International Journal of Building Pathology and Adap

Inferring the as-built air permeability of new UK dwellings

Journal:	*International Journal of Building Pathology and Adaptation*
Manuscript ID	IJBPA-02-2019-0018.R1
Manuscript Type:	Original Article
Keywords:	airtightness, mathematical modelling, compliance, inference, domestic energy use

Inferring the as-built air permeability of new UK dwellings

Abstract

Compulsory airtightness testing was introduced for new dwellings in England and Wales in 2006 and in Scotland in 2010 to ensure that they are constructed according to design air permeability targets. These targets are set to limit heat loss through air infiltration. Previous work examining the large ATTMA dataset of UK airtightness test data suggested that, in a proportion of dwellings, the targets were being met by post-completion sealing as opposed to airtight construction, but did not quantify the prevalence of this practice.

In this paper, the distribution of as-built airtightness and the proportion of dwellings undergoing post-completion sealing are estimated from the ATTMA dataset covering 2015-2016. This is carried out by Bayesian statistical modelling, using the dataset of recorded test results and a modelled representation of the testing process. This analysis finds the mode of the as-built distribution of air permeability as 4.38 ± 0.01 m^3/m^2h. It predicts that 39% of dwellings aiming for one of the five most common design targets have sealing interventions at the point of pressure testing to meet their target. The as-built distribution of the ATTMA data is compared to airtightness test data obtained from just before compulsory testing was introduced, showing an improvement in the modal air permeability of 3.6 m^3/m^2h since testing became mandatory.

This article has ~~investigated more deeply than the reported data in order to estimate~~ investigated the available data beyond simply what is reported, to estimate what the real levels of airtightness in the UK new build stock may be.

Keywords

Airtightness; distributions; compliance; sealing; Bayesian inference

1. Introduction

Airtightness testing of new dwellings became mandatory in England and Wales in 2006 (Office of the Deputy Prime Minister, 2006) and in Scotland in 2010 (Scottish Government, 2010). This was partly as a result of the limited data available at the time showing that over 75% of dwellings were not as airtight as their design specification (Office of the Deputy Prime Minister, 2004); therefore the stated aim of this new testing requirement was to reduce the number of poorly sealed buildings (Office of the Deputy Prime Minister, 2004).

Under the UK's compulsory testing regime, a design air permeability target is first set which enables a dwelling to meet an overall CO_2 emission target. An airtightness test is then carried out on the completed dwelling on a sample basis, and if the result is higher (i.e. leakier) than the target, the dwelling must undergo remedial work and be re-tested until its permeability is less than or equal to the target. Alternatively, as long as the dwelling meets the minimum statutory airtightness target of 10 m^3/m^2h at 50 Pascals, the design target may be relaxed if an improvement elsewhere in calculated carbon emission performance can be made to compensate or if the dwelling would still meet the target carbon emissions using the higher air permeability test result (HM Government, 2016)~~. (Government, 2016).~~

The airtightness test protocol in the UK is managed by the Air Tightness Testing and Measurement Association (ATTMA) (ATTMA, 2016~~.~~). ATTMA additionally runs an approved Competent Persons Scheme for airtightness testers whose requirements include lodgement of all regulatory tests, meaning a large dataset of test results is now available. Analysis of this data in a previous article (Love et al., 2017) indicated unusual structures in the distribution of airtightness results, with a disproportionately high number of measurements just meeting the targets and a sharp drop immediately in excess of them. From combining the data with the results of small scale case studies finding evidence of post-completion temporary and secondary sealing, it was proposed that such distributions may be partly attributable to in-test interventions being made in order to pass the regulatory test. This was framed as *'hitting the target and missing the point'*, since if the proposition is correct, then the presence of a target air permeability does not lead ~~lead not~~ to most dwellings achieving their target as intended through improvements to the primary air barrier. Instead, a significant proportion of dwellings are built leakier and require the use of temporary or potentially short-lived remedial measures to attain the target. This conflicts with the intended outcome of the regulations: a stock of dwellings with durable airtightness performance.

The previous analysis of the ATTMA dataset focussed on collecting evidence for the existence of mechanisms allowing dwellings to pass their tests sub-optimally, but did not attempt any quantitative estimation of the fraction of homes meeting air permeability design targets at the initial test compared to those meeting the design targets through sealing either during or after a regulatory test. This is important as it is useful to know how intrinsically airtight dwellings are at the point of completion using current construction processes, as well as how prevalent the occurrence of secondary sealing is in the cases where the homes do not initially meet their targets.

An estimation of these values was attempted in this paper using statistical modelling of the airtightness test results in the ATTMA dataset. A mathematical model has been developed of the test procedure and impact of post-completion interventions on the distribution of airtightness test results. Fitting the model to the observed data enabled estimation of key modelled parameters: the shape of the distribution of air permeability when new homes are built, and the fraction of dwellings which did not meet their target airtightness and therefore underwent some kind of later sealing. The best estimate of the as-built distribution of airtightness test results was then compared to a dataset of airtightness measurements taken before the introduction of compulsory testing (Grigg, 2004). This analysis enabled insights to be drawn about the impact of the introduction of this compulsory airtightness testing regime on the as-built airtightness of the stock.

The rest of this paper is structured as follows. Section 2 provides a description of the data and its key structure. Section 3 introduces the statistical modelling methodology to be used. Sections 4 and 5 develop the model conceptually and mathematically and Section 6 describes the analysis. The key values derived are presented in Section 7 and discussed in Section 8. Finally, the Conclusion in Section 9 summarises the key learnings.

2. Data

This paper uses two datasets, one each from after and before the introduction of compulsory testing. The ATTMA data, testing procedures, metadata and data cleaning process are reported in detail by Love et al (~~Love et al., ~~2017); with key features summarised in section 2.1. The pre-regulatory testing dataset used for comparison is described in section 2.2.

2.1 ATTMA dataset

Approximately 130,000 air permeability test results from the UK are lodged with the ATTMA database annually; this paper presents the analysis of data from a 1.5 year period, from August 2015 to December 2016. In addition to the test results, the lodged data also includes a small amount of metadata about the building and the test, as well as the inputs to the air permeability calculation. It does not include a unique building number or test timestamp.

The dataset was cleaned to remove physically unreasonable results (for example flow exponent below 0.5) and erroneous completion of the fields (for example missing envelope area). An algorithm was then applied to identify unique dwellings and detect the last test per dwelling in instances of multiple tests recorded per dwelling. This process reduced the dataset from 192,731 to 144,024 results and introduced a small error associated with the imperfect identification of unique dwellings and tests being mislabelled as last tests. The reader is referred to (Love et al., 2017) for a fuller description of the cleaning and processing steps.

The test data and the design targets taken from the metadata are shown in two different ways in Figure 1-. The boxplot on the right displays the full range of values, while the histograms on the left display only values between 0 and 10 m^3/m^2h at 50 Pascals where the majority of the data lies. Note that for the rest of this article, all design and measured air permeability values are given in 10 m^3/m^2h at 50 Pascals but quoted simply as 10 m^3/m^2h for brevity. ~~shows the frequency distribution of air permeability test results using 0.01 m^3/m^2h bins, the highest resolution common to the whole dataset. The distribution of declared design targets taken from the metadata is shown in the lower subplot.~~

