Analysis of air leakage measurements of US houses

Wanyu R. Chan*, Jeffrey Joh, Max H. Sherman

Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Mailstop 90R3058, Berkeley, CA 94720, USA

ABSTRACT

Building envelope airtightness is important for residential energy use, occupant health and comfort. We analyzed the air leakage measurements of 134,000 single-family detached homes in US, using normalized leakage (NL) as the metric. Weatherization assistance programs (WAPs) and residential energy efficiency programs contributed most of the data. We performed regression analyses to examine the relationship between NL and various house characteristics. Explanatory variables that are correlated with NL include year built, climate zone, floor area, house height, and whether homes participated in WAPs or if they are energy efficiency rated homes. Foundation type and whether ducts are located outside or inside the conditioned space are also found to be useful parameters for predicting NL. We developed a regression model that explains approximately 68% of the observed variability across US homes. Of these variables considered, year built and climate zone are the two that have the largest influence on NL. The regression model can be used to predict air leakage values for individual homes, and distributions for groups of homes, based on their characteristics. Using RECS 2009 data, the regression model predicts 90% of US houses have NL between 0.22 and 1.95, with a median of 0.67.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Residential energy efficiency and weatherization assistance programs (WAPs) have led to many measurements of air leakage being made in the US in recent years. We gathered these data to characterize the air leakage distribution of homes in the US. Uncontrolled airflow through the building envelope has important implications to energy consumption in residences. Most US homes depend on air infiltration as the dominant mean of ventilation, so air leakage also impacts the indoor environmental quality of homes. It is the goal of this regression analysis to identify housing characteristics that can explain the observed variability in air leakage of single-family detached homes. Using the regression results and US housing data, we estimated an air leakage distribution that is representative for the current housing stock.

In 2011, we gathered a large number of air leakage measurements from more than 100,000 US homes. These measurements were added to data that were previously analyzed [1,2] to form the Lawrence Berkeley National Laboratory Residential Diagnostics Database (ResDB). Previous versions of ResDB were dominated by a few data sources. As such, the data were not representative of the US. The vast majority of the data were provided by an income-qualified WAP in Ohio. At that time, the dataset was also dominated

by energy-efficient homes that were built for the extreme weather in Alaska. Furthermore, all of the ResDB data previously analyzed were collected in 2001 and earlier. Therefore, there is a need to update the database to include homes that are built more recently, especially because many residential analyses by other researchers [3–5] has since relied on that dataset dated 2001 as one of the model inputs.

In response to changes in building codes, recent studies have evaluated the energy use and other performance aspects of new US homes [6–8]. These studies suggest a general trend that new homes are being built tighter in some parts of the US. But many factors influence the air leakage of homes. In the presence of considerable house-to-house variability that is inherent in a housing stock, a large dataset is necessary for the regression analysis to evaluate the associations of air leakage with a number of housing characteristics. The approach used in this work largely follows previous regression analyses [1,2]. Recent studies in other countries have also found meaningful associations of air leakage with various housing characteristics: e.g., differences by construction and structural types [9–11], dwelling age and size [12]. In Canada, a study of 100 newly constructed homes that are representative of the new home market of 2008 found attached houses to have higher air leakage than detached houses, using average ACH_{50} as the metric of comparison [9]. Houses with a garage or are multi-story, for example, also tend to higher ACH_{50} on average. However, a similar comparison among 230 new Finnish single-family houses and apartments built in mid-2000 found the

* Corresponding author. Tel.: +1 510 4866570; fax: +1 510 4866658.
E-mail address: wrchan@lbl.gov (W.R. Chan).

reverse [10], where the airtightness of apartments was better than that of the single-family houses. The study identified the joint between the exterior wall and the ceiling as being the most common source of air leakage, and houses with a concrete ceiling are more airtight than with a timber-frame ceiling. Both of these studies only compared the average air leakage values of homes with different characteristics, and did not evaluate the relationship of house characteristics and air leakage using statistical method.

Multivariate linear regression and other statistical techniques were used to analyze the relationship between house characteristics and air leakage among 287 dwellings in UK [11] and 483 single-family dwellings in France [12]. The regression models resulted from these fairly small datasets explained roughly half of the variability in air leakage, where the R^2 equals 0.5 and 0.4 for the homes studied in the two countries, respectively. The UK dwellings [11] include detached, semi-detached, and apartments that were built by three companies in different regions after 2006. Similar to the Finnish study, apartments showed better airtightness than houses. In addition, there are notable differences between the three builders, and the construction types (e.g., precast concrete panel the most airtight, followed by timber frame). The French database [12] contains houses that are more diverse in year built, which allows the relationship with air leakage be captured also in the regression model.

The objective of this work is to characterize the air leakage distribution of homes in the US. A large air leakage dataset was analyzed by regression to identify housing characteristics that are correlated with the air leakage measurements. The housing characteristics considered are descriptive parameters that are available from various US housing surveys. This enables the estimate of air leakage distribution representative of the US housing stock, as well as subgroups of homes based on their characteristics, as presented in this paper. This approach is similar in concept to other predictive methods recently reviewed [13], but different in that the parameters considered are not component based (i.e., summing of the total air leakage area of window and door frames, electrical outlets, etc.), but rather it is largely based on more general descriptions of the whole house (e.g., climate zone, year built, floor foundation type, etc.). Results of this regression can be used to expand the scope of energy use and occupant health and comfort analysis from small-scale studies of a few homes [14] to a housing stock-scale evaluation [15].

1.1. Air leakage measurements

Air leakage is quantified by measuring the airflow through the building envelope, Q (m^3/s), as a function of the pressure across the building envelope, ΔP (Pa). This relationship fits a power law [16], as described in Eq. (1).

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

where C (m^3/s Pa^n) is the flow exponent, and n is the pressure exponent. E779-10 is the measurement standard most commonly used in the US [17]. Typically, airflow is measured using a blower door at $\Delta P = 50$ Pa. This pressure difference is low enough for standard blower door devices to achieve in most houses. At the same time, it is high enough to be reasonably independent of weather influences. For a more detail discussion of the blower door measurement technique that is commonly used to collect air leakage data, see [18].

Air leakage measurements are converted to normalized leakage (NL) for this analysis, as follows:

$$NL = 1000 \left(\frac{ELA_4}{Area}\right) \left(\frac{H}{2.5 \text{ m}}\right)^{0.3}$$

$$ELA_4 = \sqrt{\frac{\rho}{2(4 \text{ Pa})}} Q_{50} \left(\frac{4 \text{ Pa}}{50 \text{ Pa}}\right)^{0.65} \tag{2}$$

where ELA_4 (m^2) is the effective leakage area at 4 Pa, Q_{50} (m^3/s) is the airflow at 50 Pa, $Area$ (m^2) is the floor area, H (m) is the house height, and $\rho = 1.2$ kg/m^3. ELA_4 is a measure of air leakage, which represents the area of an orifice that would result in the same airflow through the building envelope at a pressure difference of 4 Pa. Other commonly used metrics of air leakage include air changes per hour at 50 Pa (ACH$_{50}$), which equals Q_{50} divided by the house volume. The conversion of NL to ACH$_{50}$ can be easily performed using Eq. (2) by first estimating Q_{50}, and then divide Q_{50} by the house volume. Roughly speaking, NL = 0.55 corresponds to ACH$_{50}$ = 10. NL is a useful and convenient metric to describe the air leakage of buildings of different sizes because $Area$ and H often are known parameters or they can be measured quite easily. NL is used in this analysis also for the sake of consistency with earlier work on the air leakage of US homes [1,2,19,20].

