

Risk Analysis of Housing Energy Efficiency Interventions under Model Uncertainty

Zaid Chalabi^{1,2,*}

Payel Das^{2,}**

James Milner¹

Mike Davies²

Ian Hamilton⁴

Benjamin Jones⁵

Clive Shrubsole²

Paul Wilkinson¹

¹ Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK

² UCL Institute for Environmental Design and Engineering, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

⁴ UCL Energy Institute, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

⁵ Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham NG7 2RD, UK

***Author for correspondence: Dr Zaid Chalabi, Department of Social and Environmental Health Research, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK. Email: zaid.chalabi@lshtm.ac.uk ; Phone: +44(0)20 7927 2453.**

**** Present address: Rudolf Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK**

Abstract

Mathematical models can be used to evaluate the health impacts of housing energy efficiency interventions. However by their nature, models are subject to uncertainty and variability, which are important to quantify if used to support policy decisions. Models that are used to assess the impacts on health of housing energy efficiency interventions are likely to be based on a pair of linked component models: a building physics model which calculates changes in exposures and whose outputs then feed into a health impact model. Current methods to propagate uncertainty in a series of models, where the outputs of one model are inputs to another, invariably use Monte Carlo (MC) numerical simulation. In this paper, two methods are used to quantify the uncertainty in the impact of draught proofing on childhood asthma: the MC simulation method and a semi-analytical method based on integral transforms. Both methods give close results but it is argued that the semi-analytical method has some advantages over the MC method, particularly in quantifying the uncertainties in the main outputs of the building physics model before propagating them to the health model.

Highlights:

- Uncertainty in health risks associated with draught proofing are quantified
- Monte Carlo method and a semi-analytical method are used to quantify uncertainty
- Both methods give approximately the same result
- The semi-analytical method has some advantages over the MC method
- It quantifies uncertainty in mould exposure before propagation to health impact

Key words: Uncertainty propagation; risk analysis; modelling; energy efficiency.

1. Introduction

Quantitative health impact assessment of housing policies and interventions requires the use of building physics and health models [1-3]. By their nature, any mathematical model is subject to uncertainty which could be attributed either to the uncertainty in its parameters or its structure. When assessing the health risks associated with a policy or an intervention, it is important that uncertainties and variability in the models are taken into account to aid robust decision-making. Although uncertainty and variability are often treated the same mathematically, they are different conceptually; with new evidence uncertainty is likely to decrease whereas variability either remains the same or even increases. The aim of this paper is to quantify the uncertainty in health impacts associated with draught proofing. This is done using two methods: the classical Monte Carlo (MC) method and a semi-analytical method. The focus is on handling parametric uncertainty and variability; structural uncertainty will be addressed in a separate study. The terms uncertainty and variability are used interchangeably in this paper unless otherwise specified.

The semi-analytical method combines the MC method with an integral transform method. The integral transform method for handling uncertainty is based on the algebra of random variables [4-5] and has been applied previously in engineering [6-8]. It is not as popular as the MC method because of the difficulty in calculating the integral transforms analytically. However recent advances in mathematical software for symbolic processing would enable the calculation of the integrals analytically.

The framework of analysis of this study is shown in Figure 1. The first method treats the building physics model and the health model as one model and applies the MC method fully to the combined model. The parameters of the combined building physic-health model are drawn from their respective distributions and the simulations are repeated as in any MC simulation. This method is completely numerical. The second method however applies the MC method only to the building physics model and propagates the uncertainty between the

output of the building physics model (mould exposure) and the health model analytically using integral transforms. This method is called semi-analytical because it combines a numerical method with an analytical method. The analytical component of the method entails symbolic (or algebraic) processing to calculate integral transforms.

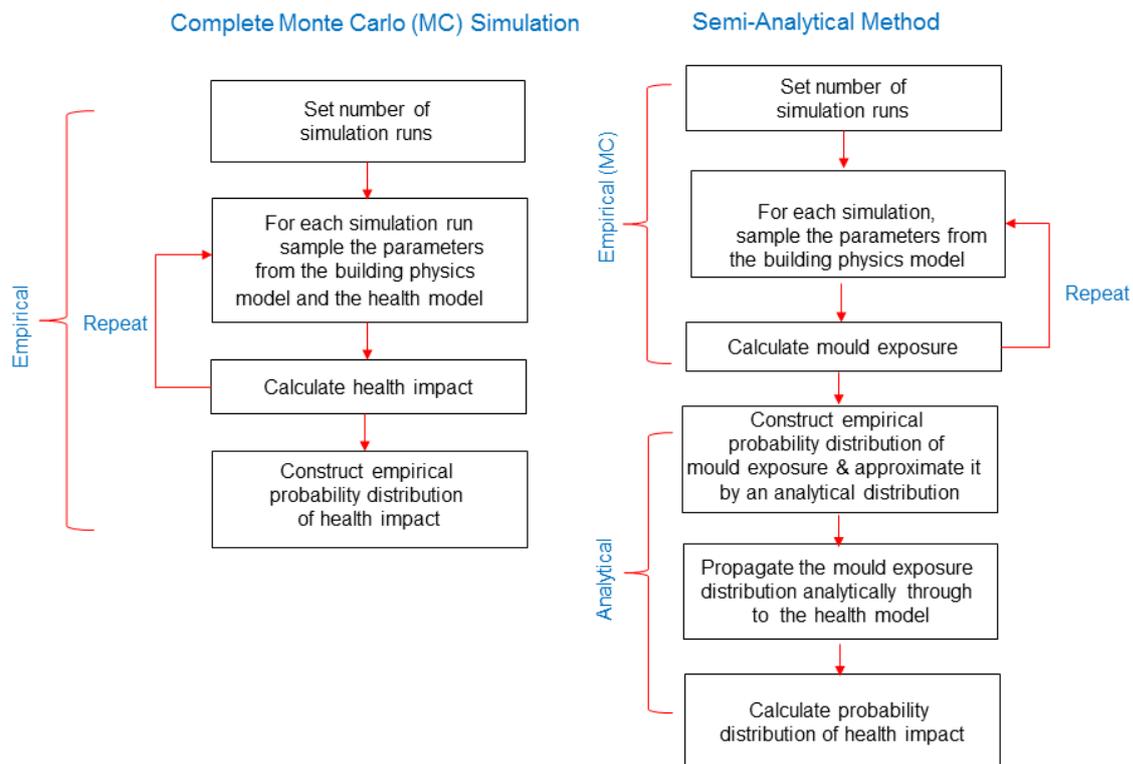


Figure 1. Framework of analysis. Two methods are used: a complete MC simulation and a semi-analytical method which combines a MC method for the building physics model and a semi-analytical method for propagating the uncertainty between the building physics model and the health model.

The outline of the paper is as follows. The second section describes the methods consisting of (i) the building physics model which simulates changes in the indoor environmental exposures post-intervention, (ii) the health model which maps changes in the indoor exposures to health outcomes, (iii) quantifying the uncertainty in the building physics and health models, and finally, (iv) propagating the uncertainty between the two models. The

third section gives the results of the uncertainty analysis and compares the results of the full probabilistic Monte Carlo (MC) method with the proposed semi-analytical method. The discussion section gives the main findings and debates the strengths and weaknesses of our uncertainty approach. The last section concludes. In order to make the paper self-contained, four appendices are added for the purpose of providing (A) the theoretical details of the building physics model, (B) practical details on the building physics modelling software tool used, (C) background material on the integral (Mellin) transform, and (D) definitions of mathematical functions referred to in the paper.