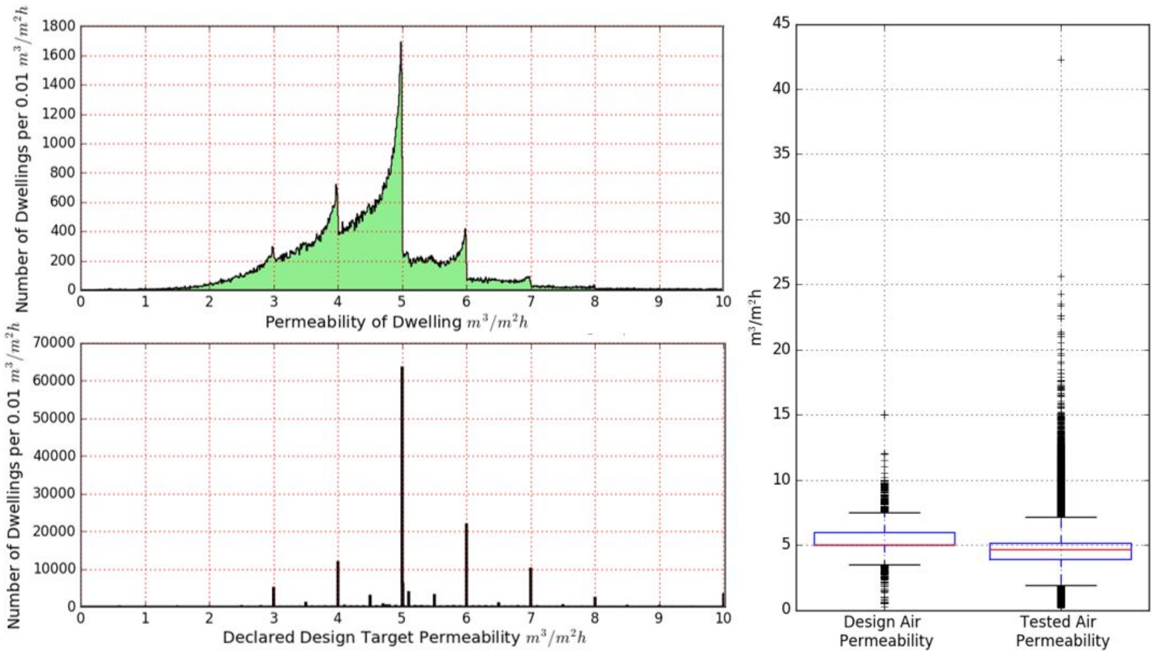

Figure 1. Top left: d~~D~~istribution of last tests. Bottom left: distribution of design targets. Right: both of these distributions in boxplot form ~~(top) and declared design targets (bottom) from the ATTMA dataset~~. All air permeabilities are reported at 50 Pa.

~~Figure 1~~

~~Figure 1 shows the frequency distribution of air permeability test results using 0.01 m^3/m^2h bins, the highest resolution common to the whole dataset. The distribution of declared design targets taken from the metadata is shown in the lower subplot.~~The distribution of ~~last~~ test results recorded in the ATTMA database was discussed in detail in previous work (Love et al., 2017). Three particularly relevant observations for the analysis presented in this paper are discussed below.

Firstly, there is a clear mismatch between the range of reported permeabilities, from 0.17 to 42.27 m^3/m^2h at 50 Pascals, and the range of reported design targets, from 0.24 to 15.0 m^3/m^2h. A significant minority of dwellings have last test results higher than their~~; some test results exceed their~~ design target; the reason for this is unclear. Furthermore it is not known why a small number of dwellings have a design target above 10 m^3/m^2h. Secondly, there are 450 identifiable unique design targets, the most common being 5.0 m^3/m^2h for 63,467 dwellings (44% of the dataset). Figure 1 highlights the large peaks followed by very sharp drops in last test results associated with common design targets, discussed in Section 1. As the 15 most common design targets in the metadata are associated with 95% of the dwellings, such trends dominate the visual trends observed in ~~the~~ Figure 1. Third~~Final~~ly, 137 design targets have only one associated dwelling; 100 of these dwellings have the design target exactly equal to the test result. It is therefore possible that such design targets were set after the test, or changed in response to the test result.

It is not possible to infer the original design and build intent for ~~such~~ situations where tested permeability remains higher than design permeability, or where the target may have been changed; better metadata would be required to resolve these issues. ~~For these reasons~~ Iit appears that the stated design target in the database is not always a reliable indicator of the actual airtightness level originally being aimed for. Therefore, in the analysis presented in this work the design target for each dwelling is not taken from the database and is instead inferred from the data.

2.2 Comparison dataset

One component of the analysis described in this paper is a comparison of the ATTMA data to a dataset collected before the introduction of mandatory testing. This comparison dataset was ~~of the as-built permeability distribution derived from the ATTMA data to that from an air permeability dataset of dwellings constructed and tested before the introduction of mandatory testing in England and Wales in 2006, which were~~ collected by Grigg for the Energy Saving Trust (Grigg, 2004)~~. This comparison dataset will~~ and thus will be referred to as the ‘EST data’.~~,~~ It~~and~~ contains air permeability test results carried out in 2004 for 99 new dwellings in England under the same test protocol as the ~~. The airtightness test protocol used for the EST data was the same as that used to collect the~~ ATTMA data. However, the EST tests were carried out to ascertain the state of airtightness in new dwellings at the time, as opposed to demonstrating compliance with building regulations. The EST dataset therefore represents the as-built state of new dwellings built in accordance with the April 2002 edition of Approved Document L1 and would therefore have not undergone any test-related sealing or remedial work.

3. Methodology

The approach taken in this paper is *statistical modelling*, whereby a model is applied to a dataset with the aim of extracting useful information about the data-generating process (Konishi and Kitagawa, 2008, Anderson, 2007). Using models to understand data is a standard technique in the field of energy and buildings.~~:~~ For example, at the simplest level it could entail linear regression (see e.g. ~~(~~Summerfield et al., 2010~~)~~), whilst~~;~~ other studies use more complex models (see e.g. ~~(~~Rouchier et al., 2018~~)~~). The appropriate level of model complexity should take into account the amount of structure present in the data – representing the complexity of the phenomenon being described as well as the ability to measure it – and the extent to which the phenomenon is understood and evidenced (Ashdown, 2018). While all models in the field of energy and buildings are abstractions of reality, some explicitly attempt to describe causal processes such as physical laws (e.g. ~~(~~Biddulph et al., 2014~~)~~) and others encapsulate a large number of unknown physical and social processes without trying to model them explicitly (e.g. ~~(~~Huebner et al., 2016~~)~~). The approach in this paper is situated in between these two approaches: we develop a mathematical representation of the socio-technical process of airtightness testing.

Statistical modelling is applied as follows. We begin with a dataset of test results and a conceptual model of the test procedure outlined in Section 4, indicating that an unknown proportion of dwellings have already undergone remedial works before their recorded test. This insight is then formalised into a mathematical model of the test procedure in Section 5, and applied to the data in Section 6. The useful information to be extracted is how airtight the dwellings were before remedial works were carried out and the proportion requiring these works.

An essential component of statistical modelling is model specification: choosing a mathematical description of the system which can reproduce the main features of the data (Lehmann, 1990). Statistical models are expressed as probability distributions because they incorporate some random variation due to limited information (Davison, 2003, Pawitan, 2001). That is, they describe a system in which the origin of each data point does not need to be specified deterministically by the model, but the distribution of all data points should be able to be recreated. This is a suitable modelling approach for this analysis because of a lack of knowledge of the characteristics of each individual dwelling. For example, it is not known whether the airtightness test result for a particular house came about from passing its regulatory test first time or from undergoing remedial works. However, it is assumed that in the entire sample of 140,000 homes, a large number of processes (involving design, construction and the test procedure) combine to generate a small set of well-defined probability distributions. The intention of the analysis is then to determine the nature of these distributions and the parameters that control them. There are several established techniques for comparing models and learning parameters from data (MacKay, 2003). ~~; i~~In this paper Bayesian inference is used (Congdon, 2007).

4. Conceptual model of airtightness testing

A simple conceptual model of the current airtightness testing process, shown in Figure 2, is used as the basis for the analysis in this paper. The conceptual model is derived from simplifying Figure 8 in Love et al (~~Love et al.,~~ 2017), and attempts to encapsulate the range of possible testing processes that may occur.

The two halves of the model in Figure 3 differ according to whether a full test is actually completed before any interventions to the building fabric take place. The left hand side of the model depicts a cycle in which a full test is carried out first, and if the result is a fail, then interventions to the building fabric are then carried out, with feedback on the level of improvement provided to the tester through re-testing. This process is termed here as ‘loose feedback’, as remedial work is carried out each time without knowledge of exactly how much effect the remedial work will have on decreasing air permeability. These interventions may therefore lead to a better result than the design target, in some cases by a large amount. This will still lead to a sharp cut-off at the design target value, but will not result in an accumulation of dwellings immediately below the design target.