2. Data description

ResDB contains air leakage data from 147,000 US homes (Fig. 1), of which 92% are from single-family detached homes that will be analyzed here. The remaining data are mostly from manufactured homes (5% of the data) that participated in WAPs, and also from single-family attached homes and multi-family housing units but in fewer numbers. Because there are potential differences in the air leakage characteristics of these other housing types, this analysis will focus on single-family detached homes only. Approximately two-fifth of the air leakage data were added to ResDB in 2011. Data sources contributed voluntarily to ResDB. Therefore, even though the sample size is large, these self-selected data do not form a representative sample of US homes. In addition, there are also many missing data in ResDB. The handling of these missing data, such as year built, foundation type, and duct location, will be described below. Other detail characteristics, such as frame, wall, and roof materials and construction types, are available in too few of the homes to be considered in the regression analysis. In the US, stick frame structures are the most common, so it is likely that they are also the most represented in ResDB.

Income-qualified WAPs contribute half of the data in ResDB. In addition to the Ohio data that dominated the previous analyses, states with WAP data include Arkansas, California, Iowa, Idaho,
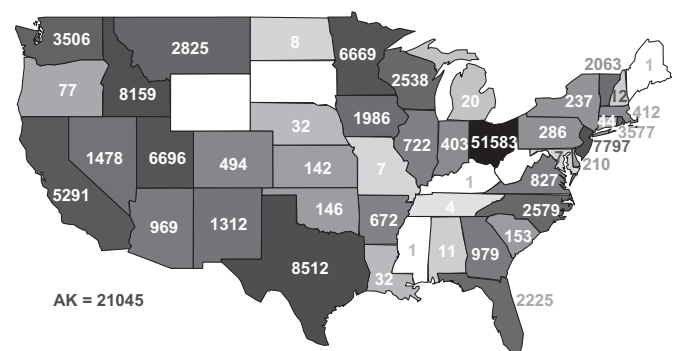

Fig. 1. Number of homes represented in ResDB. Counts include all single-family detached and attached homes, multi-family housing units, and manufactured homes.

Ohio, Minnesota, Montana, Pennsylvania, Utah, Virginia, Washington, and Wisconsin. WAPs require the air leakage of homes be tested twice, once before and once after weatherization, to demonstrate airtightness improvements from air sealing and other measures. Since WAPs are administered by the states, there are many differences in how the weatherization measures were performed and the way data were collected [21]. Some of the air leakage data are provided in the form of a database by state agencies that are responsible for WAP, while others are contributed directly by contractors who performed the weatherization.

Residential energy efficiency programs are another major sources of data. For example, the Home Performance with ENERGY STAR program for existing homes is implemented in over 30 states in the US [22]. Many utility sponsored programs also offer incentives for energy efficiency upgrades. Energy auditors who collected the measurements contributed the majority of the energy efficiency program data. Some energy efficiency programs also provided pre- and post-retrofit air leakage measurements. New Jersey and Minnesota are the two states with the most number of pre- and post-data. There are also a large number of data from residential energy efficiency programs in Vermont, Indiana, California, and Georgia.

Approximately 30% of the recently added data are from homes built after 2006. Many of these new homes were tested for air leakage in order to obtain an energy efficiency rating, or to meet the airtightness guidelines of building codes. These data were contributed mostly by energy auditors who performed the tests, or by a third-party verifier. In addition, there are also a few research programs that collected data on new homes, such as the US Department of Energy's Building America Program. California, North Carolina, Nevada, Texas, and Washington are some of the states with many data from new homes in ResDB.

Energy-efficient homes from Alaska were present in large numbers in the previous version of ResDB. The recently added data also contain some homes that are rated for energy efficiency, but in fewer numbers. There are four data contributors that tested approximately 8000 energy-efficient homes. Florida, North Carolina, and Washington are the leading states with the most number of energy-efficient homes represented in ResDB. Homes are identified as rated for energy efficiency according to the program descriptions from the data contributors. Consequently, there are differences in how the ratings are defined among programs. But minimizing air leakage is a common strategy in energy efficient homes, so it is reasonable to expect some differences between these homes as a group relative to the general housing stock.

2.1. Normalized leakage

Single-family detached homes in ResDB have NL that is log-normal distributed (Fig. 2). Typical values of NL range from 0.16 to 2.23 (5th and 95th percentiles; roughly correspond to 3–40 ACH_{50}). Homes from 43 states are represented. Approximately half of the data are from WAPs, and one-fifth are energy efficiency rated homes. The median floor area is $140\,m^2$ (interquartile 100–$195\,m^2$) and median year built in 1970 (interquartile 1932–1999). This is fairly typical of the US housing stock. The American Housing Survey 2009 reports that the median floor area is $160\,m^2$ and the median year built is 1974 [23].

When computing NL from measurements of Q_{50} using Eq. (2), if H is not provided, we assumed 2.5 m for each story, and an additional 0.5 m for ground level and inter-floor framing. In about 10% of the homes where both the number of story and house height are unknown, we followed the same assumption from previous analyses of ResDB [1,2] that houses $<200\,m^2$ are single-story, and $>200\,m^2$ are two-story. In US, about 80% of single-story detached houses are $<200\,m^2$, but only half of the multi-story detached
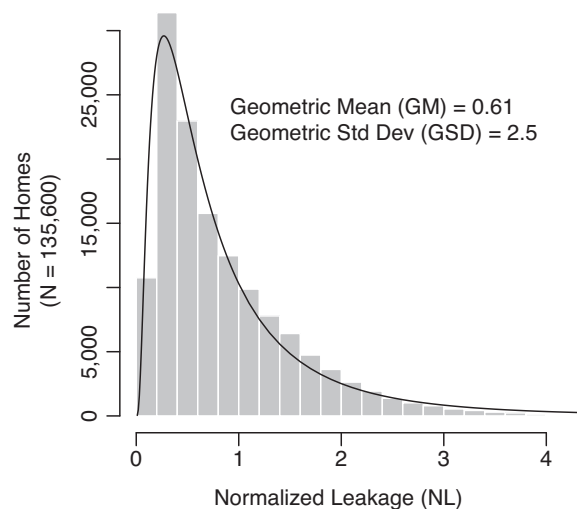

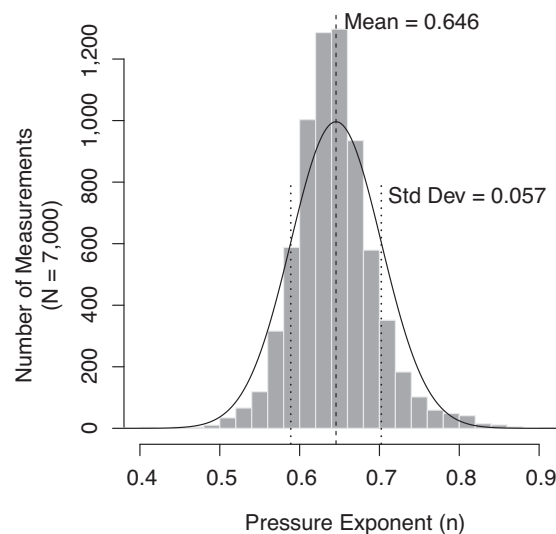**Fig. 2.** Normalized leakage distribution of single-family detached homes in ResDB.