2. Methods

Without loss of generality, we illustrate the methods on a case-study dwelling which is a flat (apartment) with two exposed walls (Figure 2) and an extract fan in the kitchen and bathroom to meet current UK building regulations [9].



Figure 2. Layout of zones within case-study dwelling. The pale blue rectangles show where windows are placed, yellow circles show airflow paths enabling flow of air and moisture between zones, dark blue squares show zones where moisture is generated, and green squares show zones with extract fans installed.

The purpose of the analysis is to determine the uncertainty in the health risks (or benefits) associated with an energy efficiency housing intervention. Draught-proofing is chosen as an example of an intervention that is likely to have an impact on health. Draught proofing increases the air tightness of the dwelling which, among other effects, can influence both indoor temperature and, through changes in indoor temperature and reduced ventilation, mould risk. Air tightness is the resistance of a building's fabric to infiltration and exfiltration where infiltration is the uncontrolled ventilation into a building and exfiltration is the

uncontrolled ventilation out of a building [10]. Presence of mould in dwellings is known to be associated with respiratory symptoms in children [11-12]. The aim is to quantify the uncertainty in the health risks associated with this intervention.

Again, without loss of generality, only the mould-respiratory disease pathway is considered. Mould severity index (MSI) is commonly used to quantify the mould exposure in dwellings [13]. MSI is based on reported visible mould defined within the English Housing Survey [14]. If MSI is greater than unity, this indicates the presence of mould. Building physics models are used here to determine the change in the likelihood of MSI exceeding unity due to draught-proofing in the case-study dwelling. Health models are then used to associate the change in mould exposure with asthma in children. Health impacts are expressed in Quality Adjusted Life Years (QALYs). The QALY is a health metric which is widely used in health impact evaluations, combining survival and quality of life lived [15].

2.1 Building physics (exposure) model

Details of the building physics model are given in Appendix A. The likelihood of MSI exceeding unity ($\mathcal{L}_{MSI>1}$) in the living room is estimated from a combination of (i) a stand-alone indoor air quality simulation model and (ii) empirical relationships derived from the national evaluation study of the “Warm Front” home energy intervention scheme in England [16]. Warm Front was a large programme in England whose aim was to reduce fuel poverty by improving energy efficiency in dwellings through the introduction of a number of housing interventions such as cavity wall insulation, loft insulation and draught proofing.

The CONTAM indoor air quality simulation model [17] was used in this study to model the infiltration and exfiltration through adventitious openings, doors, and windows as well as

room-to-room airflows in the selected flat archetype. Information on the CONTAM setup for simulating the indoor environment of UK households are given elsewhere [18-20] and the main CONTAM input parameters of relevance to this study are given in Appendix B. The airflows are a result of the wind pressures acting on the building envelope, and buoyancy effects induced by differences between internal and external temperatures. The dwelling is assumed to be ventilated through extract fans in bathroom and kitchen and by natural means (i.e. through cracks in the building envelope and opening of windows and doors without the aid of mechanical air movement systems). Moisture is modelled in this study as a non-trace pollutant and is assumed to be produced by occupants, cooking, showers, as well as by ingress from the external environment through air exchange. From the perspective of this study, the key inputs to CONTAM include dwelling characteristics (total ground floor area, permeability of the building envelope, height and orientation), occupant behaviour (affecting indoor moisture production rate, cooking times, and operation of windows and doors) and weather information (wind speed, wind direction, external moisture level, external temperature).

The models are only run for eight months of winter, nominally defined between 1st October and 31st May, because these are the months when mould poses the biggest risk. A typical weather profile (CIBSE's London Test Reference Year Weather file¹) obtained from the Chartered Institution of Building Services Engineers [21] database is used in the simulations. Moisture is assumed to be produced in the kitchen during cooking, in the bathroom during use of shower and toilet, and in the bedrooms during sleeping times. It is also assumed that the kitchen and bathroom do not have windows but have extract fans which are switched on

¹ <http://www.cibse.org/Knowledge/CIBSE-other-publications/CIBSE-Weather-Data-Current,-Future,-Combined-DSYs>

during times of cooking and use of bathroom, and that windows in the remaining zones are closed during winter months.

2.2 Health model

The relative risk R of incidence of asthma in children associated with mould is given by²:

$$R = \exp(\Delta\mathcal{L}_{MSI>1} \times \log(r)) \quad (1)$$

where $\Delta\mathcal{L}_{MSI>1}$ is the change in the likelihood of MSI exceeding unity due to the housing intervention and r is the asthma risk coefficient per unit change in the likelihood of MSI exceeding unity. For small values of $\Delta\mathcal{L}_{MSI>1} \times \log(r)$, R can be approximated to a first order Taylor series expansion around $\Delta\mathcal{L}_{MSI>1} = 0$ by

$$R \cong 1 + \Delta\mathcal{L}_{MSI>1} \times \log(r) \quad (2)$$

Denote the baseline prevalence of asthma in children by p , the number of dwellings which is to receive draught proofing by n , the average number of children aged 14 and under per dwelling by c . The health impact associated with the change in the likelihood of MSI is then given by [22]:

$$h = (1 - R) \times ((1 - w) \times p \times n \times c) = -\Delta\mathcal{L}_{MSI>1} \times (1 - w) \times p \times n \times c \times \log(r) \quad (3)$$

where h is the health impact (health gain for a negative $\Delta\mathcal{L}_{MSI>1}$ and health burden for a positive $\Delta\mathcal{L}_{MSI>1}$) and w is the quality of life weight for asthma. Because the epidemiological evidence is often expressed in terms of the logarithm of the risk coefficient, $q = \log(r)$ is replaced in Equation (3). If $u = 1 - w$, Equation (3) becomes:

² To avoid possible confusion in a long mathematical expression, the symbol “ \times ” is sometimes used to denote multiplication

$$h = -\Delta\mathcal{L}_{MSI>1} \times q \times u \times p \times n \times c \quad (4)$$

which gives an expression of health impact in terms of the change in the likelihood of MSI exceeding unity.

2.3 Characterisation of uncertainty and variability

The uncertainty and variability in the parameters of the models will be characterised using probability density functions (pdfs). As noted in the Introduction, the main difference between uncertainty and variability is that uncertainty in a parameter decreases with new evidence and additional information, whereas variability in a parameter does not necessarily decrease and may even increase. Their mathematical characterisation however is the same.

For the changes in health-related exposures, generated from the building physics model, we used MC simulations to capture the uncertainty and variability in the building physics model. There are many sources of uncertainty and variability in the inputs to CONTAM. Naturally not all of them were considered in this analysis because a full treatment of uncertainty was not the focus of this work. For example, only one weather scenario is used to define the external weather conditions.

To make the case-study dwelling realistic, the variability in the dwelling characteristics (of the type shown in Figure 1) is derived from real dwellings in London using the English Housing Survey [14]. Data in the EHS includes information on the dwelling age, dwelling type, height, and ground floor area amongst many other characteristics of the dwelling. The EHS would be used as a database from which to sample randomly the characteristics of the dwellings which are similar to the case study. Variation in height, total ground floor area and dwelling age are sampled jointly by randomly selecting an EHS database entry that matches the case-study dwelling. The fraction of the total ground area contributed by each of the

zones is kept constant, and these fractions are given in Appendix B. Because there is no information on dwelling orientation in the EHS, orientation is assumed to vary uniformly between 0° and 360°. The wind direction is set by the selected weather file, which is not varied as part of the uncertainty analysis. The permeability of the dwelling is estimated from the age of the dwelling using SAP, a UK government-approved tool for calculating notional energy demand and efficiency characteristics [23].