The right hand side of the model in Figure 3 shows a scenario in which a full test is not carried out initially. Instead, an initial ‘check test’ is performed in which the blower door fan is left running to give a 50 Pascal pressure difference in order for the tester to determine from the fan flow rate whether the dwelling would pass a full test. If not, in-test sealing and remedial measures are carried out with the fan still running, with the tester able to continuously monitor the fan flow rate to immediately determine the impact of any remedial measures on the airtightness level. W~~Only w~~hen the tester is confident that the dwelling will pass, a ~~is a~~ full test is then then carried out. This process is termed here as ‘tight feedback’, as the tester has constant knowledge of the air permeability at 50 Pascals, and remedial measures can stop immediately upon reaching the design target and will therefore lead to a large number of dwellings with an air permeability immediately below the design target. One of the key differences between the proposed “loose feedback” and “tight feedback” mechanisms is that only one full regulatory test would be carried out in the “tight feedback” case, whereas the number of number of full regulatory tests in the “loose feedback” case could range from one test to a number of tests.

Unfortunately, it is not possible to separate out these two different feedback mechanisms from this data alone, as it is impossible to know how much of which feedback mechanism is being used. We therefore expect a permeability distribution which is a mixture of the as-built, tight and loose feedback distributions. The distribution for a particular design target is expected to be similar to the as-built distribution up to the design target with a sharp peak just below the design target if there is significant tight feedback.

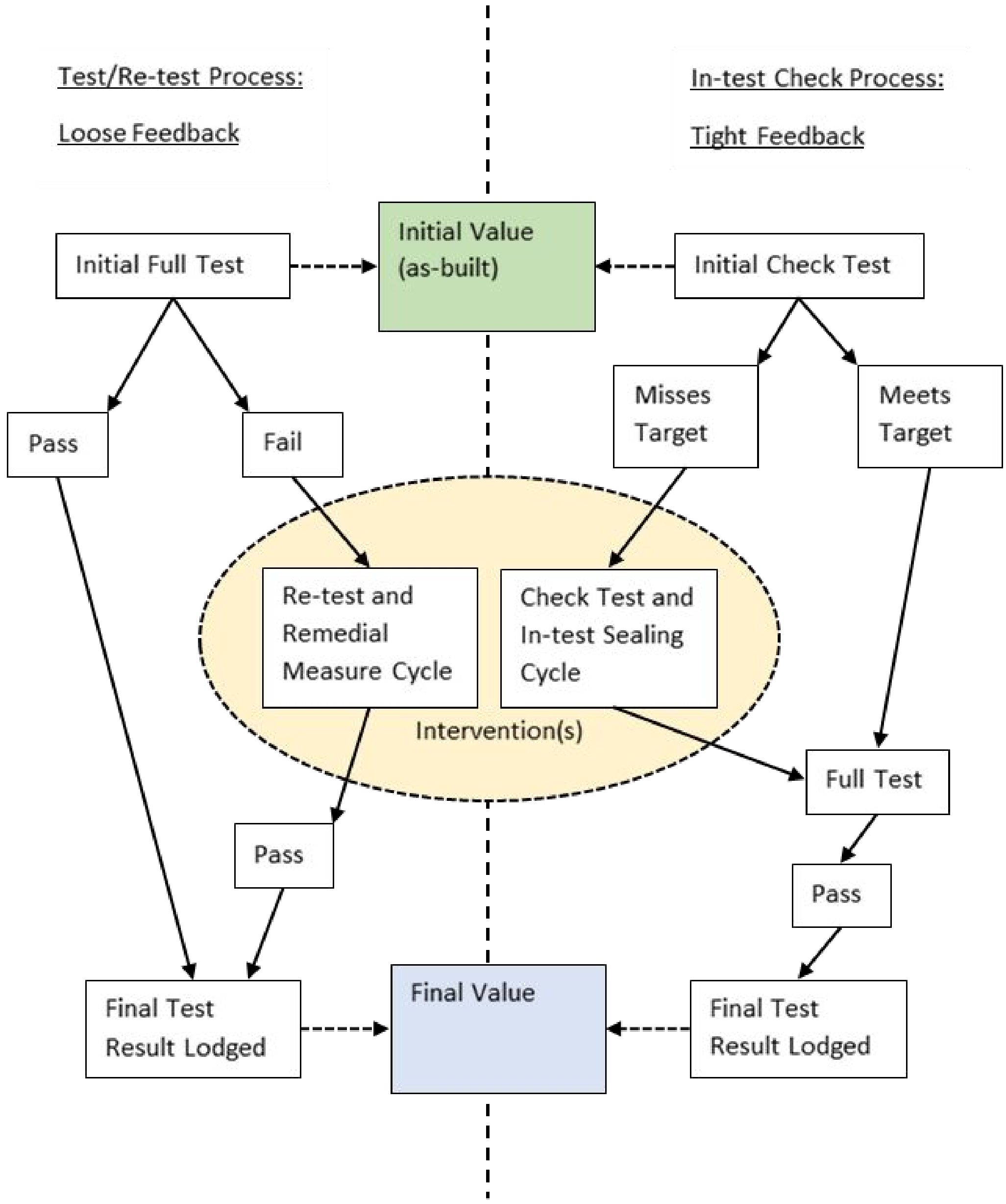

Figure 2. Conceptual model of the airtightness testing process, showing 'loose feedback' (left) and 'tight feedback' (right).

The aim of the analysis in this article is to use the distribution of 'Final Value' data, shown at the bottom of Figure 2, to estimate the initial as-built air permeability distribution of the sample, shown at the top of Figure 2, and to estimate what proportion of dwellings have undergone sealing interventions post-completion to enable them to meet their airtightness target. To undertake this, the conceptual model in Figure 2 is first expressed as a statistical model, below.

5. Statistical model of airtightness testing process

We now express the above conceptual model of the testing process in terms of probability distributions. Two distinct phases of the process for constructing dwellings to a certain airtightness have been identified above and are treated separately in this model: dwelling construction itself, and post-construction sealing and remedial measures.

It is assumed that dwelling construction results in an air permeability taken from an initial as-built *background distribution*, as shown in the top box in Figure 2. Previous work on Finnish dwellings by Vinha et al (~~Vinha et al., ~~2015) discovered a non-normal distribution for the background distribution, but did not propose a specific mathematical form. Therefore, a lognormal distribution with parameters μ and σ is selected here, since it has characteristics consistent with the expected distribution of measurements discussed in previous work (Love et al., 2017);~~:~~ it is broad, smooth, has a single mode, is always positive and has a long tail. The background probability distribution of any air permeability, $perm$, is therefore given by:

$$P_b(0 < perm) = \frac{1}{perm}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{ln(perm)-\mu}{\sigma}\right)^2} \quad (1)$$

Equation 1 represents the as-built distribution of air permeability of dwellings before any testing or intervention. The use of a single mode distribution implies a simplistic assumption that all dwellings are originally built to the same standard regardless of the design target associated with the building; the validity of this assumption is returned to later.

Using the background distribution, it is then possible to determine the fraction of dwellings, *F*, that do not meet their airtightness target when they are first measured by the tester. This is the fraction of dwellings that would have a permeability above a design target t_n, expressed in Equation 2 using the error function ERF.

$$F(t_n < perm) = \frac{1}{2}\left(1 + ERF\left(\frac{\mu - log(t_n)}{\sigma\sqrt{2}}\right)\right) \quad (2)$$

Next, it is hypothesised that dwellings whose as-built air permeability is above their design target undergo an intervention to bring the air permeability down to the target or below; this mechanism is represented in Figure 2 as the 'interventions' box. This is modelled using an *intervention distribution* for each design target, of exponential form, with a *rate parameter* λ m²h/ m³. The intervention distribution for each design target is given by:

$$P_I(0 < perm \leq t_n) = \lambda\left(1 + e^{-\lambda t_n}\right)e^{(-\lambda(t_n - perm))} \quad (3)$$

The exponential distribution attempts to capture both the tight feedback and loose feedback mechanisms described in Section 4, as it allows re-tests or in-test checks to fall either exactly on the target or to overshoot it. Note that a normalisation factor $\left(1 + e^{-\lambda t_n}\right)$ has been included to account for the tail of the exponential distribution which extends beyond zero.