Fig. 3. Distribution of pressure exponent reported from blower door measurements in ResDB.

houses are $<200\,m^2$ [23]. This simple method of using the house size to approximate number of story is reasonable, but it is a source of uncertain in the data analysis.

In most cases, $n = 0.65$ is assumed in computing NL. This commonly used value is based on the assumption that air leakage pathway in residences are dominated by developing flows in cracks, such that $n = 0.67$ (or 2/3) [24]. ResDB contains 7000 measurements of n (Fig. 3), most of these are data added to ResDB in 2011. The distribution of n is roughly normal, with a mean at the expected value of 0.65. The value of n is interesting from a diagnostic perspective because it provides an indication of the relative size of the dominant leaks. If the leakage pathways are small and long, then n increases to approach 1; but if leakage pathways are dominated by specific openings such as a flue, then n decreases to approach 0.5.

3. Method

The multivariate regression considers the relationship between NL and these housing characteristics that are available from ResDB. I's are indicator variables that have values of either 1 or 0, depending if a condition is true or false.

(a) Year built ($\overleftrightarrow{I_{year}}$): before 1960, 60–69, 70–79, 80–89, 90–99, 2000 and after

(b) IECC climate zones[1] ($\overleftrightarrow{I_{cz}}$): twelve combinations of zones 1–8 in four climates: humid (A), dry (B), marine (C), and Alaska (AK)

(c) Homes participated in WAP (I_{LI})

(d) Homes rated for energy efficiency (I_e)

(e) Floor area, *Area* (m^2), and house height, H (m)

(f) Foundation type: slab (I_{slab}), conditioned basement/unvented crawlspace (I_{floor1}), unconditioned basement/vented crawlspace (I_{floor2})

(g) Duct location: conditioned space (I_{cond}), unconditioned attic/basement (I_{duct1}), vented crawlspace (I_{duct2})

Some data in ResDB give the exact year built of the homes tested, while others provided a range. To maximize the number of available data in the regression analysis, year built is considered as a categorical parameter as described above. There are approximately 98,000 measurements of NL where housing parameters (a–e) are known. Year built is missing from the remaining 36,000 homes, and most of the data (>90%) are missing parameters (f) and (g). To handle these missing data, we preformed several multiple regression analyses and combined the results as follows. We first performed the regression between ln(NL) and parameters (a–e) with the 98,000 measurements, as shown in Eq. (3).

$$\ln(\text{NL}) = \beta_y Y + \overleftrightarrow{\beta_{cz}}\overleftrightarrow{I_{cz}} + \beta_{LI}I_{LI} + \beta_e I_e + \beta_{area}Area + \beta_h H \tag{3}$$

where Y is an integer between 5 (oldest) and 0 (newest) corresponding to the six year built categories as described above. Eq. (3) gives $\beta_y = 0.141$ (95% confidence interval = 0.138–0.143), meaning that ln(NL) decreases at a rate of 0.14 per decade. Using β_y and estimates of the other coefficients (β's) from Eq. (3), homes with missing year built are imputed by selecting one of the six year built categories to minimize the difference between NL predicted by Eq. (3) and the measured value.

Next, a regression is performed as follows that considered the entire dataset on 134,000 homes including the imputed data.

$$\ln(\text{NL}) = \overleftrightarrow{\beta_{year}}\overleftrightarrow{I_{year}} + \overleftrightarrow{\beta_{cz}}\overleftrightarrow{I_{cz}} + \beta_{LI}I_{LI} + \beta_e I_e + \beta_{area}Area + \beta_h H \tag{4}$$

where $\overleftrightarrow{I_{year}}$ are a set of six indicator variables that represents the as-reported and imputed year built. Twelve IECC climate zones are modeled: five in humid climate (A 1–2, 3, 4, 5, and 6–7), three in dry climate (B 2–3, 4–5, and 6), two in marine climate (C 3 and 4), and two in Alaska (AK 7 and 8). Some climate zones are combined (e.g., A 1–2) because there is insufficient data in ResDB to model them separately. The way that the twelve climate zones are defined using a set of twelve indicator variables means over parameterization (i.e., the value of one of the variables is defined by the other eleven). To resolve this, we set the coefficient of one of the climate zone to zero. In this case, we selected A 6–7, but any choice would give the same relative results.

In Eq. (4), the coefficient estimates $\overleftrightarrow{\beta_{year}}$ are not confined to follow $\beta_y = 0.14$ used in the imputation, but it is determined by the least-square regression. Fig. 4 compares the predicted NL using the coefficient estimates obtained (i) with and (ii) without the imputed data. The general trend that newer homes tend to have lower NL is clear, regardless if the imputed data are used in the regression or not. The imputation method used has the potential of underestimating the differences between the observed and predicted values. In this case, however, the fit of regression model with ($R^2 = 0.683$) and without ($R^2 = 0.682$) the imputed data was
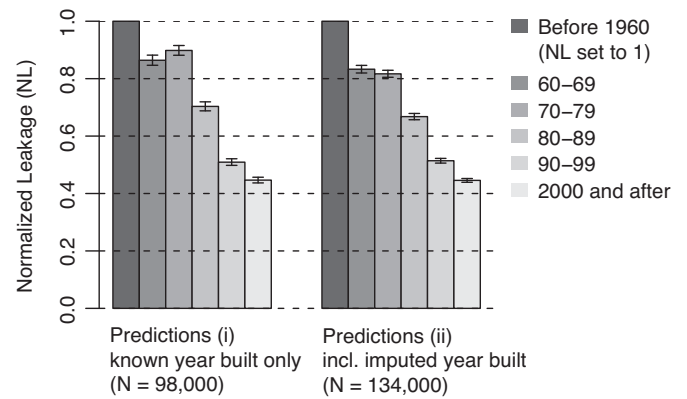
Fig. 4. Predicted normalized leakage as a function of year built from the regression model Eq. (4) considering homes with (i) known year built only, and (ii) includes imputed data.

essentially unchanged. The imputed data have minor effects on the coefficient estimates of the other parameters. Values of β_{area} and β_{IL} are altered only slightly by 6%. Changes to β_h and β_e are larger at −15% and +28%, respectively. But the significance of the imputation is that it allows more data to be included in the regression model. Otherwise, homes in the dry climate zones B 4, 5, and 6, and in the marine climate zones C 3 and 4, would not be adequately represented in the regression.

