Lawrence Berkeley National Laboratory

Title
Analysis of U.S. residential air leakage database

Permalink
https://escholarship.org/uc/item/6pk6r4gs

Authors
Chan, Wanyu R.
Price, Phillip N.
Sohn, Michael D.
et al.

Publication Date
2003-07-01

Peer reviewed

Analysis of U.S. Residential Air Leakage Database

Wanyu R. Chan*, Phillip N. Price, Michael D. Sohn, Ashok J. Gadgil
Indoor Environment Department
Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, California 94720-8132

* Corresponding author (Mail Stop: 90R3058;
Tel: 510-495-2459; Fax: 510-486-6658; E-mail: wchan@lbl.gov)

July 2003

LBNL Report Number 53367

ABSTRACT

The air leakage of a building envelope can be determined from fan pressurization measurements with a blower door. More than 70,000 air leakage measurements have been compiled into a database. In addition to air leakage, the database includes other important characteristics of the dwellings tested, such as floor area, year built, and location. There are also data for some houses on the presence of heating ducts, and floor/basement construction type.

The purpose of this work is to identify house characteristics that can be used to predict air leakage. We found that the distribution of leakage normalized with floor area of the house is roughly lognormal. Year built and floor area are the two most significant factors to consider when predicting air leakage: older and smaller houses tend to have higher normalized leakage areas compared to newer and larger ones. Results from multiple linear regression of normalized leakage with respect to these two factors are presented for three types of houses: low-income, energy-efficient, and conventional. We demonstrate a method of using the regression model in conjunction with housing characteristics published by the US Census Bureau to derive a distribution that describes the air leakage of the single-family detached housing stock. Comparison of our estimates with published datasets of air exchange rates suggests that the regression model generates accurate estimates of air leakage distribution.

Keywords: Air leakage, Blower door, Fan pressurization measurements, Infiltration

1 INTRODUCTION

Air infiltration in residences is one of the primary factors affecting indoor air quality and energy consumption. For example, high infiltration rates can decrease harmful exposures to pollutants of indoor origin, or they can increase exposures to pollutants that originate outdoors. High infiltration rates can also cause high energy consumption when heating or cooling system is needed to condition the infiltrating air.

Air infiltration data are costly to obtain, and few measurements have been made. The existing databases of air infiltration often contain information on only a few house types or for a few regions in the United States. Larger datasets are available on air exchange rates (e.g. Murray and Burmaster [1995]) but these do not provide full information on leakage characteristics or the influence of building factors.

In this report, we analyzed most of the existing air leakage databases, which amounted to some 70,000 data points (Sherman and Matson [2002]), to investigate air leakage for the US housing stock. We relate leakage distributions of houses to the age of the building, and other building characteristics. We also identify which information was most helpful in predicting air leakage and what could significantly improve the quality of the predictions.

Sherman and Matson [2002] analyzed the air leakage data for newly constructed buildings contained in the databases. They found that newly constructed houses are significantly tighter than what is believed to be typical for the general US housing stock. We will revisit this conclusion using data not only from newly constructed houses, but all the validated data points.

In the sections that follow, we begin by briefly defining air leakage and explaining the experiments employed to measure it. This discussion is followed by an exploratory analysis and summary of the validated data.

2   BACKGROUND

Air leakage in buildings is typically measured using the ASTM Standard E779 "Standard Test Method for Determining Air Leakage by Fan Pressurization," which describes a technique commonly referred to as a "blower door" field test. The method is applicable for buildings with small temperature differences between the inside and outside of the building. The method also assumes that during the test the pressure on the building exerted by wind is minimal. We review the major aspects of the method here, but the reader is referred to Sherman [1995] for a thorough review and history.

Blower door tests measure the amount of air flow needed to pressurize a building to various indoor-outdoor pressure differences. Leakier buildings will require higher airflow rates to pressurize the building, whereas tighter buildings will require lower flow rates. Using several experiments, Sherman [1980] related airflow rate Q_f to indoor-outdoor pressure difference P_f, using a two-parameter power law relationship[1]:

$$Q_f = \kappa P_f^n \tag{1}$$

Eq. (1) reduces to the so-called orifice equation (Eq. (2)), a one-parameter model, if leakage is limited by the flow resistances at the entrance and the exit of the cracks (Sherman [1980]). When applied to buildings, the proportionality constant in Eq. (2) is called the effective leakage area (ELA).

$$Q_f = ELA \cdot \sqrt{\frac{2P_f}{\rho}} \tag{2}$$

Data from blower-door experiments allows one to determine κ and n in Eq. (1). If $n = 0.5$, then equating Eq. (1) and Eq. (2) shows that $ELA = \kappa \cdot \sqrt{\frac{\rho}{2}}$. In general, however, n is observed to vary in the housing stock, typically between 0.5 and 0.75. In this case, the

---

[1]See *Appendix A List of Symbols* for variables used.