The combination of the background and intervention distributions then gives a probability distribution with a sharp peak at the design target. This is given for one design target t_n by :

$$P_n(0 < perm \leq t_n) = P_b + FP_I$$

$$P_n(t_n < perm) = 0 \quad (4)$$

The effect of Equation 4 is that all dwellings in the background distribution which are greater than their design target are relocated to the intervention distribution. This is illustrated graphically in Figure 3 for an example target of 7 m³/m²h, with arbitrary parameters for the background distribution.

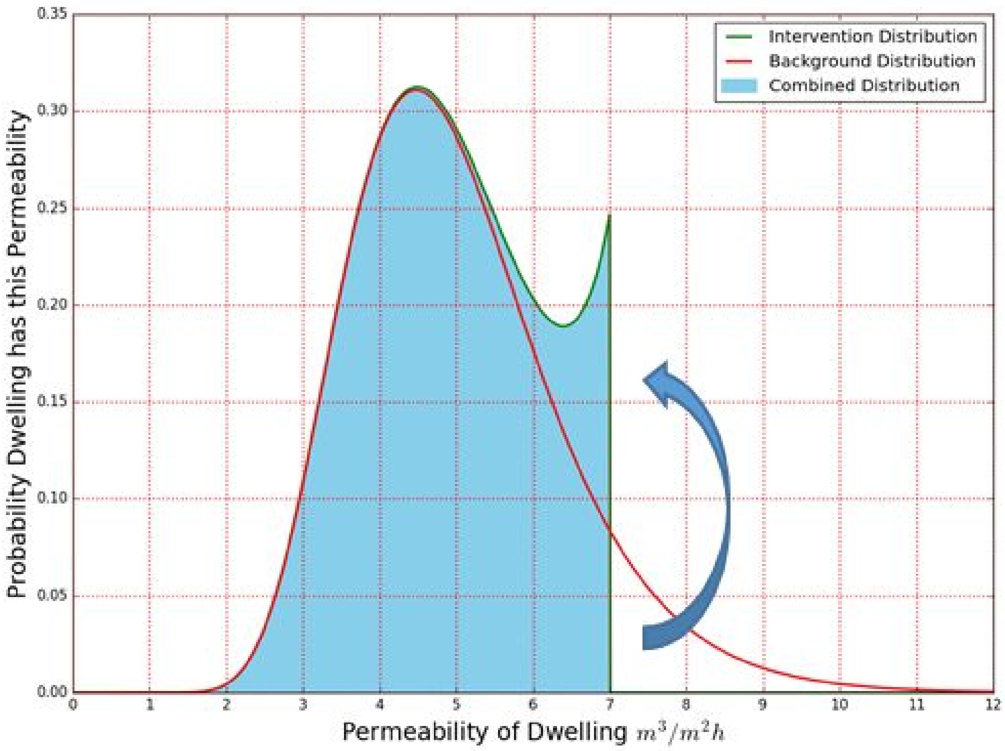

Figure 3. Illustration, using arbitrary parameters, of dwellings moving from the background to the intervention distribution, simulating the intervention part of the airtightness testing process.

To complete the mathematical description, the background and intervention distributions at different design targets are combined. Each design target has a fraction f_n of the total number of dwellings aiming for it. The set of design targets, and the fraction of dwellings aiming for each one, are predicted by the model, i.e. they are outputs not inputs. The ATTMA metadata contains a design target for each dwelling, but as stated in Section 2.1 there is evidence that this is often not the design target the builder was working towards. Allowing the model to predict the design targets and associated fractions of dwellings removes the need to trust the declared design targets.

Equation 5 gives the complete simulated distribution of recorded test results. The $f_c P_b$ term is included as a *catch-all distribution* for those design targets not included in the model. We assume that these dwellings are randomly taken from the background distribution, P_b. Equation 5 then represents the distribution of simulated test results, or the 'Final Value' box in Figure 2.

$$P(0 < perm) = \sum_{n=1}^{N} f_n P_n + f_c P_b \quad (5)$$

Finally, the rate parameter λ is not necessarily the same for dwellings aiming for different design targets. Dwellings which are constructed relatively airtight are expected to have smaller and potentially fewer defects (e.g. large cracks) than those that are leaky, reducing the impact of interventions on the permeability. For this reason, the modelled rate parameter was allowed to vary for different design targets. However, in order not to introduce too many extra parameters, we specify that the rate parameter must be a linear function of design target, determined entirely by two parameters: gradient *m* and intercept *c*. *m* and *c* are then *hyperparameters*.

In summary of the above, the parameters associated with the mathematical model are listed in Table 1.

Table 1. Parameters of model of airtightness testing process.

Parameter name	Interpretation
μ,σ	Parameters of the lognormal background distribution. Units: $\ln(m^3/m^2h)$
m_λ, c_λ	Hyper-parameters representing the gradient and intercept of the rate parameter λ of the intervention distribution. Units: h^2/m^2 and m^2h/m^3
$t_1, t_2 \ldots t_N$	Design targets. Units: m^2h/m^3
$f_1, f_2, \cdots, f_N$	Fraction of dwellings aiming for each design target in the set of *N* targets
f_c	Fraction of dwellings that have a design target not included in the model, so are instead in the catch all distribution. Note: $f_c = 1.0 - \sum_{n=1}^{N} f_n$

6. Analysis

The aim of the analysis was to estimate the following outputs:

- The mode of the background distribution representing the construction process;
- The fraction of dwellings which met their target in their as-built state, and those which did not meet the target and then underwent extra sealing;
- The improvement of the ATTMA background distribution from the comparison dataset collected before compulsory testing.

In order to estimate the above outputs, the parameters of the mathematical model listed in Table 1 must first be estimated from the data. Bayesian inference was used to estimate the most likely values of the parameters using a method closely following that applied by Biddulph et al (2014~~Biddulph et al., 2014~~) and Elwell et al (~~Elwell et al.,~~ 2015). Model parameters, θ, may be estimated from their joint probability distribution, $P(\theta|y,H)$, given the applied model, *H*, and the data, *y*, using the Bayes equation:

$$P(\theta|y,H) = \frac{P(y|\theta,H) \times P(\theta|H)}{P(y|H)} \quad (6)$$

$P(y|\theta,H)$ is the likelihood function, the probability of measuring the recorded data given the estimated parameters and model. $P(\theta|H)$ is the prior distribution, the estimated initial probability distribution of the parameters and $P(y|H)$ is the evidence, the probability of observing the recorded data given the model.

To compare model and data required the ability to compare a continuous function (the model) with a large set of discrete number in bins (the data). ~~Since the model is continuous and the data is a histogram of a large set of discrete numbers of dwellings falling into each permeability bin of width 0.01 m³/m²h, the~~ This was possible using the Binomial distribution ~~was used~~ as the likelihood function. This ~~to~~ related the number of dwellings within each bin in the data, *k*, to the average predicted number of dwellings given by the model over the bin's range, *a*:

$$P(y|\theta,H) = \frac{a^k e^{-a}}{k!} \quad (7)$$

Flat non-informative uniform priors with realistic ranges were used, since there was no information available on the values of μ, σ and λ. In principle, priors for f_n could have been formed using the number of dwellings associated with each declared design target. However, due to the large size of the dataset, the relative influence on the analysis of priors is very small, so uniform priors were used for simplicity.

A maximum a posteriori probability (MAP) estimate, the mode of the posterior distribution space, was obtained. This gives the most likely values of the parameters listed in Table 1, and their statistical error. However, the statistical error only relates to the particular model applied and does not mean that a model is 'right', or the best way of generating the data. Although there is no way to tell if a model is 'right', i.e. represents the actual processes occurring in the data, it is possible to compare alternative models against each other to find the most appropriate out of those tested. Specifically, the most appropriate model is that which explains the greatest amount of structure in the data with the lowest number of parameters. Within the Bayesian analysis framework, this is known as Bayesian Model Comparison . This enables comparison of the relative probability of different models describing the observed data, accounting for the increased size of the posterior parameter space in more complex models by penalising such models, and thus reducing the risk of choosing models which overfit the data.