There are 12,500 houses with known foundation types, and only 526 houses with known duct locations. Because there are relatively few data on these two parameters, we assumed that the coefficient estimates from Eq. (4) apply, and considered separately the effects of foundation type and duct location on the model residuals NL' as follows:

$$\ln(\text{NL}') = \ln(\text{NL}) - [\overleftrightarrow{\beta_{year}}\overleftrightarrow{I_{year}} + \overleftrightarrow{\beta_{cz}}\overleftrightarrow{I_{cz}} + \beta_{LI}I_{LI} + \beta_e I_e$$
$$+ \beta_{area}Area + \beta_h H] \tag{5}$$

Foundation type : $\quad \ln(\text{NL}') = \beta_{slab}I_{slab} + \beta_{floor1}I_{floor1} + \beta_{floor2}I_{floor2}$
$$\tag{6}$$

Duct location : $\quad \ln(\text{NL}') = \beta_{cond}I_{cond} + \beta_{duct1}I_{duct1} + \beta_{duct2}I_{duct2}$
$$\tag{7}$$

Categories of foundation type and duct location are selected based on similarity in the values of ln(NL'), as shown in Fig. 5. After adjusting for the other parameters, homes with slab have the lowest NL, followed by homes with either a conditioned basement or an unvented crawlspace, i.e. $I_{floor1} = 1$. Homes with an unconditioned basement or a vented crawlspace i.e. $I_{floor2} = 1$, tend to have the highest NL. Similar reasoning is used to select the duct location categories, but using another subset of the data where this information is available. Homes with ducts located inside the conditioned space have the lowest NL, followed by homes with ducts located in the unconditioned attic or basement, i.e. $I_{duct1} = 1$, and homes with ducts located in the vented crawlspace, i.e. $I_{duct2} = 1$, have the highest NL.

4. Regression results

The regression model explains 68% of the observed variability. Table 1 shows the coefficient estimates that resulted from the regression analyses. Residuals of the regression model are roughly normal distributed with a mean close to zero (6.2e−17) and a variance of 0.203, if only parameters (a–e) are considered (Eq. (4)). Parameters (f) and (g), modeled using Eqs. (5)–(7),
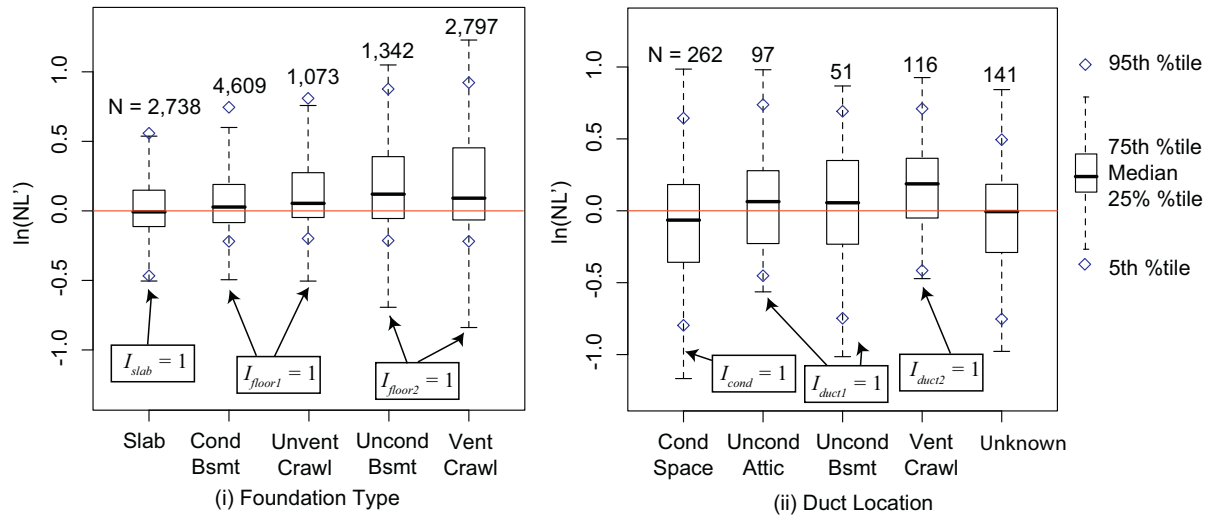
Fig. 5. Model residuals, ln(NL′), computed using Eq. (5) for homes with known (i) foundation type and (ii) duct location (N = house counts). ln(NL′) > 0 means that houses have NL higher than is predicted by Eq. (4) without considering foundation type or duct location.

lower the variance of the residuals slightly to 0.200. Inclusion of these two parameters also modestly altered the residual mean to −0.0094. For the purpose of applying these results to model the air leakage distribution of US homes, it is reasonable to assume a lognormal distribution, where the geometric mean is predicted by the coefficient estimates shown in Table 1, and the geometric standard deviation equals 1.56 as determined by the residual variance of 0.2.

4.1. Year built and climate zones

Year built and climate zones are the two most influential parameters on NL. For example, the difference in predicted NL between the two extreme climate zones in US, A 1–2 (humid and warmest)

and AK 8 (coldest) is a factor of 2.7, based on the coefficient estimates from Table 1.

$$\frac{NL_{A\ 1-2}}{NL_{AK\ 8}} = \exp(\beta_{cz,A\ 1-2} - \beta_{cz,AK\ 8}) = 2.678 \tag{8}$$

Similarly, the difference in NL between homes that are built before 1960 and after 2000 is a factor of 2.2. In humid areas of the US, the regression results suggest a general trend that homes located in the warmer climate zones tend to have higher NL than those in the colder climate zones. This trend also holds for homes in Alaska, i.e. $NL_{AK\ 7} > NL_{AK\ 8}$. However, the reverse is found for homes in the marine climate zones, i.e. $NL_{C\ 3} < NL_{C\ 4}$. Data from climate zones C 3 and 4 are dominated by homes from California and Washington, respectively. Differences between the two states in construction

Table 1
Results of regression models, described in Eqs. (4)–(7), that relate ln(NL) and various housing characteristics.

Explanatory variable	Coefficient estimates	Standard error	Pr(>|t|)	95% confidence interval (C.I.)
(a) Year built $\overleftrightarrow{\beta_{year}}$				
Before 1960	−0.250	0.00705	<2e−16	−0.264; −0.236
1960–69	−0.433	0.00811	<2e−16	−0.449; −0.417
1970–79	−0.452	0.00762	<2e−16	−0.467; −0.437
1980–89	−0.654	0.00836	<2e−16	−0.670; −0.637
1990–99	−0.915	0.00816	<2e−16	−0.931; −0.899
2000 and After	−1.058	0.00748	<2e−16	−1.073; −1.043
(b) Climate zone $\overleftrightarrow{\beta_{cz}}$				
Humid A 1–2	0.473	0.01015	<2e−16	0.453; 0.493
A 3	0.253	0.00653	<2e−16	0.240; 0.266
A 4	0.326	0.00586	<2e−16	0.315; 0.338
A 5	0.112	0.00551	<2e−16	0.101; 0.123
A 6–7	0	–	–	–
Dry B 2–3	−0.038	0.00759	7.57e−07	−0.052; −0.023
B 4–5	−0.009	0.00684	2.00e−01	−0.022; 0.005
B 6	0.019	0.00988	4.91e−03	0.00008; 0.039
Marine C 3	0.048	0.01407	6.02e−04	0.021; 0.076
C 4	0.258	0.01133	<2e−16	0.236; 0.281
Alaska AK 7	0.026	0.00589	1.42e−05	0.014; 0.037
AK 8	−0.512	0.00938	<2e−16	−0.530; −0.439
(c) WAP homes (pre-weatherization): β_{IL}	0.420	0.00428	<2e−16	0.411; 0.428
(d) Energy efficiency rated homes: β_e	−0.384	0.00453	<2e−16	−0.393; −0.375
(e) Floor area: β_{area} (m^{-2})	−0.00208	0.0000179	<2e−16	−0.00211; −0.00204
House height: β_h (m^{-1})	0.064	0.00125	<2e−16	0.061; 0.066
(f) Foundation type – slab: β_{slab}	−0.037	0.00709	1.85e−07	−0.051; −0.023
Conditioned basement/unvented crawlspace: β_{floor1}	0.109	0.00492	<2e−16	0.099; 0.118
Unconditioned basement/vented crawlspace: β_{floor2}	0.180	0.00577	<2e−16	0.169; 0.192
(g) Duct location – conditioned space: β_{slab}	−0.124	0.0255	1.53e−06	−0.174; −0.074
Unconditioned attic/basement: β_{duct1}	0.071	0.0339	3.59e−02	0.005; 0.138
Vented crawlspace: β_{duct2}	0.181	0.0383	2.98e−06	0.106; 0.256