ELA must be defined at a specific pressure difference, P_r. By equating Eq. (1) and Eq. (2) at pressure P_r, we have:

$$Q_f = \kappa \cdot P_r^n = ELA\sqrt{\frac{2P_r}{\rho}}$$

$$\kappa = \frac{ELA\sqrt{\frac{2P_r}{\rho}}}{P_r^n}$$

(3)

Since κ is constant, this expression can be substituted into Eq. (1) to yield a general relationship between infiltration and effective leakage area:

$$Q_f = ELA \cdot \sqrt{\frac{2P_r}{\rho}} \cdot \left(\frac{P_f}{P_r}\right)^n$$

(4)

Commonly, the reference pressure P_r is set to be 4 Pa. Leaky single-family residences have ELA of approximately 0.3 m^2 and tight residential buildings have ELA of about 0.04 m^2. Infiltration models use the ELA to predict the air leakage as a function of wind speed and internal-external temperature difference. For example, the LBL infiltration model (Eq. (5)), describes airflow as follows:

$$Q_f = ELA \cdot s$$

$$s = \sqrt{f_s^2 \cdot \Delta T + f_w^2 \cdot v^2}$$

(5)

In the discussion to follow, the effective leakage area is normalized with the building floor area and a correction factor for the building height (Eq. (6)). We have found that this helps to describe the relative leakage for a wide range of building sizes. We chose to use normalized leakage (NL) in this study so as to be consistent with earlier analyses of the leakage database. Normalized leakage is also the measure used for verification when the

data were collected and compiled.

$$NL = 1000 \cdot \frac{ELA}{A_f} \cdot \left(\frac{H}{2.5m}\right)^{0.3} \tag{6}$$

Whenever ELA and NL are mentioned in the rest of the paper, the two quantities are specified at the reference pressure of 4 Pa.

3 Exploratory Analysis of Dataset

3.1 Data Collection

Sherman and Matson [2002] have compiled some 70,000 entries in their air leakage database. We expect the database to continue to expand since the data collection efforts remain active. The three largest contributors to the database are the Ohio Weatherization Program (77% or 51,300 measurements), an energy-efficiency program in Alaska, AKWarm (11% or 7,200 measurements), and the Wisconsin Energy Conservation Corporation (3% or 2,200 measurements). Thirty-one other organizations[2] contributed the remaining 5,700 measurements covering 30 states.

3.2 Data Processing

We performed several data validation, verification, and relevance tests on the database. One requirement for all data points was that the normalized leakage could be calculated. Over 90% of the measurements reported the amount of air flow, Q_f, needed to pressurize the house to 50 Pa. Referring to Eq. (4), ELA was calculated by assigning the reference pressure to 4 Pa and assuming the flow exponent n to be 0.65. When only the air exchange rate (ACH [h^{-1}]) of the house at 50 Pa was reported, we multiplied ACH by house volume [m^3] to obtain Q_f in units [$\frac{m^3}{h}$].

---

[2]See *Appendix B* for full list of contributors.

Normalizing from ELA to NL requires the knowledge of the floor area and the height of each house. Almost all data points report the floor area of the house. However, building height is often not reported, except those from the AKWarm (Alaska) dataset. For all other data, we assumed a building height according to the floor area of the house. We assumed houses with floor areas less than 92 m² (about 1000 ft²) were single storey houses. We assumed houses with larger floor areas were 1.5 storey houses. While it might be the case that some of the 1.5-storey houses are actually two storeys since two-storey homes are more common than split-level or bi-level houses, this uncertainty does not constitute a large source of error because NL only varies in proportion to $H^{0.3}$ as shown in Eq. (6). From the number of storeys, building height is approximated by assuming each storey to be about 2.5 m and adding 0.5 m to the total for roof space.

In addition to air leakage and house size, the remaining information in the database describes the house age, location and construction type. Among these, data on the presence of a thermal duct system and the floor/basement construction type were the most scarce.

Data processing yielded about 66,500 data points that we considered valid or acceptable for further data analysis.