In terms of parametric uncertainty, only three sources are considered. The first is the uncertainty in the maximum moisture production rate used in CONTAM. The moisture production rate in each room is expressed as a fraction of this maximum rate. The maximum rate is assumed to be uniformly distributed within $\pm 10\%$ of its baseline value. The second source of uncertainty is the estimate of the change in permeability and E-value associated with draught-proofing. These are based on the changes in infiltration rate measured as a result of draught stripping of windows, reducing floor infiltration, and any other infiltration adjustments in an air tightness investigation in the Warm Front study [24]. The change in infiltration associated with these three components is assumed to follow a uniform distribution bounded by the 95% confidence intervals found in the study. The third source of uncertainty is that associated with the parameters of the fitted empirical relations between indoor temperature and E-value, and likelihood of mould and relative humidity, also determined in the Warm Front study [13] [16]. The uncertainties in these two relations are accounted for by generating a different realisation of the original data used to construct the relations for each Monte Carlo simulation, assuming Gaussian errors. The new realisations are each fitted with smoothing splines that are slightly different of each other, therefore propagating the uncertainty in the relation to the predicted E-value and mould likelihood.

For the health impacts, a disease risk coefficient based on a meta-analysis of the relationship between mould and asthma risk [11] is used. It is assumed that uncertainty in this estimate is

represented by a uniform distribution bounded by $\pm 10\%$ around the central estimate. It is assumed further that the uncertainty in prevalence is represented by a uniform distribution prior bounded by $\pm 10\%$ around the central estimate. Although it can be argued that variability in the quality weighting of any disease could be taken into account because this represents different utilities attached to the burden of the disease, it was not considered for simplicity of exposition. Furthermore no consideration was taken of any uncertainty or variability in the number of dwellings affected by the interventions or in the average number of children living in a dwelling.

2.4 Propagation of uncertainty in multiplicative models

Equation (4) is a multiplicative model with three independent random (uncertain) variables q , p and $\Delta\mathcal{L}_{MSI>1}$. u, n, c are treated as constant. All the variables can be considered to be positive. q and p are strictly positive because q is a risk coefficient and p is prevalence. $\Delta\mathcal{L}_{MSI>1}$ can always be assumed to be positive because health impact can be formulated as health gain if $\Delta\mathcal{L}_{MSI>1} < 0$ and as health burden if $\Delta\mathcal{L}_{MSI>1} > 0$.

As shown in Figure 1, two methods were employed to quantify the uncertainty in the health impact (h). In the first method (fully probabilistic), we sample the values of the building physics model parameters and the health model parameters simultaneously and independently from their respective pdfs and then construct the empirical pdf of h from its sample values obtained by multiplying through the variables on the right hand side of Equation (4). In the second method (semi-analytical), we use algebraic methods based on Mellin integral transforms [25] to propagate the distribution of the mould exposure analytically through to the health model and obtain the distribution of the health impact.

The Mellin transform (MT) of a random variable maps the random variable in probability space into an algebraic expression and one of its key properties is that the MT of the product

of independent random variables is the product of the MTs of the random variables (Appendix C). Using this property and ignoring for the time being the constants in Equation (4), we express the MT of the health impact as the product of the MTs of the terms on the right hand side of Equation (4):

$$M_{f_h}(s) = M_{f_{\Delta\mathcal{L}_{MSI>1}}}(s) \times M_{f_q}(s) \times M_{f_p}(s) \quad (5)$$

where $f_{\Delta\mathcal{L}_{MSI>1}}$, f_q and f_p are respectively the pdfs of $\Delta\mathcal{L}_{MSI>1}$, q and p ; $M_{f_{\Delta\mathcal{L}_{MSI>1}}}$, M_{f_q} and M_{f_p} are respectively the MT of $\Delta\mathcal{L}_{MSI>1}$, q and p .

Based on Equation (5) we use the following steps to calculate the uncertainty in h .

- Calculate the pdf of $\Delta\mathcal{L}_{MSI>1}$ via probabilistic simulations and then calculate its MT
- Determine the pdfs of q and p based on information from the literature and calculate their respective MTs
- Calculate the product of the MTs using equation (5)
- Calculate the exact analytical expression of the inverse MT, if possible. If not, calculate the exact mean, variance and higher order moments of h .

3. Results

The results are presented in the chronological order of the steps of the method described above, starting with the calculations of the change in the likelihood of MSI exceeding unity.

3.1 Probability density function of the likelihood of change in mould severity index exceeding unity

As noted above, there are several sources of variability and uncertainty which contribute to the total variation in $\Delta\mathcal{L}_{MSI>1}$. We will not be able to show the uncertainty in each of the

sources. As an example, we show below the uncertainty associated with one of the fitted empirical function. Figure 3 shows the baseline fit of the likelihood of mould severity index exceeding unity with saturated vapour pressure.

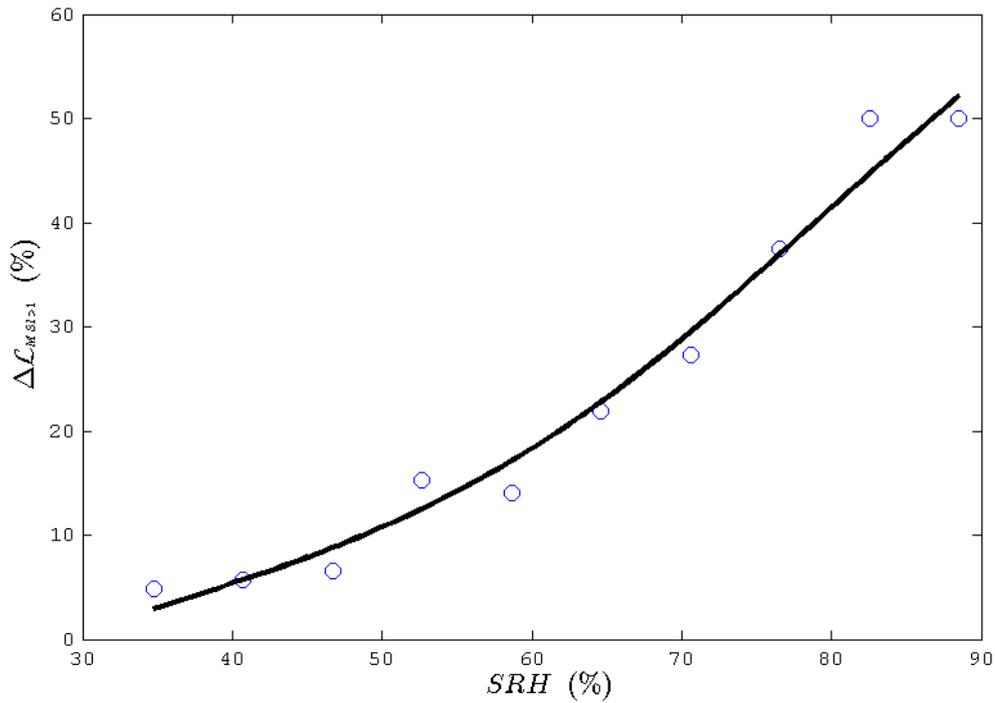


Figure 3. The observations from the Warm Front data (circles) along with the fitted line of the likelihood of mould severity index exceeding unity with saturated vapour pressure.