practices and building codes are the likely reasons that can explain this result.

The regression model predicts similar NL for homes in all dry climate zones of US. These homes tend to be quite airtight relative to those in the humid and marine climate zones. The coefficient estimate of climate zone B 4–5 is not statistically significant (see Table 1), meaning that NL of homes in B 4–5 tend to be less than in A 6–7, but the difference is small, and our analysis cannot exclude the possibility that this apparent difference occurs only by chance in the data. We tried grouping homes from the two climate zones B 4–5 and A 6–7 together and performed the regression again, but this did not change the overall model fit or other coefficients significantly. Since the two climate zones are geographically far apart, for completeness all twelve climate zones are kept in the model.

4.2. Weatherization assistance programs (WAPs)

The regression model suggests that homes that participated in WAPs tend to have NL 50% higher per-weatherization than comparable homes. This is computed by the coefficient estimate β_{IL} from Table 1.

$$\frac{NL_{WAP}}{NL_{non-WAP}} = \exp(\beta_{IL} - 1) = 0.522 \tag{9}$$

Eligibility to WAPs is based on household income. In 2009, WAPs used 200% of the poverty line as the eligibility criteria [25], but over the years this had varied between 125% and 150%. It is reasonable to assume that the coefficient estimate β_{IL} applies more broadly to homes that are occupied by low-income households in general, where the construction quality and maintenance is limited by resources to a greater extent. Compared to previous analyses [1,2] where the NL of WAP homes are twice the values of other similar homes, the current result is likely more representative of the US because it includes data from 11 other states instead of solely from Ohio.

4.3. Energy efficient homes

Homes that are rated for energy efficiency tend to have NL 30% lower than comparable homes. Again, the current result is expected to be more representative of the US because it is based on data from many states, and not just from Alaska alone. Energy efficiency rating guidelines for new homes have changed over time. For example, between 1995 and 2006, ENERGY STAR Version 1 was used [26]. Version 2 became effective in 2007. The current Version 3 specifies ACH_{50} to be less than 3–6, depending on the climate zone. Follow roughly this timeline when the different versions of ENERGY STAR were adopted, we introduced additional indicator variables to represent energy efficiency program implemented in these three time periods: pre-1995, 1995–2007, and post-2007. However, we found that this refinement does not improve the model fit compared to using a single variable as in Eq. (4). Further, the coefficient estimates for the three indicator variables, ranging from −0.36 to −0.40, are very similar in magnitude compare to the single-parameter model ($\beta_e = -0.38$, see Table 1). It appears that homes that are rated for energy efficiency continue to be built with a more airtight building envelope than the average housing stock.

4.4. Floor area and house height

Even though NL is already normalized by both Area and H, there is a correlation between Area and H with NL that is roughly the same as estimated from previous analyses [1,2]. Area is negatively correlated with NL, while H is positively correlated with NL. For example, the regression model predicts that homes that are 100 m^2 larger in floor area tend to have NL 20% lower than comparable homes; homes that are two-story instead of one-story (i.e., H to increase by 2.5 m) tend to have NL 17% higher than comparable homes. These results suggest that air leakage scales more strongly than linear with Area^{-1} and H as used in the normalization (Eq. (2)).

4.5. Foundation type

Homes built on a concrete slab are common in warmer climates of US. Nationally, 38% of single-family detached homes have a slab foundation, followed by 34% having a basement (20% heated and 14% unheated), and 28% have a crawlspace [27]. Our regression results suggest that homes with either a conditioned basement or an unvented crawlspace tend to have NL 16% (95% confidence interval: 14–18%) higher than homes on slab. Homes with either an unconditioned basement or a vented crawlspace tend to have NL 24% (95% confidence interval: 22–27%) higher than homes on slab. Other combinations of foundation types are possible, as previous analyses [1,2] combined slab and conditioned basement as having no floor leaks, and the all other foundation types as having floor leaks. Here, these three categories of foundation types are selected to give the best-fit model with the highest R^2 value using Eq. (6).

4.6. Duct location

The data on whether a home has duct or not was found not to be a useful indicator variable in previous analyses [1,2]. This is likely because knowing the presence or absence of ducts alone is insufficient to explain meaningful relationships with air leakage. There are other characteristics of the duct systems that matter, such as if the ducts are located within or outside of the conditioned space. The current regression suggests homes with ducts inside the conditioned space tend to have NL 18% lower (95% confidence interval: 11–24%) compared to ducts located in unconditioned attic or basement. Homes with ducts in vented crawlspace tend to have NL 12% higher (95% confidence interval: 1–23%) than houses with ducts located in unconditioned attic or basement. The uncertain estimates are large because the regression is based on only 526 homes. Half of these homes are from just two states, South Carolina and Minnesota, and the remaining half are mostly from these five states: Utah, Indiana, Illinois, Massachusetts, and Virginia. More data and better spatial coverage would improve the confidence of this analysis.

4.7. Model prediction and measurement comparison

The regression analysis suggests that much of the variability observed in NL is associated with (a) year built, (b) climate zone, and whether the houses are participants in (c) WAPs or are (d) energy efficiency rated homes. The remaining factors, namely (e) floor area and house height, (f) foundation type, and (g) duct location, each explain minor differences in NL in the 10–20% range. Thus in comparison, their importance is secondary for predicting NL. Fig. 6 shows the observed and predicted geometric mean of NL for homes separated by the four characteristics (a–d) that are the most important for predicting air leakage of US homes. Separate comparisons are plotted for homes in these three types: (i) homes that are neither WAP nor energy efficiency rated (37% of the data), (ii) WAP homes (39%), and (iii) energy efficiency rated homes (14%). In each of the three plots, the observed and predicted geometric mean of NL are calculated for 72 possible groups of homes from the combinations of six year built categories and twelve climate zones. Only groups with at least ten homes or more are included in Fig. 6.