3.3   Data Summary

3.3.1   Spatial Coverage

Most of the measurements are from houses located in Ohio, Alaska and Wisconsin (Fig. 1). The most sampled region is therefore the Midwest. Arizona, California, and Washington together make up the second most sampled region (West). This is followed by the South region, which consists of data from Florida, North Carolina, Arkansas, and Oklahoma. Data in the North region are mostly from houses measured in Vermont and Rhode Island.

In terms of the nine divisions in the US Census classification, North East Central is the most substantially represented region because it includes both Ohio and Wisconsin (Table
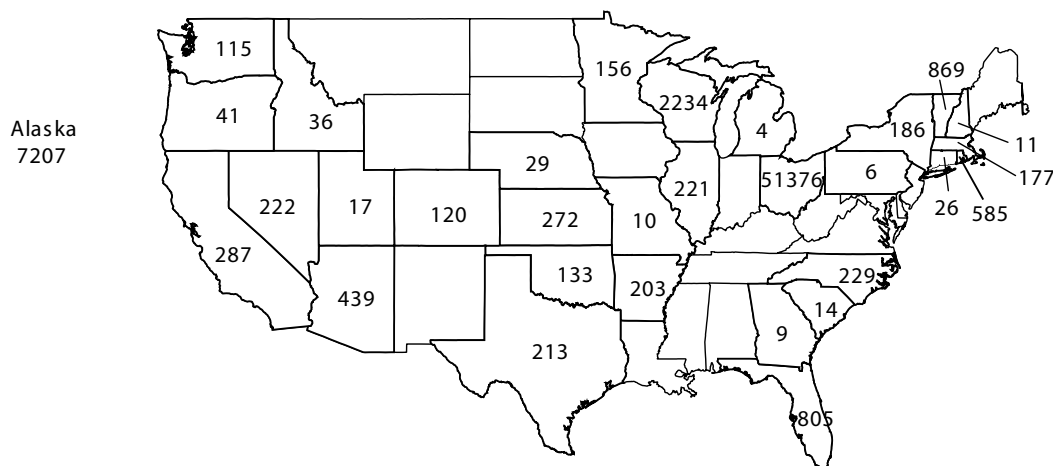
Figure 1: Number of valid blower-door measurements from each state.

1). There are a few hundred measurements from each of the remaining divisions except for the South East Central region, where no data are available.

Over 94% of the data points include zip codes. Some include just city names. Sherman and Matson [1997] estimated the average normalized leakage for each county in the US using a subset of the dataset collected earlier. Instead of reporting normalized leakage directly as a function of location, we will use commonly available building characteristics (e.g. year structure built, floor area) to explore what factors determine normalized leakage.

3.3.2   House Size

The US Census Bureau conducts the American Housing Survey (AHS) to obtain up-to-date housing statistics for the Department of Housing and Urban Development. The 1999 national survey is a sample resulting from 58,400 interviews of homeowners. The raw numbers are reported as scaled estimates according to the total number of homes reported by the 1990 Census Decennial Survey for a given region. We compared some of the housing characteristics in the database with those reported by the 1999 AHS to test how well the air leakage database represents the US housing stock.

Table 1: Number of air leakage measurements tabulated by state, US Census division and region. Total number of data points sums to 66,536. Excluded are apartments and invalid entries. Low-income (all from Ohio) or energy-efficiency program participants are listed separately from conventional houses.

