010. "Understanding the Health Impacts of Air Pollution in London." *Kings College London* .
- Wright, A. J., A. N. Young, and S. Natarajan. 2005. "Dwelling temperatures and comfort during the August 2003 heat wave." *Building Services Engineering Research and Technology* 26 (4): 285–300.
- ZCH. 2015a. "Defining Overheating: Evidence Review." *Zero Carbon Hub* .
- ZCH. 2015b. "Overheating in Homes - The Big Picture ." *Zero Carbon Hub* .