For this source, we take into account the uncertainty in the fitted parameters.

MC simulations were carried out to quantify the uncertainty in $\Delta\mathcal{L}_{MSI>1}$ (Figure 4)

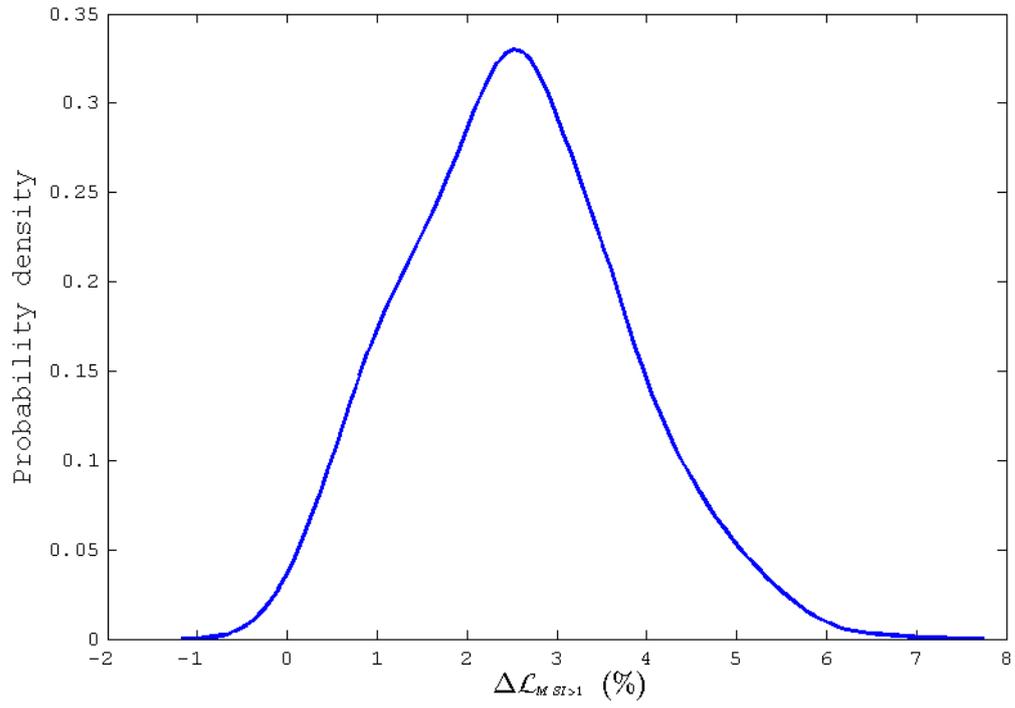


Figure 4. Empirical probability density function of $\Delta\mathcal{L}_{MSI>1}$.

A normal and a log-normal distribution are made to fit the cumulative probability distribution function of $\Delta\mathcal{L}_{MSI>1}$ (Figure 5)

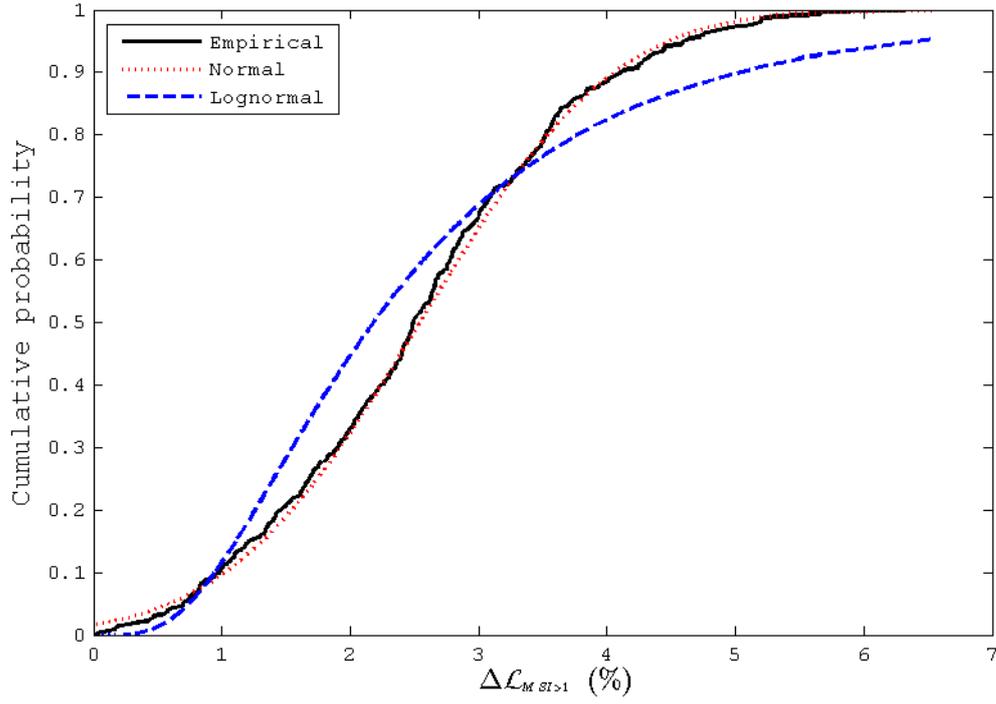


Figure 5. The empirical cumulative probability density function of $\Delta\mathcal{L}_{MSI>1}$ along with best fitted normal and log-normal distributions.

It is clear that the normal distribution provides a better fit to the pdf. The fitted normal distribution is given by the following equation:

$$f_{\Delta\mathcal{L}_{MSI>1}}(\Delta\mathcal{L}_{MSI>1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\Delta\mathcal{L}_{MSI>1} - \mu)^2}{2\sigma^2}\right) \quad (6)$$

where μ is the mean and σ^2 is the variance.

Using *Mathematica* [26], we derive an analytical expression for the MT of $f_{\Delta\mathcal{L}_{MSI>1}}$ is derived:

$$M_{f_{\Delta L_{MSI>1}}}(s) = \frac{2^{-2+\frac{s}{2}}}{\sqrt{\pi}\sigma} \left(\frac{1}{\sigma^2}\right)^{-\frac{s}{2}} \left(\sqrt{2} G\left(\frac{s}{2}\right) \times F\left(\frac{1-s}{2}, \frac{1}{2}, -\frac{\mu^2}{2\sigma^2}\right) + 2\mu \sqrt{\frac{1}{\sigma^2}} \times G\left(\frac{1+s}{2}\right) \right. \\ \left. \times F\left(1 - \frac{s}{2}, \frac{3}{2}, -\frac{\mu^2}{2\sigma^2}\right) \right) \quad (7)$$

where $G(\cdot)$ and $F(\cdot)$ are the Gamma and Hypergeometric functions respectively (Appendix D).