Model comparison was applied in this case by creating a suite of nested models each incorporating a different number of design targets. For example, the simplest model was not allowed to incorporate any design targets, the second simplest was allowed to incorporate one target, and the most complex model was allowed to incorporate 15. As stated above, the targets were not provided as inputs but were instead found by the models. However, once a target had been found by the model with n parameters, the model with n+1 parameters also assumed the presence of this target. The Bayesian model evidence was calculated for each variant of the model, allowing quantitative comparison of which number of design targets best describes the data without overfitting it.

7. Results

The most suitable model is shown in Section 7.1. Its parameters are given in Section 7.2 which gives insight into the current status of airtightness construction in the UK. The progress made since compulsory testing began is then explored in Section 7.2.

7.1 Model comparison

The model using 5 design targets was shown to be the one with the highest Bayesian model evidence, as shown in Figure 4.

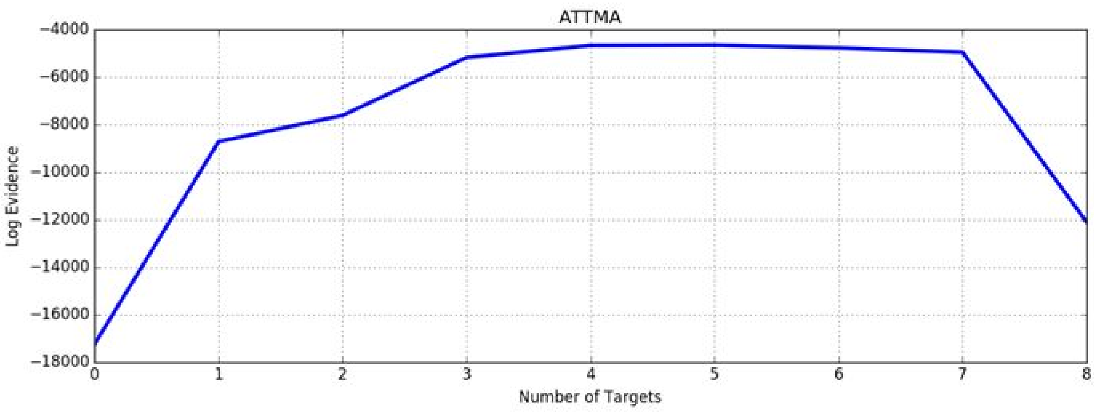

Figure 4. Model evidence for models with different numbers of parameters (design targets).

The set of design targets found by the model was, in order of number of dwellings associated with each target,~~:~~ ~~{~~5,6,7,4~~,~~3~~}~~ m^3/m^2h. Table 2 shows the model's prediction that 85% of dwellings are aiming for one of these 5 targets. This leaves 15% of dwellings associated either with a different target or no target. Whilst it is clear from the data that additional design targets do exist, the result of the model comparison shows that the data available does not justify the use of a model with the additional complexity introduced by more targets.

Table 2. Set of 5 targets as predicted by the best performing model. Modelled proportion of dwellings aiming for each target, and statistical error on this estimate.

Target Value (m^3/m^2h)	Modelled percentage of data associated with target (%)	Statistical error (%)
5	44	0.3
6	18	0.3
7	18	0.4
4	4	0.1
3	0	0.7
% dwellings aiming for one of above targets	85	
% dwellings not aiming for one of above targets	15	

7.2 Values of parameters using best model

Using the most probable model from the above model selection, the best estimate of the parameters are given in Table 3 with their statistical error. A plot of the model using these parameter estimates is shown in Figure 5, superposed on the data to allow visual comparison between the two.~~:~~ T~~t~~he fit is very good.

Table 3. Estimated background lognormal distribution and the intervention exponential distribution parameters for the model form the analysis.

Parameter	Value	Statistical error	Units
μ	1.60	0.00	$\log(m^3/m^2h)$
σ	0.35	0.00	$\log(m^3/m^2h)$
m_λ	-1.80	0.00	m^2h/m^3

c_λ	12.76	0.02	h^2/m^2

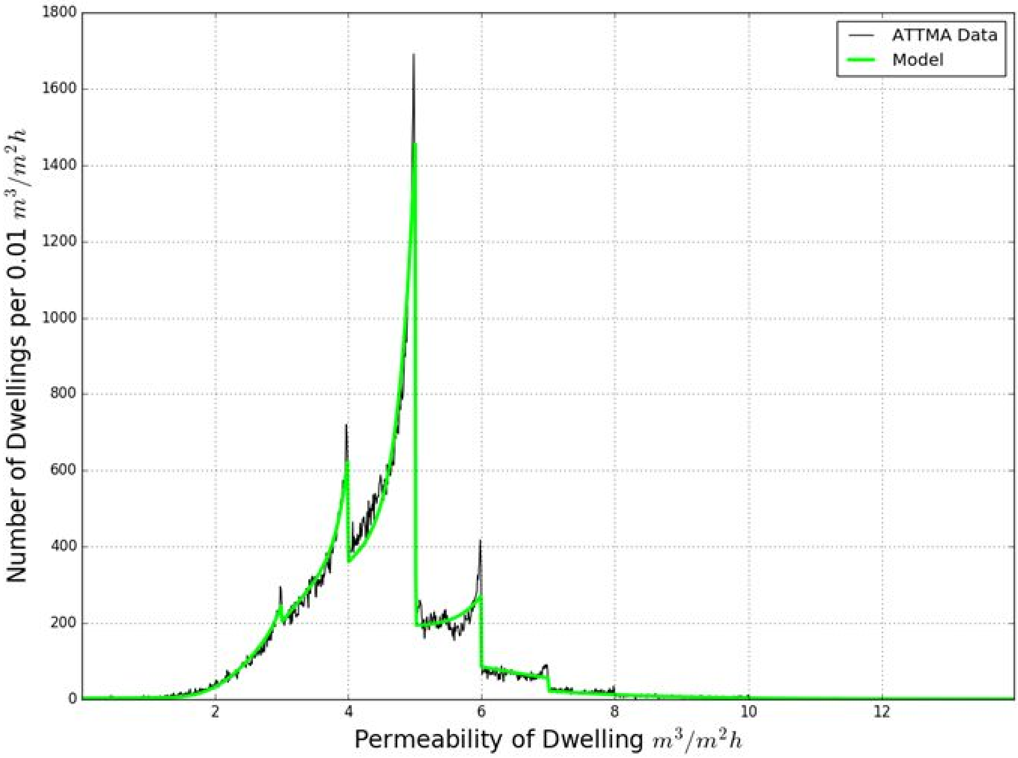

Figure 5. Model (green) superposed on data (black), showing the model for the air permeability of dwellings fit to the ATTMA data.

One important component of the model superposed on Figure 5 is the background distribution, representing our best estimate of the distribution of dwelling airtightness at the point of building completion. This is shown on its own in Figure 6. The background distribution, assumed to be log-normal, is defined by the two parameters μ and σ; however, these do not have a convenient physical interpretation. It is more meaningful to use these parameters to calculate the mode of the distribution, the mean, and ~~a descriptor of the width of the distribution:~~ the confidence interval[1] (a descriptor of the width of the distribution). The mode of the distribution is 4.38 $\pm$0.01 m^3/m^{2}h, the mean is 4.95 $\pm$0.01 m^3/m^{2}h, and the 67% confidence interval is 3.0-6.3 m^3/m^{2}h.