|	State	Division	Region	No. in State	No. in Division	No. in Region	Convent- ional	Energy Efficient
1	OH	NE Central	MW	51376	54302	55688		
2	WI			2234			2121	113
3	IL			221				221
4	MI			4			4	
5	KS	NW Central		272	467			272
6	MN			156			49	107
7	NE			29			29	
8	MO			10			10	
9	AK	Pacific	W	7207	7650	8484	199	6408
10	CA			287			251	36
11	WA			115			115	
12	OR			41			41	
13	AZ	Mountain		439	834		101	338
14	NV			222			24	198
15	CO			120			41	79
16	ID			36			31	5
17	UT			17			17	
18	VT	New England	NE	869	1668	1860	69	800
19	RI			585			564	21
20	MA			177			13	164
21	CT			26			19	7
22	NH			11				11
23	NY	Mid Atlantic		186	192		150	36
24	PA			6				6
25	FL	S Atlantic	S	805	1260	1403	339	466
26	NC			229			172	57
27	SC			14			6	8
28	GA			9			1	8
29	TX	SW Central		213	549		111	102
30	AR			203			203	0
31	OK			133			108	25
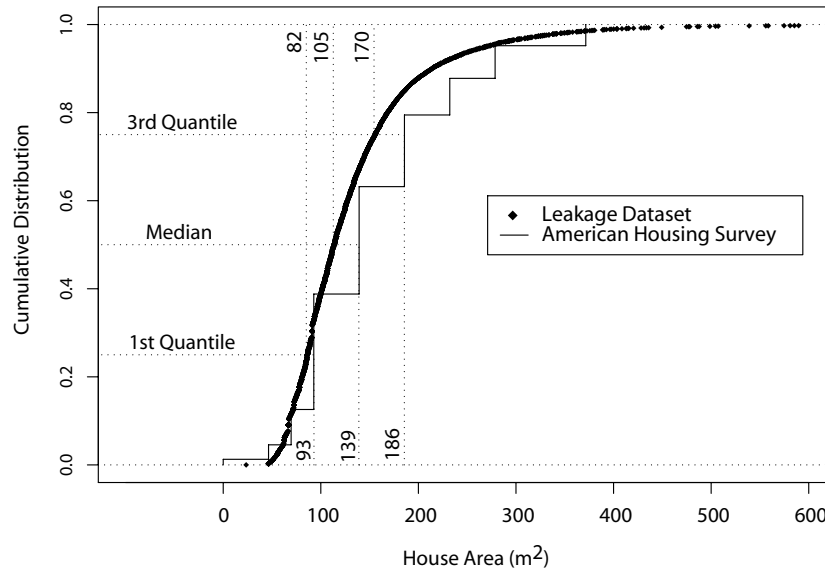
Figure 2: Comparison of the cumulative distribution function of the floor area of houses in the leakage database to the American Housing Survey 1999. Highlighted for each distribution are the 1st and 3rd quartiles (i.e., 25th and 75th percentiles), and the median (i.e., 50th percentile). The dotted line presents a smoothed estimate for the cumulative distribution function of the 1999 AHS.

The floor areas of houses in the leakage database are generally smaller than those reported in the AHS (Fig. 2). This is because the leakage database is dominated by measurements from the Ohio Weatherization Program. This program only included residences with household incomes lower than 125% of the poverty guideline. Consequently, the houses tested under this program have floor areas smaller than the national average (Fig. 3). On the other hand, there are also some states where the houses tested have floor areas much larger than the national average, like Utah and New Hampshire.

3.3.3   House Year-Built

Comparison with the AHS data shows that houses in the air leakage database are slightly older than houses in the AHS dataset, though they span similar ranges (Fig. 4). Fig. 5 shows the distribution of year built plotted by state. Many of the houses in the dataset are quite old because they are low-income houses tested in the Ohio Weatherization Program.
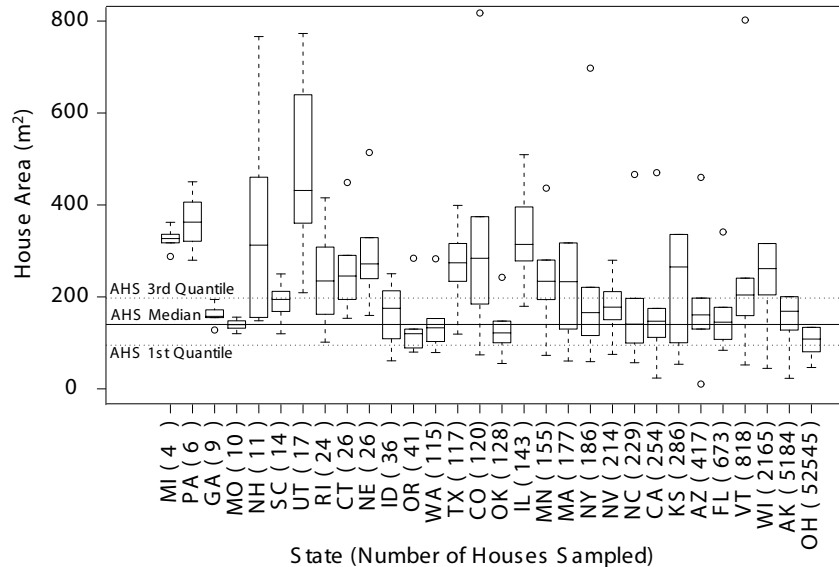
Figure 3: The floor area of the houses in the air leakage database sorted by the state where the measurement was taken. The solid line across the graph is the median floor area from the AHS 1999, and the two dashed lines are the 1st and 3rd quartiles. In parentheses are the number of data points collected from each state.