3.2 Probability density functions of health-related parameters

As stated above, it is assumed that the central estimates of the risk coefficient q is uniformly distributed between q_{min} and q_{max} , i.e. its probability density function is given by:

$$f_q(q) = \left\{ \begin{array}{l} \frac{1}{q_{max} - q_{min}} \quad \text{if } q_{min} \leq q \leq q_{max} \\ 0 \quad \text{if } q < q_{min} \text{ or } q > q_{max} \end{array} \right\} \quad (8)$$

and that the prevalence of asthma is uniformly distributed between a minimum p_{min} and a maximum p_{max} :

$$f_p(p) = \left\{ \begin{array}{l} \frac{1}{p_{max} - p_{min}} \quad \text{if } p_{min} \leq p \leq p_{max} \\ 0 \quad \text{if } p < p_{min} \text{ or } p > p_{max} \end{array} \right\} \quad (9)$$

Using *Mathematica*, the MTs of q and p are derived respectively as:

$$M_{f_q}(s) = \frac{q_{max}^s - q_{min}^s}{s (q_{max} - q_{min})} \quad (10)$$

$$M_{f_p}(s) = \frac{p_{max}^s - p_{min}^s}{s (p_{max} - p_{min})} \quad (11)$$

3.3 Probability density function of health impact

Using Equation (5) the MT transform of the health impact is given by:

$$\begin{aligned}
 M_{f_h}(s) &= \frac{2^{-2+\frac{s}{2}}}{\sqrt{\pi} \sigma} \left(\frac{1}{\sigma^2} \right)^{-\frac{s}{2}} \left(\sqrt{2} G\left(\frac{s}{2}\right) \times F\left(\frac{1-s}{2}, \frac{1}{2}, -\frac{\mu^2}{2\sigma^2}\right) + 2\mu \sqrt{\frac{1}{\sigma^2}} G\left(\frac{1+s}{2}\right) \right. \\
 &\quad \left. \times F\left(1 - \frac{s}{2}, \frac{3}{2}, -\frac{\mu^2}{2\sigma^2}\right) \right) \times \left(\frac{q_{max}^s - q_{min}^s}{s(q_{max} - q_{min})} \right) \\
 &\quad \times \left(\frac{p_{max}^s - p_{min}^s}{s(p_{max} - p_{min})} \right) \tag{12}
 \end{aligned}$$

Because of the complexity of the expression on the right-hand side of Equation (12), it is not possible to obtain an analytical solution of the inverse MT. Although there are numerical procedures to approximate the solution of an inverse MT [27], we opted instead to get analytical expressions for the mean, and higher order moments of the health impact. The variance is used as a measure of uncertainty. The mean (\bar{h}) and variance (v_h) of the health impact are given respectively by (Appendix C)

$$\begin{aligned}
 \bar{h} = M_{f_h}(2) &= \frac{1}{2\sqrt{\pi} \sigma} \left(\frac{1}{\sigma^2} \right)^{-1} \left(\sqrt{2} G(1) \times F\left(-\frac{1}{2}, \frac{1}{2}, -\frac{\mu^2}{2\sigma^2}\right) + 2\mu \sqrt{\frac{1}{\sigma^2}} G\left(\frac{3}{2}\right) \right. \\
 &\quad \left. \times F\left(0, \frac{3}{2}, -\frac{\mu^2}{2\sigma^2}\right) \right) \left(\frac{q_{max} + q_{min}}{2} \right) \left(\frac{p_{max} + p_{min}}{2} \right) \tag{13}
 \end{aligned}$$

$$v_h = M_{f_h}(3) - M_{f_h}^2(2) \tag{14}$$

where:

$$\begin{aligned}
M_{f_h}(3) = & \frac{2^{-\frac{1}{2}}}{\sqrt{\pi} \sigma} \left(\frac{1}{\sigma^2} \right)^{-\frac{3}{2}} \left(\sqrt{2} G\left(\frac{3}{2}\right) \times F\left(-1, \frac{1}{2}, -\frac{\mu^2}{2\sigma^2}\right) + 2\mu \sqrt{\frac{1}{\sigma^2}} G(2) \right. \\
& \left. \times F\left(-\frac{1}{2}, \frac{3}{2}, -\frac{\mu^2}{2\sigma^2}\right) \right) \times \left(\frac{q_{max}^3 - q_{min}^3}{3(q_{max} - q_{min})} \right) \left(\frac{p_{max}^3 - p_{min}^3}{3(p_{max} - p_{min})} \right) \quad (15)
\end{aligned}$$

and $M_{f_h}^2(2)$ is the square of the right hand side of Equation (13).

3.4 Numerical results

As stated in the previous section, the empirical pdf of $\Delta\mathcal{L}_{MSI>1}$ was obtained by Monte Carlo simulations. The best fitted normal distribution to the pdf gave a mean $\mu = 2.3752 \times 10^{-2}$ and a standard deviation $\sigma = 1.3294 \times 10^{-2}$, respectively. The lower and upper bounds of the uniform distribution of the log-risk coefficients were specified as $q_{min} = 0.3827$ and $q_{max} = 0.4678$, and the lower and upper bounds of the uniform distribution of the prevalence were specified $p_{min} = 0.0144$ and $p_{max} = 0.0176$ ($\pm 10\%$ around the mean). The remaining constants in Equation (13) are $u = 0.1$, $n = 1$ and $c = 0.4198$. Using Equations (23) and (24) gives the mean health impact per dwelling as -6.840×10^{-6} QALYs and the standard deviation as 3.729×10^{-6} QALYs (or -178 QALYs and 97 QALYs respectively for England).

For comparison purposes, we carried out a full MC simulation for the whole chain of models (i.e. building physics and health model). Figure 6 shows the estimated probability density of the health impact per dwelling. The mean and standard deviation of the health impact were -6.768×10^{-6} QALYs and 3.893×10^{-6} respectively which are close to the values obtained using Mellin transforms.

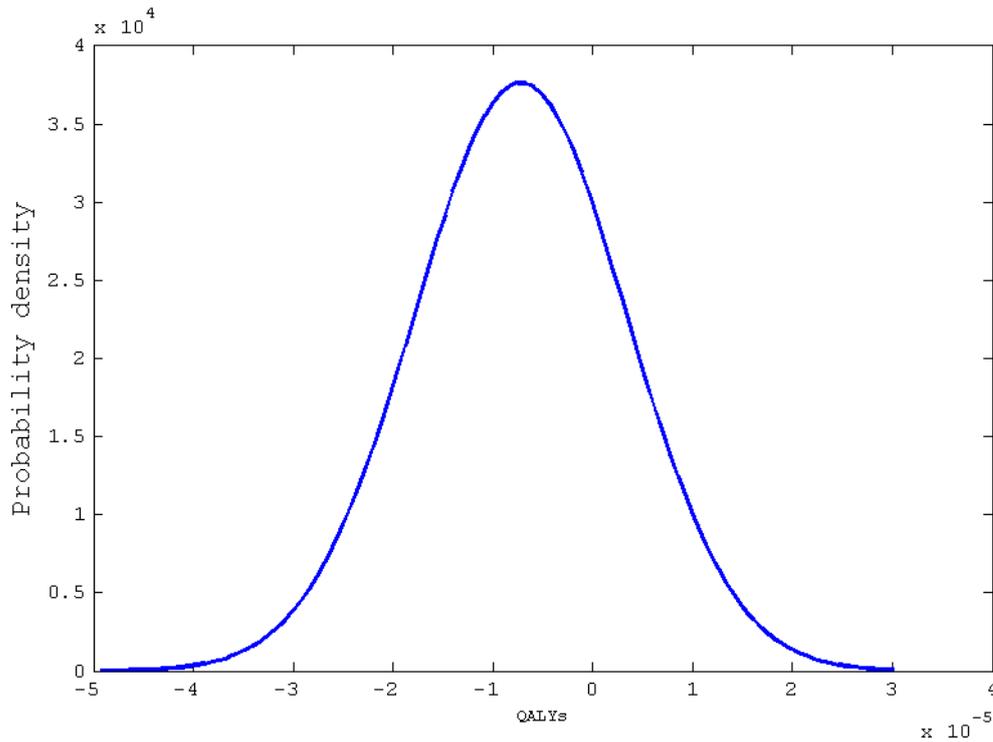


Figure 6. The empirical probability density function of the health impact.