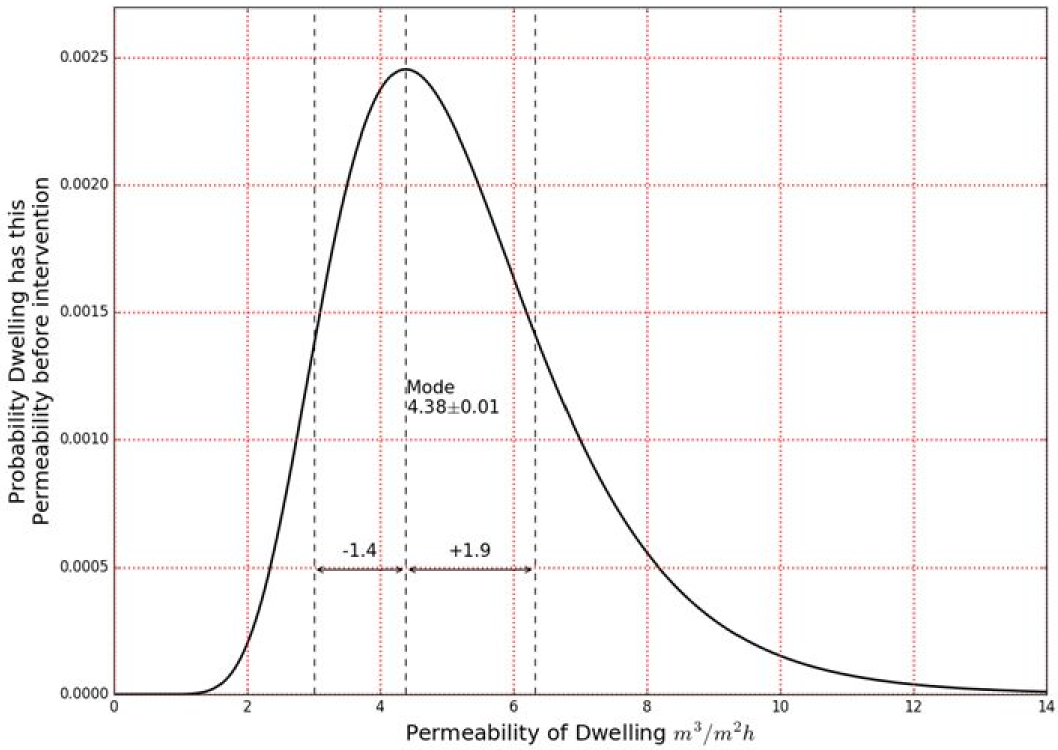

Figure 6. Predicted background distribution of dwelling airtightness, prior to any interventions. Distribution inferred from the model of the testing process and the data.

The second important component of the model is the representation of dwellings undergoing sealing interventions due to not meeting their design targets after construction. This is associated with the spikes in the model predictions in Figure 5. ~~For the subset (85%) of dwellings aiming for one of the 5 most common targets,~~ Table 4 shows the predicted result that ~~shows the model suggests that~~ 39% of the modelled sample of dwellings~~se dwellings~~ have undergone this process.

[1] Within the Bayesian framework, a 67% confidence interval is the smallest permeability interval which contains 67% of the data.

Table 4. Estimated percentage of dwellings requiring post-construction sealing. Note that this result only applies to the dwellings aiming for one of the 5 most common design targets.

Dwellings meeting target after construction (%)	Dwellings not meeting target after construction (%)	Statistical error (%)
61	39	0.4

The negative value of the parameter *m* in Table 3 suggests that a narrower intervention distribution is predicted for lower design permeability targets (the parameter m is negative). In practical terms, this could constitute evidence for the proposition in Section 5 that as buildings become more airtight, there are fewer opportunities for major additional sealing, since there are fewer obvious gaps in the building fabric. It is likely that design and construction teams that aim for targets below 4 m^3/m^2h are more knowledgeable and for the lowest levels, adopt different construction systems from those associated with the majority of dwellings.

7.2 Progress in airtightness of UK dwelling construction over time

In this section, the ATTMA background distribution inferred above is compared to the EST data described in Section 2.2. We use a very simple model developed by ~~(~~Lowe et al~~.,~~ (2000) to anticipate the effect of compulsory testing on the distribution of air permeability. Lowe et al proposed a simple scaling of the air permeability distribution, in which the mean and width both decrease. Their equation is given in Equation 8, where P_o is the distribution before introduction of compulsory testing, P is the distribution afterwards and a is a scaling factor.

$$P(perm) = aP_o(a \times perm) \qquad (8)$$

In order to utilise this model in the current study, firstly a lognormal curve was fitted to the EST data so that it had the same form as the ATTMA background distribution (as required by Equation 8). Secondly, the relationship between the EST and ATTMA lognormal curves was investigated. It was found that these curves scale very well: a single parameter (*a*=0.45) can highly accurately transform the EST distribution into the ATTMA background distribution, as is illustrated in Figure 7. Here, the EST data is displayed as a solid histogram, its lognormal curve as a solid line and the ATTMA background distribution as a dot-and-dash line. This supports the suitability of Equation 8 to describe the evolution of airtightness in the dwellings with the data analysed and models developed here.

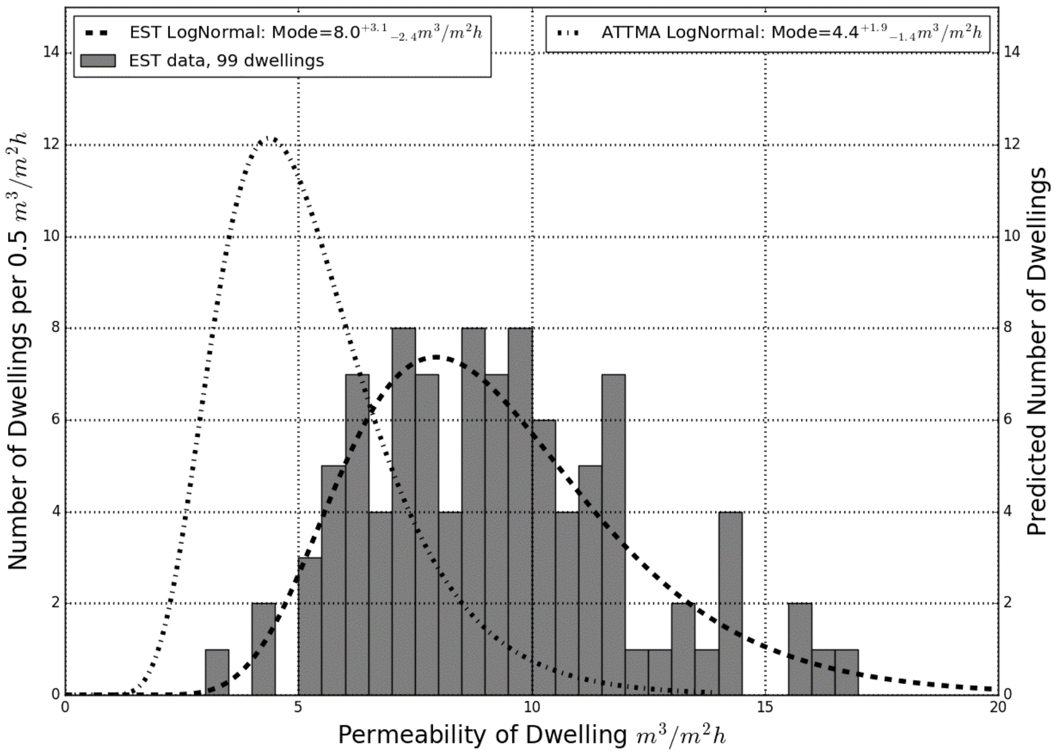

Figure 7. EST data, lognormal curve fitted to EST distribution and ATTMA background distribution. Y-axis for data on the left, y-axis for models on the right.

Figure 7 firstly shows that a lognormal curve is an appropriate distribution to represent the EST data;~~,~~ given the limited size of this dataset there is no evidence to suggest that an alternative distribution is a better representation of the data. Secondly, by comparing the EST modelled distribution with the ATTMA curve, the progress in airtightness construction can be observed. Between 2004 and 2017, the mode of the background permeability distribution changed from 8.0 to 4.4 m^3/m^2h: a decrease of 3.6 m^3/m^2h.