On the other extreme are newly constructed homes tested by various energy-efficiency programs, with samples from Alaska being an example. A few well sampled states like Wisconsin, North Carolina, and New York have houses built from several decades.

3.3.4 Other Characteristics

There are other characteristics in our database that are less commonly reported, including the presence of heating ducts, the type of floor/basement construction, and whether the house is participating in an energy-efficiency program.

The presence of a thermal distribution system can add significant leakage. Sherman and Dickerhoff [1998] report that when leaks from duct systems are measured separately, they account for almost 30% of the total leakage of the house. The AHS classifies heating equipment into several types, but the two that use ducts as part of the system are warm-air furnaces and electric heat pumps. They represent 60% and 10% of the total housing stock respectively. In the air leakage database, there are nearly 2,000 data points that record the
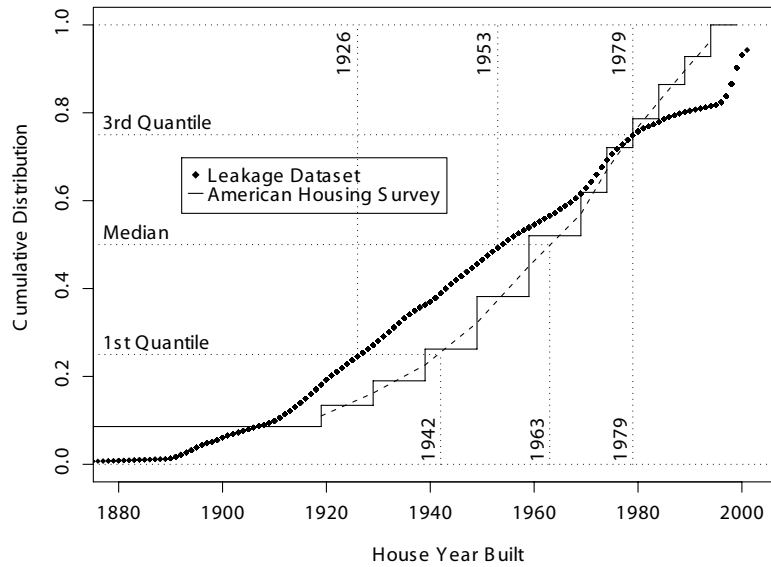
Figure 4: Comparison of the cumulative distribution function of the year built of houses in the air leakage database to the American Housing Survey 1999. Highlighted for each distribution are the 1st and 3rd quartiles (i.e., 25th and 75th percentiles), and the median (i.e., 50th percentile). The dotted line presents a smoothed estimate for the cumulative distribution function of the 1999 AHS.
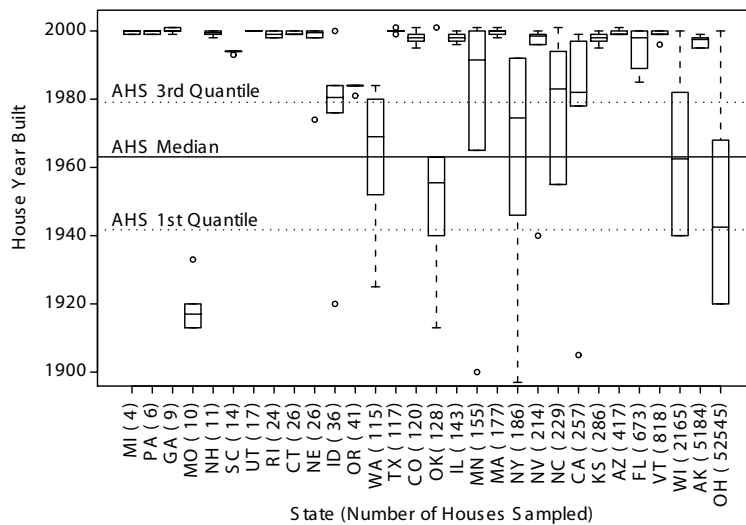