5. Discussion

Monte Carlo simulations or Latin Hypercube Sampling are widely used to quantify uncertainty in model outputs [28-29]. The classical probabilistic approach for propagating parametric uncertainty between a series of distinct models in which the output of one model is an input to another model (e.g. building physics to health) is to sample the values of the parameters of all the models simultaneously from their respective probability density functions (pdfs) and then calculate the output variable(s) of interest for each combination of parameter values by running through the chain of models. When the models take considerable computing time to calculate the baseline values of relevant outputs, meta-modelling is used to approximate the model and perform the uncertainty analysis [30-31]. Such methods have been applied for quantifying parametric uncertainty in health impact assessment of

environmental interventions (of which draught proofing is an example) where the uncertainty of the environmental exposures are propagated numerically to the health outcomes [32].

Although the above computational approach holds merit particularly because it is easy to implement in practice, it suffers from four disadvantages:

- The approach can be computationally demanding although it could be argued this is not an issue with increasing computing power, the parallelisation of simulation algorithms [33] and use of meta-modelling [30-31].
- The contribution of the uncertainty of each model is not quantified separately prior to propagating it to the next model in the chain.
- If the uncertainty in a parameter or a set of parameters of one model is revised, it is necessary to re-do probabilistic simulations for the whole series of models.
- The uncertainties in some of the models in the chain could only be available as outcome uncertainties (e.g. when they are calculated *a priori*).

To address some of the above disadvantages, we proposed an alternative semi-analytical approach for propagating uncertainty. This approach quantifies the uncertainty in each model separately (using common probabilistic methods when necessary) but the propagation of uncertainty between the chain of models is done in algebraic space using integral transforms rather than in probability space. We then compared the classical MC approach with the alternative approach we proposed. Although both approaches gave approximately the same results, the proposed approach has several merits. It isolates the uncertainty in each model before propagating the uncertainty between the series of models. As such, if the uncertainties in the parameters of one model are changed, it is not necessary to re-do the probabilistic simulations for all the models. It is sufficient to quantify the uncertainty in the affected model only and then use the MT to propagate the uncertainties. Because of its stepwise approach in

dealing with the uncertainty in each model separately and then propagating the uncertainty between the models, it can be argued that the analytical method for propagation of uncertainty is more transparent and efficient than a full numerical method which is applied across all models simultaneously.

Naturally there are disadvantages to the proposed approach too. It is not always possible to determine exactly an analytical MT or an inverse MT. Although there are symbolic processing tools such as *Mathematica* (which were used in this study) to perform symbolic processing, an analytical solution may be intractable. In this case, some analytical approximations can be made. The overall model considered in this study is a multiplicative type. However there could be situations when the model is not purely multiplicative. In this situation the analytical solution becomes more complex. The table below summarises the main advantages and disadvantages of the method:

Advantages	Disadvantages
Quantifies the uncertainty in one model output before propagating it through to the input of another model	The integral transforms could be difficult to calculate exactly
If the uncertainty in the parameters of one model are revised, it is not necessary to perform the probabilistic simulations for the whole chain of models	Ideal for multiplicative models in a chain but more difficult to use for non-multiplicative models
Less computationally demanding	Not easy to implement

Table 1. Main advantages and disadvantages of the proposed method compared to full Monte Carlo method.

Conclusions

A semi-analytical method for quantifying the uncertainty in the health impact of a housing intervention has been demonstrated. The standard method for quantifying the uncertainty in the overall output of a series of models in which the output of one model is an input to

another model is to use probabilistic simulation. In the case of parametric uncertainty, this entails sampling randomly the parameters of all the models simultaneously from their pre-defined probability density functions. An alternative method is proposed in which the propagation of the uncertainty between the models is done algebraically rather than numerically using integral transforms. Compared to MC method, the main advantage of the proposed method is that it isolates the uncertainties in the models prior to propagating them through the chain of models. The disadvantage of this method is that it requires the analytical calculation of the Mellin Transform which can be unwieldy for complex distributions.

Acknowledgment

This research was funded from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no 265325 (PURGE). The funder had no role in study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the paper for publication. We thank the reviewers for their constructive comments on the original manuscript.

Appendix A: Building Physics Model

All the calculations are made at an hourly time step but time dependence is not shown in the equations to simplify the mathematical notation. In the equations, we will differentiate between two types of parameters: physics constants and model parameters subject to either variability or uncertainty. The physics constants are denoted by the vector $\boldsymbol{\theta}$ whereas the model parameters are described by the five vectors \boldsymbol{a} , \boldsymbol{b} , \boldsymbol{c} , \boldsymbol{d} , \boldsymbol{e} representing different parameterisations. A reference to an element in a vector is denoted by the name of the vector with a subscript e.g. a_1 is an element of \boldsymbol{a} and θ_1 is an element of $\boldsymbol{\theta}$.

The CONTAM outputs are hourly values of the humidity ratio (ratio of water vapour mass to total air mass) in each zone of the dwelling. They are processed using empirical relations to calculate $\mathcal{L}_{MSI>1}$. This is done in five steps. The first step determines the indoor or internal vapour pressure excess (VPE) and then standardises it ($SVPE_5$) to “winter conditions” (defined as external temperature $\tilde{T}_e = 5$ °C and external relative humidity 80%). The second step calculates the indoor temperature standardised to winter conditions. This is known as the standardised indoor temperature (SIT_5). The third step calculates the internal and external saturated vapour pressure standardised also to winter conditions ($VP_{sat,i}$ and $VP_{sat,e}$ respectively). The fourth step calculates the internal standardised relative humidity ($SRH_{5,80}$) and the final step calculates $\mathcal{L}_{MSI>1}$. $\mathcal{L}_{MSI>1}$ is calculated before and after draught proofing of the dwelling to estimate the change in the likelihood, $\Delta\mathcal{L}_{MSI>1}$. Some of the dwelling characteristics such as the permeability and E-value (the required energy consumption by the main heating device to maintain a one degree Celsius temperature difference between inside and outside during steady state conditions and ignoring incidental heat gains and ventilation heat losses) change because of the intervention.

Starting with the first step, CONTAM is used to generate hourly concentrations of the indoor humidity ratio of the dwelling:

$$r_i = \zeta(\mathbf{a}) \tag{A.1}$$

Assuming that the total indoor and outdoor air pressure are much greater than the indoor and outdoor vapour pressure respectively, the hourly VPE is calculated using:

$$VPE = \theta_1(r_e - r_i) \tag{A.2}$$

where r_e is the outdoor humidity ratio.

Standardised vapour pressure excess $SVPE_5$ is calculated by fitting a linear regression line between VPE and external temperatures (T_e):

$$VPE = b_1 T_e + b_2 \quad (A.3)$$

and then substituting $T_e = \tilde{T}_e = 5$ °C in the regression Equation [A.3] to give $SVPE_5$.