Following Lowe et al's method~~ [22]~~, we have taken Equation 8 together with the current background distribution to predict the required mode of a distribution that would enable 90% of dwellings to pass an airtightness test. This is equivalent to asking by how much the background distribution must improve to reduce the intervention rate to 10%. Two simple scenarios were used: in the first all dwellings aim for a target of 5 m^3/m^2h and in the second the target was 10 m^3/m^2h. Figure 8 shows the results of applying Equation 8 to these scenarios. Achievement of a 10% failure rate under a universal target of 5 m^3/m^2h requires a mode of 2.8 m^3/m^2h compared to its current mode of 4.38. Otherwise stated, since the background distribution contains a certain level of spread, then in order to ensure the performance of a stock of dwellings is below a required value, the mode of the background distribution must be well below that value.

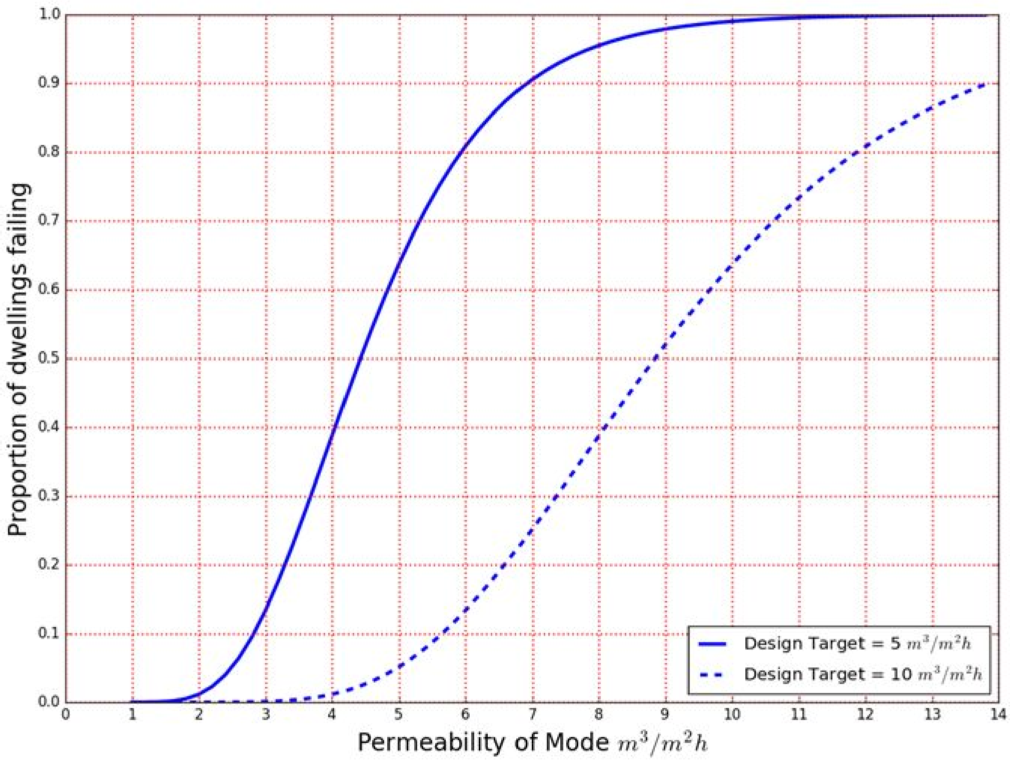

Figure 8. Application of the scaling rule to calculate the required mode to achieve different failure rates, for two design targets.

8. Discussion

It has been possible to recreate the unusual shape of the ATTMA data distribution (Figure 1) using a simple model which learns its parameters from the dataset. The model consisted of a lognormal background distribution and a set of exponential intervention distributions at five design targets located by the model. Figure 5 demonstrated that this model can recreate the main structure observed in the data, using 10 parameters.

The model recreates the sharp increase up to each main design target observed in the data, which is believed to be caused by post-construction sealing interventions. However, the model does not differentiate between types of intervention, nor the extent to which different types are permitted under the ATTMA protocol. It is likely that both types of intervention cycle shown in Figure 2 are present: the tight feedback cycle described by in-test checks and sealing, and the loose feedback cycle described by testing, remedial measures and re-testing, but we cannot determine what proportions.

8.1 Appropriate interpretation of model predictions

Before discussing the findings, it is important to highlight that these results are associated with one particular conceptual model of the airtightness testing process through one mathematical representation. Although different versions of the model were compared to select the one with the highest probability given the data, this does not render the chosen model 'valid'. Instead, we observe from Figure 5 that the chosen model provides a very good fit to the data using only 10 parameters, and therefore that the theory of interventions (possibly in-test) occurring before tests are reported is consistent with the data. However, the conceptual model in Figure 2 is a highly simplified description of the real testing process and does not capture other mechanisms likely to be present in the dataset, such as secondary sealing occurring after construction but before any airtightness testing is carried out. As such, the results are likely to be a conservative estimate of the number of dwellings undergoing secondary sealing.

With these important caveats noted, the results are now discussed.

8.2 Discussion of findings

Application of the model predicted that dwellings in the ATTMA dataset were constructed with a lognormal distribution with a modal air permeability of 4.38 $\pm$ 0.01 m^3/m^2h, and that 39% of the dwellings aiming for the five targets modelled had undergone some form of sealing intervention after an initial airtightness test. We did not estimate the proportion undergoing secondary sealing before any testing took place.

Considering first the distribution of airtightness before post-build interventions, the background distribution predicted from the ATTMA dataset may be compared to its equivalent before the introduction of mandatory testing in 2006. Results suggest that the modal air permeability in the UK has decreased by around 3.6 m^3/m^2h over the past 10 years and the distribution has narrowed by approximately 45%. This suggests that the combination of compulsory testing, design targets and the maximum regulatory air permeability limit of 10 m^3/m^2h have had a positive impact on airtightness performance.

Previous work has raised concerns about the longevity of post-construction sealing measures ((Leprince et al., 2017)Love et al., 2017).The current analysis cannot directly determine the likely future deterioration rate of remedial measures, but the widespread extent of post-construction sealing predicted by the model, together with the uncertainty about the long-term performance of such measures, presents a concern for the durability of the airtightness of the new dwelling stock.

It is expected that construction practices which ensure the effectiveness of the primary air barrier during construction are preferable to attempting to meet the airtightness target through post-construction sealing. Quality assurance within the construction process is therefore central to this performance. It is typical in manufacturing quality control and process improvement methodologies to ensure that the majority of the distribution of a measured parameter is well within any defined upper or lower performance specification limits. For example, in the Six Sigma process methodology, the ultimate objective is to achieve a process where the mean of a normally distributed process dataset is at least six standard deviations from any specification limit (Tennant, 2001, Coleman, 2008). A good example of the effective use of statistical methods in the UK construction industry is the use of random testing and statistical process control charts by Robust Details to monitor trends in the measured field performance of ~~the~~ different acoustic construction details in the Robust Details catalogue ((Wingfield et al., 2011), (Smith et al., 2006)).

If, as the model suggests, 39% of dwellings currently do not attain their targets upon construction, the QA processes of the industry are well below other sectors such as manufacturing. If a failure rate of under 10% were required, it was estimated in this analysis that the required background distribution mode would be 2.8 m^3/m^2h. Whilst this result provides a statistical estimate for illustrative purposes, it does not account for any significant changes to process in the construction of the primary air barrier (Johnston and Lowe, 2006) that could significantly change the shape of the distribution.

The analysis presented in this article indicates that the distribution of as-built air permeability for dwellings in the UK has improved, but that there is further progress to be made to ensure that most dwellings meet their airtightness targets on completion. To improve the as-built airtightness of the stock, and minimise the need for post-construction interventions, which are potentially only effective in the short term, it is suggested that testing and interventions are carried out throughout the production process rather than only measuring the dwelling at the point of completion. This pre-testing approach is normal for dwellings built to the Passivhaus standard, which recommends testing at three points in the building process: after construction of the primary air barrier, after main construction but before fitting out, and after fitting out (McLeod et al., 2014). This approach is however not common in mainstream house building in the UK.