It is clear that the regression model can capture the differences between the three types of homes (note the difference in scale of the three plots), where the NL predictions for (ii) WAP homes are the
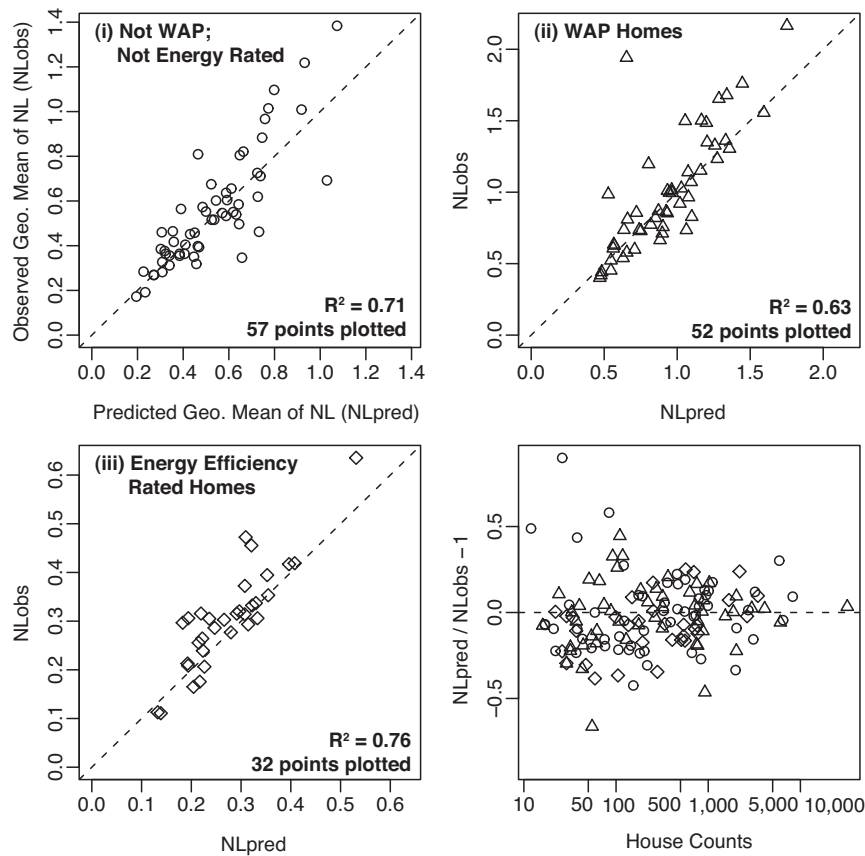
Fig. 6. Comparison of the observed geometric mean (GM) of normalized leakage (NLobs) and the corresponding values predicted by the regression model (NLpred). Each point represents the GM of a group of homes from a given year-built and climate zone. The bottom-right plot shows the difference between NLpred and NLobs for the three types of homes as a function of the number of houses represented by each data point.

highest, and lowest for (iii) energy-efficient homes. Within each of the three types of homes, the regression model also describes the differences as a result of year built and climate zone reasonably well, where the R^2 is roughly 0.7. Overall, there is a slight tendency for the regression model to underpredict the geometric mean of NL, as shown in the bottom-right of Fig. 6. The agreement between observation and prediction is better when more homes are represented in the data. About 90% of observed and predicted geometric mean of NL agrees within ±30% of each other. This shows that using the housing characteristics that are available for this analysis, the regression model captures their influences on the air leakage of US homes reasonably well.

5. Estimates of US air leakage distribution

The NL distribution of US houses can be estimated by applying the resulted regression model using housing data that are representative of the housing stock.

$$\ln(NL) = \beta_{area}Area + \beta_h H + \overset{\leftrightarrow}{\beta_{year}}\overset{\leftrightarrow}{I_{year}} + \beta_{LI}I_{LI} + \beta_e I_e$$

$$+ \overset{\leftrightarrow}{\beta_{cz}}\overset{\leftrightarrow}{I_{cz}} + \beta_{slab}I_{slab} + \beta_{floor1}I_{floor1} + \beta_{floor2}I_{floor2}$$

$$+ \beta_{cond}I_{cond} + \beta_{duct1}I_{duct1} + \beta_{duct2}I_{duct2} + e \quad (10)$$

where e is the residual term that is roughly normal distributed $N(\mu = 0, \sigma^2 = 0.2)$.

The 2009 RECS microdata [27] is an example of US housing data that can be used for this analysis. RECS provides detailed survey data from 7771 single-family detached homes, representing 71.6 million of this type of housing units in the US. This information

is collected by the US Energy Information Administration to estimate the energy costs and usage for heating, cooling, appliances, and other end uses in homes. We extracted the relevant housing characteristic data from each of the single-family detached homes surveyed, and applied the regression model to predict NL. Predictions from the individual homes are aggregated using weights from RECS to give a NL distribution that is representative for the US housing stock. We also present the distributions for homes in selected US states to illustrate the differences in NL estimates as a comparison.

Table 2 provides the names and brief descriptions of the RECS variables used for estimating NL. Many of the housing characteristics from RECS are readily available for this estimation after straightforward adjustments, as described below. In cases where additional information is needed, they are provided in Supplementary Data.

The year built ranges reported in RECS correspond directly to the regression variable $\overset{\leftrightarrow}{I_{year}}$. Climate zones are determined by the states and climate regions given in RECS (see Table S1 in Supplementary Data). I_{IL} is set to 1 if RECS indicates that household income is at or below 150% of poverty line. Energy efficiency rated homes are not modeled because even though programs like ENERGY STAR are increasing in popularity, their number remains few as a fraction of the whole housing stock [28][2]. At the time of this analysis, RECS 2009 has not yet released data on floor area of homes, so the

---

[2] As of November 2009, the total number of ENERGY STAR certified homes built is one million, which is less than 1% of the 113.6 million housing units in US. However, its market share is increasing among new homes, and has reached 25% by September 2011.

Table 2
Use of RECS 2009 data to estimate NL distribution using the regression model.

Regression variable		RECS 2009 variable and descriptions
(a)	Year built $\overset{\leftrightarrow}{I_{year}}$	YEARMADERANGE: 8 year ranges when housing unit was built (before 1950, 50–59, 60–69, 70–79, 80–89, 90–99, 2000–04, 04–09)
(b)	Climate zone $\overset{\leftrightarrow}{I_{cz}}$	REPORTABLE_DOMAIN: 27 reportable states and groups of states CLIMATE_REGION_PUB: Building America climate region (1–5)
(c)	WAP homes I_{LI}	POVERTY150: household income at or below 150% of poverty line (yes/no)
(d)	Energy efficiency rated homes I_e	Assumed negligible
(e)	Floor area $Area$ (m²)	TOTROOMS: total number (1–25) of rooms in the housing unit
	House height H (m)	STORIES: number of stories in a single-family home (1, 2, 3, 4+, split-level, others, n/a)
(f)	Foundation type: $I_{slab}, I_{floor1}, I_{floor2}$	CONCRETE: housing unit over a concrete slab (yes, no, n/a); CELLAR: basement in housing unit; BASEHEAT: heating used in basement; CRAWL: housing unit over a crawlspace
(g)	Duct location: $I_{cond}, I_{duct1}, I_{duct2}$	Weighted averages computed based on REPORTABLE_DOMAIN

regression variable $Area$ is approximated from the total number of rooms (see Tables S2–S4). House height is approximated from the number of stories (excluding basement and attic if present): $H = 3$ m for 1-story, 5.5 m for 2-story or split-level, 8 m for 3-story, and 10 m for 4+ stories. RECS provides data on the presence of concrete slab ($I_{slab} = 1$ if yes), basement, and crawlspace. The regression variable I_{floor1} is set to 1 if a basement is heated, and $I_{floor2} = 1$ if not. All houses with a crawlspace are assumed vented (i.e., $I_{floor2} = 1$), which is more common among US homes than unvented crawlspaces.