In the next step, SIT_5 in the living room is calculated from the E-value of the dwelling using an empirical relation determined from Warm Front data [16]:

$$SIT_5 = \psi(\mathbf{c}, E) \quad (A.4)$$

Only mould in the living room is considered. In the following step, the indoor and outdoor saturated vapour pressures are calculated using the following two physics-based equations respectively:

$$VP_{sat,i} = \theta_2 \exp\left(\frac{\theta_3 SIT_5}{\theta_4 + SIT_5}\right) \quad (A.5)$$

$$VP_{sat,e} = \theta_2 \exp\left(\frac{\theta_3 \tilde{T}_e}{\theta_4 + \tilde{T}_e}\right) \quad (A.6)$$

Using Equation (A.5) and (A.6), we can calculate $SRH_{5,80}$ (%) via:

$$SRH_{5,80} = 100 \times \frac{SVPE_5 + \theta_6 VP_{sat,e}}{VP_{sat,i}} \quad (A.7)$$

Finally, $\mathcal{L}_{MSI>1}$ is determined from $SRH_{5,80}$ using another empirical relationship derived from Warm Front data [13]:

$$\mathcal{L}_{MSI>1} = \varphi(\mathbf{d}, SRH_{5,80}) \quad (A.8)$$

In both Equations (A.4) and (A.8), the empirical fitting uses smoothing splines and cross-validation to fix the smoothing parameters.

The mapping of the pre-intervention dwelling characteristics (C_{pre}) to the post-intervention characteristics (C_{post}) as a result of draught proofing is determined using SAP:

$$C_{post} = \xi(\mathbf{e}, C_{pre}) \quad (A.9)$$

where \mathbf{e} is a vector of parameters representing the relation between the pre- and post-dwelling characteristics. The specific forms of the function ξ and parameter vector \mathbf{e} in Equation (A.9) are generic and their specific forms are different for the various dwelling characteristics.

Appendix B: Inputs of the CONTAM models

As discussed in the Methods section, a single-floor flat archetype is used based on one of the flats discussed at length in [18-19]. The indoor conditions between 1st October and 31st May are simulated using CIBSE's London Test Reference Year weather file at 10s intervals, outputting conditions every 15 minutes.

Dwelling geometry

The archetype consists of five zones. The total ground floor area and height of the flat is given by the randomly selected entry in the English Housing Survey, but the fractional contribution to the total ground floor area of each zone is given below:

Zones					
Room	Kitchen	Living	Bedroom	Entrance	Bathroom

Floor area (fraction of total)	0.11	0.40	0.28	0.14	0.07
---	------	------	------	------	------

Ventilation

Uncontrolled ventilation occurs between the indoor and outdoor environments through the permeability of the exposed facades. This is modelled by placing one crack at the top of the exposed facade, and one at the bottom [18]. The permeability can then be varied by the value of the crack coefficient. The windows are assumed to be not opened during the winter months. There are additional extract fans in the kitchen and bathroom due to a lack of exposed facades in these zones. The ventilation rates and schedules of these fans are given below:

Intermittent extract ventilation rates and schedules			
Zone	Extract rate (l/s)	Day	Schedule
Kitchen	60	Weekday	07:30-08:30 18:00-19:30
		Weekend	08:30-09:30 12:00-12:30

			18:00- 19:30
		Weekday	07:30- 08:30 18:00- 19:30
		Weekend	08:30- 09:30 12:00- 12:30 18:00- 19:30
Bathroom	15		

Contaminant generation

Only moisture is considered as a pollutant. The moisture production rate in each zone is expressed as a fraction of a maximum rate of 1359g/h. The fraction of this maximum allocated to each zone and the schedule for the generation is given below and is based on previous work [18]:

Moisture generation rates and schedules		
Zone	Day	Schedule and fraction of maximum rate generated
Kitchen	Weekday	07:30-07:45 (0.188)

		07:45-08:00 (0.229) 08:00-08:15 (0.262) 08:15-08:30 (0.255) 18:00-19:30 (0.631)
	Weekend	08:30-08:45 (0.188) 08:45-09:00 (0.229) 09:00-09:15 (0.262) 09:15-09:30 (0.255) 12:00-12:30 (0.483) 18:00-19:30 (0.631)
	Weekday	17:00-18:00 (0.107) 18:00-18:30 (0.067) 18:30-19:30 (0.107) 19:30-20:30 (0.114) 20:30-21:30 (0.148) 21:30-22:00 (0.041)
Living	Weekend	08:30-08:45 (0.188) 08:45-09:00 (0.229) 09:00-09:15 (0.262) 09:15-09:30 (0.255) 12:00-12:30 (0.483)

		18:00-19:30 (0.631)
Bedroom	Weekday	00:00-07:00 (0.044)
		07:00-07:45 (0.041)
	Weekend	22:00-23:00 (0.081)
		23:00-00:00 (0.044)
Bathroom	Weekday	00:00-08:00 (0.044)
		08:00-08:45 (0.041)
	Weekend	22:00-23:00 (0.081)
		23:00-00:00 (0.044)
Bedroom	Weekday	07:00-07:30 (0.926)
		07:30-08:00 (0.033)
		18:00-19:00 (0.148)
		19:00-19:30 (0.369)
		19:30-20:30 (0.845)
		20:30-21:00 (0.369)
	Weekend	08:00-08:30 (0.926)
		08:30-09:00 (0.033)
		18:00-19:00 (0.148)
		19:00-19:30 (0.369)
		19:30-20:30 (0.845)
		20:30-21:00 (0.369)

Appendix C: Mellin Transform

The Mellin transform of a positive random variable x is given by [6] [25]:

$$M_{f_x} = \int_0^{\infty} x^{s-1} f_x(x) dx \quad (C.1)$$

where f_x is the probability density function (pdf) of x . The inverse of the Mellin transform is given by:

$$f_x(x) = \frac{1}{2\pi i} \int_{\lambda-i\infty}^{\lambda+i\infty} x^{-s} M_{f_x}(s) ds \quad (C.2)$$

The Mellin transform has a number of interesting basic properties. The expected value of x^s is:

$$\bar{x^s} = \int_0^{\infty} x^s f_x(x) dx \quad (C.3)$$

From Equation (C.1), (C.3) can be written as:

$$\bar{x^s} = \int_0^{\infty} x^{(s+1)-1} f_x(x) dx = M_{f_x}(s+1) \quad (C.4)$$

which shows the relationship between the expected value of x^s and its Mellin transform. The expected value and variance of x can be easily obtained by applying Equation (C.4) for a specific value of s :

$$\bar{x} = M_{f_x}(2) \quad (C.5)$$

$$\overline{x^2} - \bar{x}^2 = M_{f_x}(3) - M_{f_x}^2(2) \quad (C.6)$$

The Mellin transform of the product of two independent random variables $z = xy$ is the product of their Mellin transforms:

$$M_{f_z} = M_{f_x} M_{f_y} \quad (C.7)$$

If z is equal to the product of several independent random variables $z_1 \dots z_k$ and a constant c :

$$z = c \prod_{i=1}^k z_i \quad (C.8)$$

then the expected value, variance and coefficient of variation of z are given by respectively

$$\bar{z} = c \prod_{i=1}^k M_{f_{z_i}}(2) \quad (C.9)$$

$$\overline{z^2} - \bar{z}^2 = c^2 \left(\prod_{i=1}^k M_{f_{z_i}}(3) - \prod_{i=1}^k M_{f_{z_i}}^2(2) \right) \quad (C.10)$$

Appendix D: Gamma and Kummer confluent hypergeometric function

The Gamma function is defined as:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (D.1)$$

For example, $\Gamma(4) = 6$.