9. Conclusion

This article investigated the premise that a proportion of the dwellings with test results recorded in the UK's largest dataset of airtightness test results, ATTMA, have already undergone post-construction sealing interventions to enable them to meet their design targets. The distribution of test data and a model of the testing process were used to infer the as-built air permeability of dwellings and the proportion not meeting their target in this state.

The analysis estimated the mode of the background distribution, representing the as-built airtightness of the new build stock, as 4.38 $\pm$ 0.01 m^3/m^2h. It was found that this distribution closely corresponds to a scaled transformation of an equivalent distribution obtained from data before design targets and compulsory testing were introduced. The improvement in air permeability between the EST data, recorded in 2004, and the ATTMA data, recorded in 2015/16 is a decrease of 45% in both mode and width, indicating more tightly controlled construction processes. However, it is estimated that 39% of properties initially fail tests and are subsequently sealed to close the performance gap. This high failure rate may be better addressed through better quality control during the construction process, including airtightness testing, to avoid the significant use of secondary sealing measures, such as caulk, which may be less robust than measures applied to the primary air barrier during construction. If the simple scaling of the airtightness distribution observed between the two datasets analysed here continues to hold, a mode of 2.8 m^3/m^2h would be required to achieve 90% of dwellings with air permeability under 5 m^3/m^2h on completion.

The statistical error on the quantitative results of this study is very low and the model was shown to perform better than a set of alternatives, but it is still a highly simplified representation of reality. The total uncertainty introduced by using this model to capture a complex and perhaps even unknown set of real world processes is unquantified. More comprehensive metadata and more reliable recording of data would allow a more sophisticated model to take other factors into account. These factors would include a better understanding of the theoretical form of the background distribution and whether it varies with different constructions, clearer understanding of the actual process involved when a dwelling fails a test and improved knowledge of how learning and experience from previous construction might influence the building process of subsequent dwellings. Airtightness performance in the UK is improving and it is important that this trajectory continues.

10. Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) funded 'Research Councils UK(RCUK) Centre for Energy Epidemiology' under EP/K011839/1, and the EPSRC London-Loughborough Centre for Doctoral Training in Energy Demand under EP/H009612/1.

11. Acknowledgements

The authors are grateful to Barry Cope from ATTMA for provision of the data.

References

ANDERSON, D. R. 2007. *Model based inference in the life sciences: a primer on evidence*, Springer Science & Business Media.

ASHDOWN, M. 2018. *Airtightness testing in the UK, across varied building typologies, before, during and after the introduction of regulatory testing: an inferential approach.* Master of Research in Energy Demand Studies, University College London.

ATTMA 2016. Technical Standard L1. Measuring air permeability in the envelopes of dwellings. September 2016 Issue. Amersham.

BIDDULPH, P., GORI, V., ELWELL, C. A., SCOTT, C., RYE, C., LOWE, R. & ORESZCZYN, T. 2014. Inferring the thermal resistance and effective thermal mass of a wall using frequent temperature and heat flux measurements. *Energy and Buildings,* 78**,** 10-16.

COLEMAN, S. 2008. Six Sigma: an opportunity for statistics and for statisticians. *Significance,* 5**,** 94-96.

CONGDON, P. 2007. *Bayesian statistical modelling*, Chichester: John Wiley & Sons.

DAVISON, A. C. 2003. *Statistical models*, Cambridge: Cambridge University Press.

ELWELL, C. A., BIDDULPH, P., LOWE, R. & ORESZCZYN, T. 2015. Determining the impact of regulatory policy on UK gas use using Bayesian analysis on publicly available data. *Energy Policy,* 86**,** 770-783.

HM GOVERNMENT, 2016. The Building Regulations 2010: Approved Document L1A: Conservation of Fuel and Power in New Dwellings. 2013 edition with 2016 amendments.

GRIGG, P. 2004. Assessment of Energy Efficiency Impact of Building Regulation Compliance. A Report Prepared for the Energy Savings Trust/Energy Efficiency Partnership for Homes, Client Report Number 219683, Building Research Establishment, Garston, Watford.

HUEBNER, G., SHIPWORTH, D., HAMILTON, I., CHALABI, Z. & ORESZCZYN, T. 2016. Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy,* 177**,** 692-702.

JOHNSTON, D. & LOWE, R. 2006. Improving the airtightness of existing plasterboard-lined load-bearing masonry dwellings. *Building Services Engineering Research and Technology,* 27**,** 1-10.

KONISHI, S. & KITAGAWA, G. 2008. *Information criteria and statistical modeling*, Springer Science & Business Media.

LEHMANN, E. L. 1990. Model specification: the views of Fisher and Neyman, and later developments. *Statistical Science*, 160-168.

LEPRINCE, V., MOUJALLED, B. & LITVAK, A. Durability of building airtightness, review and analysis of existing studies. 38th AIVC Conference: Ventilating healthy low-energy buildings, 2017 Nottingham, UK.

LOVE, J., WINGFIELD, J., SMITH, A., BIDDULPH, P., ORESZCZYN, T., LOWE, R. & ELWELL, C. 2017. 'Hitting the target and missing the point': Analysis of air permeability data for new UK dwellings and what it reveals about the testing procedure. *Energy and Buildings,* 155, 88-97.

LOWE, R., JOHNSTON, D. & BELL, M. 2000. Review of possible implications of an airtightness standard for new dwellings in the UK. *Building Services Engineering Research and Technology,* 21, 27-34.

MACKAY, D. J. 2003. *Information theory, inference and learning algorithms*, Cambridge: Cambridge university press.

MCLEOD, R., JAGGS, M., CHEESEMAN, B., TILFORD, A. & MEAD, K. 2014. Passivhaus primer: Airtightness Guide, Airtightness and air pressure testing in accordance with the Passivhaus Standard. BRE.

ODMP 2004. Proposals for amending Part L of the Building Regulations and Implementing the Energy Performance of Buildings Directive: A consultation document. London: ODPM.

ODPM 2006. Approved Document L1A: Conservation of Fuel and Power in New Dwellings. 2006 edition. London: Office of the Deputy Prime Minister,.

PAWITAN, Y. 2001. *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press.

ROUCHIER, S., RABOUILLE, M. & OBERLÉ, P. 2018. Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling. *Building and Environment,* 134, 181-190.

SCOTTISH GOVERNMENT 2010. Building Standards technical handbook 2010: domestic. Livingston: Scottish Government Building Standards Division.

SMITH, S., BAKER, D., MACKENZIE, R., WOOD, J. B., DUNBAVIN, P. & PANTER, D. 2006. The development of robust details for sound insulation in new build attached dwellings. *Journal of Building Appraisal,* 2, 69-85.

SUMMERFIELD, A. J., LOWE, R. J. & ORESZCZYN, T. 2010. Two models for benchmarking UK domestic delivered energy. *Building Research & Information,* 38, 12-24.

TENNANT, G. 2001. *Six Sigma: SPC and TQM in manufacturing and services*, Gower Publishing, Ltd.

VINHA, J., MANELIUS, E., KORPI, M., SALMINEN, K., KURNITSKI, J., KIVISTE, M. & LAUKKARINEN, A. 2015. Airtightness of residential buildings in Finland. *Building and Environment,* 93, 128-140.

WINGFIELD, J., BELL, M., MILES-SHENTON, D., SOUTH, T. & LOWE, R. 2011. Evaluating the impact of an enhanced energy performance standard on load-bearing masonry domestic construction: Understanding the gap between designed and real performance: lessons from Stamford Brook.

~~GRIGG, P. 2004. Assessment of energy efficiency impact of building regulations compliance. Watford: Building Research Establishment.~~

~~HM GOVERNMENT 2016. The Building Regulations 2010: Approved Document L1A: Conservation of Fuel and Power in New Dwellings. 2013 edition with 2016 amendments.~~

~~LEPRINCE, V., MOUJALLED, B. & LITVAK, A. Durability of building airtightness, review and analysis of existing studies. 38th AIVC Conference: Ventilating healthy low-energy buildings, 2017 Nottingham, UK.~~