RECS only reports if ducts are used for space heating and cooling, but it does not report duct locations. The Home Energy Saver (HES) [29] is a tool developed at Lawrence Berkeley National Laboratory to provide "asset-based" analysis for single-family homes. This tool has been widely used in home inspections and energy audits. We obtained data from 1427 inspection records and 44,584 energy audit records to determine the state-by-state prevalence of duct locations. Inspection records entered by certified professionals are used as the primary source of data. Energy audit records, many are entered by homeowners, are used only if there is insufficient home inspection records (<30 homes). From these data records (see Table S5), there are clear geographical differences in the prevalence of duct locations. Homes with ducts inside the conditioned space ($I_{cond} = 1$) are common in the colder areas, such as Michigan and Wisconsin. Ducts are more commonly located in the attic or basement ($I_{duct1} = 1$) among the Northeast states (e.g., Massachusetts and New York), and also in Arizona, California, Florida, and Texas. Homes with ducts located in the crawlspace are common only in a few states, e.g., Tennessee and Washington. Using this HES dataset, the effect of duct locations on NL is estimated by the sum of the regression coefficients (β_{cond}, β_{duct1}, and β_{duct2}) weighted by the percentages of home having ducts in each of the three locations, as shown in Table S5.

Estimates of the US air leakage distribution are shown in Fig. 7 nationwide, and for selected states as examples. The predicted
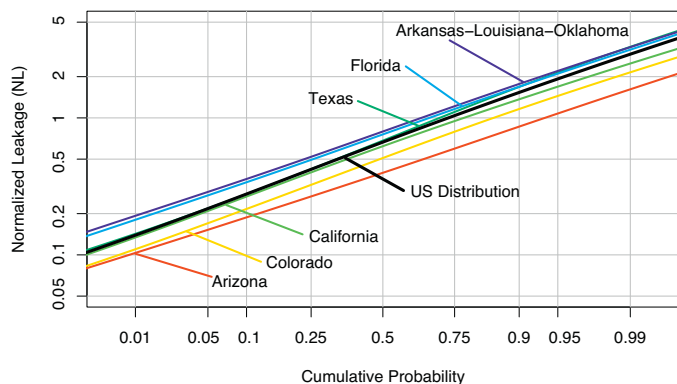
median NL is 0.67, which corresponds roughly to 12 ACH$_{50}$. The regression model, together with housing characteristics based primarily on RECS 2009, predicts that most US homes have NL between 0.22 and 1.95 (5th to 95th percentiles, roughly corresponding to ACH$_{50}$ between 4 and 35).

There are some between-state differences that can be explained by the climate zones that the modeled homes are located in. For example, homes in Florida (climate zone A 1–2) are predicted to have NL higher than homes in Arizona (B 2–3 and 4–5) because $\beta_{cz,A\,1–2}$ is greater than the coefficient estimates in other climate zones. Moreover, homes in Arizona also tend to be newer (median year built = 1990, based on RECS 2009) than in Florida (median year built = 1980). As a result, NL predictions for homes in Arizona are among the lowest in US from the regression model. On the other hand, homes in Arkansas–Louisiana–Oklahoma are relatively older (median year built = 1975). In addition, there is a large fraction (26%) of housing units that would be eligible for WAPs in those three states, as determined based on the household income criteria from RECS 2009. In comparison, the eligibility fraction in the other five states shown in Fig. 7 ranges between 9% (Colorado) and 19% (California). As a result, the predicted NL distribution for Arkansas–Louisiana–Oklahoma is higher than the US average.

6. Discussion

We made a number of assumptions in analyzing air leakage measurements collected from a wide range of data sources, and in our evaluation of the relationship between NL and housing characteristics. These include the method used to impute year built, and also performing the regression analysis step-wise then combining the coefficient estimates to model the effects of foundation type and duct location. As a follow-up to this analysis, the uncertainty of our methodology can be evaluated by using other statistical techniques to handle missing data, and see if there are significant changes to the regression results. There are also air leakage measurement errors [30] and other errors in data reporting that are not considered in our analysis, that may have impacted the reliability of the regression model. Because there are very few data on duct location, our estimate of its relationship with NL is likely the most uncertain. The coefficient estimates of the effect of foundation type, and in some climate zones where the number of homes are relatively fewer in numbers (e.g., A 1–2, B 6, C 3, C 4, and AK 8 all have home counts between 1000 and 3000), are also relatively uncertain.

Overall, the regression suggests coherent relationships between air leakage and housing characteristics. Homes that have higher NL tend to be older, occupied by low-income household, smaller in floor area, multistory, and with likely air leakage through its foundation and ducts. Conversely, homes that are newer, rated for energy efficiency, larger in floor area, single-story, with minimal leakage through its foundation and ducts located inside the conditioned space have lower NL. These results show that many factors

Fig. 7. Estimated NL distributions for single-family detached homes in the US nationwide and in selected states.

can influence building envelope airtightness, including the construction and design of the homes, and possibly how well the homes are maintained. The two factors that have the largest impact on air leakage, year built and climate zones, further point out that airtightness may be impacted by weathering of the building envelope leading to settling and leaks to develop over time. The longevity of the different components of a home, such as insulation and air sealing, may also be a contributing factor that explains the relationship between NL and year built.

The linear combination of the explanatory variables considered in the regression model explains a significant portion (68%) of the variability observed in the air leakage measurements. Analysis of variance suggests that it is reasonable to treat the variables as independent of one another. All the variables considered are necessary for modeling air leakage, leaving out one or more would lead to a worse regression fit. There are additional air leakage analyses that can be performed using data from ResDB, such as more detailed analyses of the potential differences within states that have many data, and the potential seasonal effect on measurement bias. This analysis made used of all the available housing characteristics that are commonly available in ResDB. But, there are subsets of the data that can be used to evaluate the improvements in airtightness from retrofit, for example, which will be useful for estimating energy savings by WAPs and various residential energy efficiency programs.

7. Conclusions

Building envelope air leakage is a key factor in determining air infiltration, which provides most of the ventilation in existing US dwellings. Drafty homes are uncomfortable to live in, and use more energy to heat and cool. On the other hand, homes that are built with a very tight envelope may require mechanical ventilation to maintain good indoor air quality. To characterize the US housing stock, we analyzed the air leakage data of 134,000 single-family detached homes. There are substantial differences in the air leakage of homes that can be explained by these factors: year built, climate zone, whether homes participated in WAPs or if they are energy efficiency rated, floor area, house height, foundation type, and duct location. The regression model from this analysis can be used to estimate the US distribution, such as by using housing data from RECS as performed here. Predicted NL varies by an order of magnitude among 90% of US homes. The variability is significant, and it is important to capture in residential energy analyses and health exposure assessments.