The Kummer confluent hypergeometric function is defined by [34]:

$$\Psi(a, b, z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{zt} t^{a-1} (t-1)^{b-a-1} dt \quad (D.2)$$

As an example, $\Psi(2,4,3) = 5.575$

References

- [1] Wilkinson P, Smith KR, Davies M, Adair H, Armstrong BG, Barrett M, Bruce N, Haines A, Hamilton I, Oreszczyn, Ridley I, Tonne C, Chalabi Z. Public health benefits of strategies to reduce greenhouse-gas emissions: household energy. *Lancet* 2009; 374: 1917-1929.
- [2] Das P, Chalabi Z, Jones B, Milner J, Shrubsole C, Davies M, Hamilton I, Ridley I, Wilkinson P. Multi-objective methods for determining optimal ventilation rates in dwellings. *Building and Environment* 2013; 66: 72-81.
- [3] Milner J, Shrubsole C, Das P, Jones P, Ridley I, Chalabi Z, Hamilton I, Armstrong B, Davies M, Wilkinson P. Home energy efficiency and radon related risk of lung cancer: modelling study. *British Medical Journal* 2014; 348, f7493. doi: 10.1136/bmj.f7493
- [4] Springer MD. *The algebra of random variables*. New York: Wiley; 1979.
- [5] Glen AG, Leemis LM, Drew JH. Computing the product of two continuous random variables. *Computational Statistics & Data Analysis* 2004; 44 (3): 451-464.
- [6] Tung Y-K. Mellin transform applied to uncertainty analysis in hydrology/hydraulics. *Journal of Hydraulic Engineering* 1990; 116 (5): 659-674.
- [7] Hu X. *Probability methods of industrial situations using transform techniques*. MSc, Industrial and Manufacturing Systems Engineering, Ohio University; 1995.
- [8] Tyagi A, Hann CT. Reliability, risk, and uncertainty analysis using generic expectation functions. *Journal of Environmental Engineering* 2001; 127 (10): 938-945.

[9] HM Government. The building regulations 2000 (2010 edition): Approved document F1: means of ventilation; 2010.

http://www.planningportal.gov.uk/uploads/br/BR_PDF_ADF_2010.pdf (last accessed July 2015).

[10] Jones BM, Das P, Chalabi Z, Davies M, Hamilton I, Lowe R, Milner J, Ridley I, Shrubsole C, Wilkinson P. The effect of party wall permeability on estimations of infiltration from air leakage. *International Journal of Ventilation* 2013; 12: 17-29.

[11] Fisk WJ, Lei-Gomez Q, Mendell MJ. Meta-analyses of the association of respiratory health effects with dampness and mould in homes. *Indoor Air* 2007; 17(4): 284-296.

[12] Mendell MJ, Mirer AG, Cheung K, Tong M, Douwes J. Respiratory and allergic health effects of dampness, mould, and dampness-related agents: a review of the epidemiological evidence. *Environmental Health Perspectives* 2011; 119 (6): 748-756.

[13] Oreszczyn T, Ridley I, Hong SH, Wilkinson P. Mould and winter indoor relative humidity in low income households in England. *Indoor and Built Environment* 2006; 15 (2): 125-135.

[14] EHS. English Housing Survey; 2009.

<http://discover.ukdataservice.ac.uk/catalogue?sn=6804> (accessed January 2014)

[15] Haddix AC, Teutsch SM, Corso PS. Prevention effectiveness. A guide to decision analysis and economic evaluation. Second Edition. Oxford: Oxford University Press; 2003.

[16] Oreszczyn T, Hong SH, Ridley I, Wilkinson P. Determinants of winter indoor temperatures in low income households in England. *Energy and Buildings* 2006; 38 (3): 245-252.

- [17] Walton GN, Dols WS. CONTAM 2.4 User guide and program documentation. Gaithersburg, MD, US; 2008.
- [18] Shrubsole C, Ridley I, Biddulph P, Milner J, Vardoulakis S, Ucci M, Wilkinson P, Davies M. Indoor PM_{2.5} exposure in London domestic stock: modelling current and future exposures following energy efficient refurbishment. *Atmospheric Environment* 2012; 62:336-343.
- [19] Das P, Chalabi Z, Jones B, Milner J, Shrubsole C, Davies M, Hamilton I, Ridley I, Wilkinson P. Multi-objective methods for determining optimal ventilation rates in dwellings. *Building and Environment* 2013; 66:72-81.
- [20] Hamilton I, Milner J, Chalabi Z, Das P, Jones B, Shrubsole C, Davies M, Wilkinson P. Health effects of home energy efficiency interventions in England: a modelling study. *BMJ Open* 2015; doi:10.1136/bmjopen/-2014-007298.
- [21] CIBSE. Guide J – Weather, solar and illuminance data, 7th Edition, London, CIBSE Publications; 2002.
- [22] Davies M, Ridley I, Moore G, Chalabi Z, Wilkinson P, Hutchinson E, Sands J, Hillyard R. Costs and benefits of fitting trickle ventilators in replacement windows. Report BD2625, Department for Communities and Local Government, London; 2010.
- [23] BRE. The Government's Standard Assessment Procedure for energy ratings of dwellings, 2009 Edition, incorporating RdSAP 2009. Building Research Establishment. http://www.bre.co.uk/filelibrary/sap/2009/sap-2009_9-90.pdf (accessed January 2014).
- [24] Hong SH, Ridley I, Oreszczyn T. The impact of energy efficient refurbishment on the airtightness in English dwellings, in: 25th Air Infiltration and Ventilation Centre Conference: Ventilation and Retrofitting, Prague Czech Republic, pp 7-12; 2004.

- [25] Bertrand J, Bertrand P, Ovarlez J-P. Mellin transform. In *The transform and applications handbook*, Third Edition, Ed. AD Poularkis, Chapter 12, pp 12.1-12.38, CRC Press, Boca Raton FL.; 2010.
- [26] Wolfram S. *Mathematica* 10; 2014. (<http://www.wolfram.com/>)
- [27] Tagliani A. Recovering a probability density function from its Mellin transform. *Applied Mathematics and Computation* 2001; 118: 159-159.
- [28] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* 2003; 81: 23-69.
- [29] Janssen H. Monte-Carlo based uncertainty analysis: sampling efficiency and sampling convergence. *Reliability Engineering & System Safety* 2013; 109: 123-132.
- [30] Allaire D, Wilcox K. Surrogate modeling for uncertainty assessment with application to aviation environmental system models. *American Institute of Aeronautics and Astronautics Journal* 2010; 48 (8): 1791-1803.
- [31] Manfren M. Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. *Applied Energy* 2013; 103: 627-641.
- [32] Mesa-Frias M, Chalabi Z, Vanni T, Foss AM. Uncertainty in environmental health impact assessment: quantitative methods and perspectives. *International Journal of Environmental Health Research* 2012; DOI: 10.1080/09603123.2012.678002.
- [33] Feyen L, Vrugt JA, Nualláin BÓ, Van Der Knijff J, De Roo A. Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISAFLOOD model. *Journal of Hydrology* 2007; 332: 276-289.

[34] Weisstein E. Confluent hypergeometric function of the first kind. MathWorld- A Wolfram web source; 2013.

<http://mathworld.wolfram.com/ConfluentHypergeometricFunctionoftheFirstKind.html>