We discussed and provided plausible explanations for the relationships between air leakage and a number of housing characteristics. While the relationships examined here appear rational, our interpretations on why certain factors tend to be correlated with higher or lower NL are speculative. The regression method is appropriate for identifying variables that are useful to consider when estimating NL, but other studies are better suited to demonstrate cause and effect, such as by comparing the air leakage of a selective group of homes with just one key difference but are otherwise similar in characteristics.

This work only considered the air leakage of single-family detached homes in US. ResDB also contains some data on manufacture homes, and to lesser extent single-family attached and multifamily homes as well. Analysis of these data will provide valuable information on these other types in US homes, where very few data exist. Even though these other house types remain in the minorities for US as a whole, they are important in urban areas. In terms of energy use and occupant health and comfort, there are issues that are of particular concerns in these other home types, such as the air leakage to outside versus to adjacent units. Analysis

of these other home types will likely require the collection of additional data with more detailed data fields to support the work.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.enbuild.2013.07.047.

References

[1] W.R. Chan, W.W. Nazaroff, P.N. Price, M.D. Sohn, A.J. Gadgil, Analyzing a database of residential air leakage in the United States, Atmospheric Environment 39 (19) (2005) 3445–3455.
[2] J. McWilliams, M. Jung, Development of a Mathematical Air-leakage Model from Measured Data, in: LBNL Report 59041, Lawrence Berkeley National Laboratory, Berkeley, CA, 2006.
[3] W.R. Chan, W.W. Nazaroff, P.N. Price, A.J. Gadgil, Effectiveness of urban shelter-in-place—II: Residential districts, Atmospheric Environment 41 (33) (2007) 7082–7095.
[4] A. Persily, A. Musser, S.J. Emmerich, Modeled infiltration rate distributions for U.S. housing, Indoor Air 20 (6) (2010) 473–485.
[5] M.S. Breen, M. Breen, R.W. Williams, B.D. Schultz, Predicting residential air exchange rates from questionnaires and meteorology: Model evaluation in Central North Carolina, Environmental Science and Technology 44 (24) (2010) 9349–9356.
[6] F.J. Offermann, Ventilation and Indoor Air Quality in New Homes, in: Collaborative Report, California Air Resources Board and California Energy Commission, PIER Energy-Related Environmental Research Program, 2009.
[7] J. Proctor, R. Chitwood, B.A. Wilcox, Efficiency Characteristics, Opportunities for New California Homes, Proctor Engineering Group, Ltd., Chitwood Energy Management, Inc. Bruce A. Wilcox, California Energy Commission, 2011.
[8] B.D. Nelson, Successful implementation of air tightness requirements for residential buildings, in: Best2 Conference—A New Design Paradigm for Energy Efficient Buildings, Atlanta, Georgia, 2012, p. 12.
[9] J. Harris, Air Leakage in Ontario Housing, DSG Home Inspections Inc./Aubrey LeBlanc Consulting Inc, 2009, pp. 18.
[10] M. Korpi, J. Vinha, J. Kumitski, Airtightness of single-family houses and apartments, in: 8th Nordic Symposium on Building Physics Symposium, Copenhagen, 2008.
[11] W. Pan, Relationships between air-tightness and its influencing factors of post-2006 new-build dwellings in the UK, Building and Environment 45 (11) (2010) 2387–2399.
[12] M.I. Montoya, E. Pastor, F.R. Carrie, G. Guyot, E. Planas, Air leakage in Catalan dwellings: developing an airtightness model and leakage airflow predictions, Building and Environment 45 (6) (2010) 1458–1469.
[13] T.-O. Relander, S. Holøs, J.V. Thue, Airtightness estimation, A state of the art review and an en route upper limit evaluation principle to increase the chances that wood-frame houses with a vapour- and wind-barrier comply with the airtightness requirements, Energy and Buildings 54 (2012) 444–452.
[14] S. Nabinger, A. Persily, Impacts of airtightening retrofits on ventilation rates and energy consumption in a manufactured home, Energy and Buildings 43 (11) (2011) 3059–3067.
[15] H.R.R. Santos, V.M.S. Leal, Energy vs. ventilation rate in buildings: a comprehensive scenario-based assessment in the European context, Energy and Buildings 54 (2012) 111–121.
[16] I.S. Walker, M.H. Sherman, D.J. Wilson, A comparison of the power law to quadratic formulations for air infiltration calculations, Energy and Buildings 27 (3) (1998) 293–299.
[17] ASTM, E779-10 Standard Test Method for Determining Air Leakage Rate by Fan Pressurization, 2010.
[18] M. Sherman, The use of blower-door data, Indoor Air 5 (3) (1995) 215–224.
[19] M.H. Sherman, D.J. Dickerhoff, Airtightness of US dwellings, ASHARE Transactions 104 (2) (1998) 1359–1367.
[20] M.H. Sherman, N.E. Matson, Air tightness of new houses in the U.S, in: 22nd Air Infiltration and Ventilation Centre Conference, Bath, UK, 2001.
[21] J.F. Eisenberg, Weatherizaiton Assistance Program Technical Memorandum Background Data and Statistics, Oak Ridge National Laboratory, Oak Ridge, TN, 2010.

[22] USEPA, States with Home Performance with ENERGY STAR program, http://www.energystar.gov/index.cfm?fuseaction=hpwes_profiles.showsplash (accessed 03.10.12).

[23] AHS, U.S. Census Bureau Current Housing Reports, American Housing Survey for the United States: 2009, in: H150/09, U.S. Department of Housing and Urban Development, Washington, DC, 2011.

[24] M. Orme, M.W. Liddament, A. Wilson, An Analysis and Data Summary of the AIVC's Numerical Database, AIVC Technical Note, Air Infiltration and Ventilation Centre, 1994.

[25] T. Russell, 2010 Poverty Income Guidelines and Definition of Income, Department of Energy, Washington, DC, 2010, pp. 4, http://www1.eere.energy.gov/wip/pdfs/wpn_10-18.pdf (accessed 11.10.12).

[26] USEPA History of the ENERGY STAR Guidelines for New Homes, http://www.energystar.gov/index.cfm?c=new_homes.nh_history (accessed 03.10.12).

[27] EIA, Residential Energy Consumption Survey Public Use Microdata File, U.S. Administration, Washington, DC, 2009, pp. 2011.

[28] USEPA ENERGY STAR Major Milestones, http://www.energystar.gov/index.cfm?c=about.ab_milestones (accessed 15.10.12).

[29] HES Home Energy Saver data from home inspection and energy audit records, http://homeenergysaver.lbl.gov/consumer/ (accessed 05.26.12).

[30] M.H. Sherman, L. Palmiter, Uncertainty in fan pressurization measurements, in: M.P. Modera, A. Persily (Eds.), Airflow Performance of Envelopes, Components and Systems, American Society for Testing and Materials, Philadelphia, 1994, pp. 262–283.