Figure 5: The year built of the houses in the air leakage database sorted by the state where measurement was taken. The solid line across the graph is the median year-built of the houses surveyed, and the two dashed lines are the 1st and 3rd quartiles. In parentheses are the number of data points collected from each state.
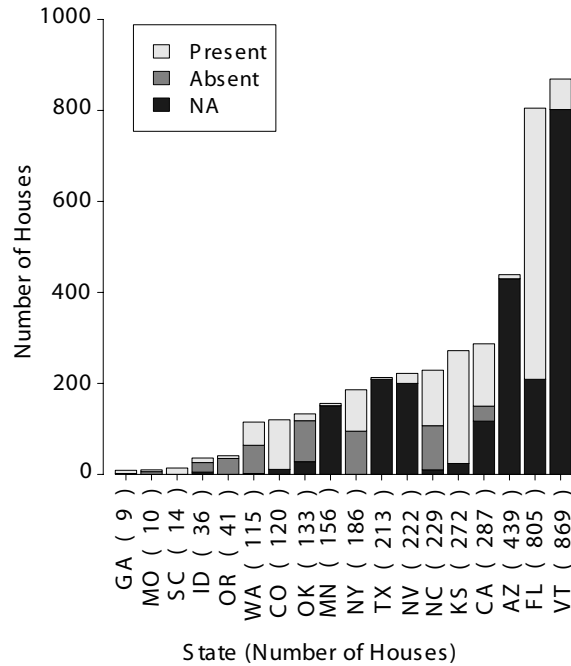
Figure 6: Histogram of the presence of thermal duct systems of houses in the leakage database sorted by the state where the measurement was taken. States that do not specify the presence or absence of a duct system are not shown. In parentheses are the number of data points in each states including NA (Not Available) data points.

presence or absence of duct systems. Of these 2,000 houses, 76% have a thermal distribution system. In states like North Carolina and New York (Fig. 6), there are many houses in each category which allow the effect of ducts on air leakage to be studied.

Sherman and Dickerhoff [1998] point out that the normalized leakage of houses with a slab-on-grade foundation is significantly less than for houses with a crawlspace or an unconditional basement. Unfortunately, we know the underfloor construction for fewer than 10% of the houses in the database. Fig. 7 shows the different underfloor types. The 1999 AHS reports the presence of a slab, basement, or crawlspace to be about 20%, 31%, and 18% in the US housing stock. However, the survey did not differentiate between conditioned and unconditioned basements. The leakage database, in comparison, reports about 8%, 45%, and 41% of the houses having a slab, basement, or crawlspace, respectively.

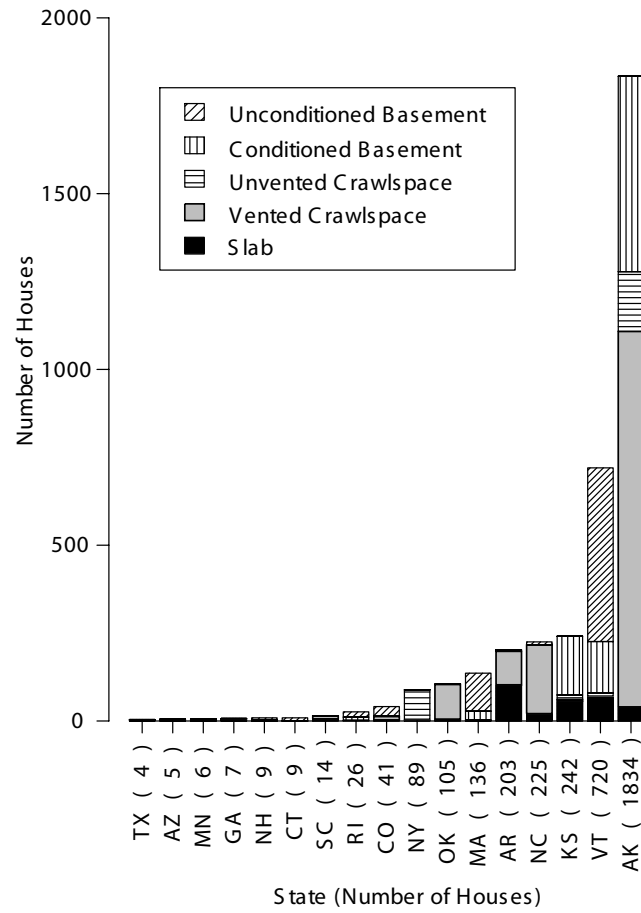Lastly, houses that are participants in energy-efficiency programs are designed to be

Figure 7: Histogram of the underfloor construction of houses in the leakage database, sorted by the state where measurements were taken. Only the data points with known floor/ basement type are shown. In parentheses are the number of data points in each state.
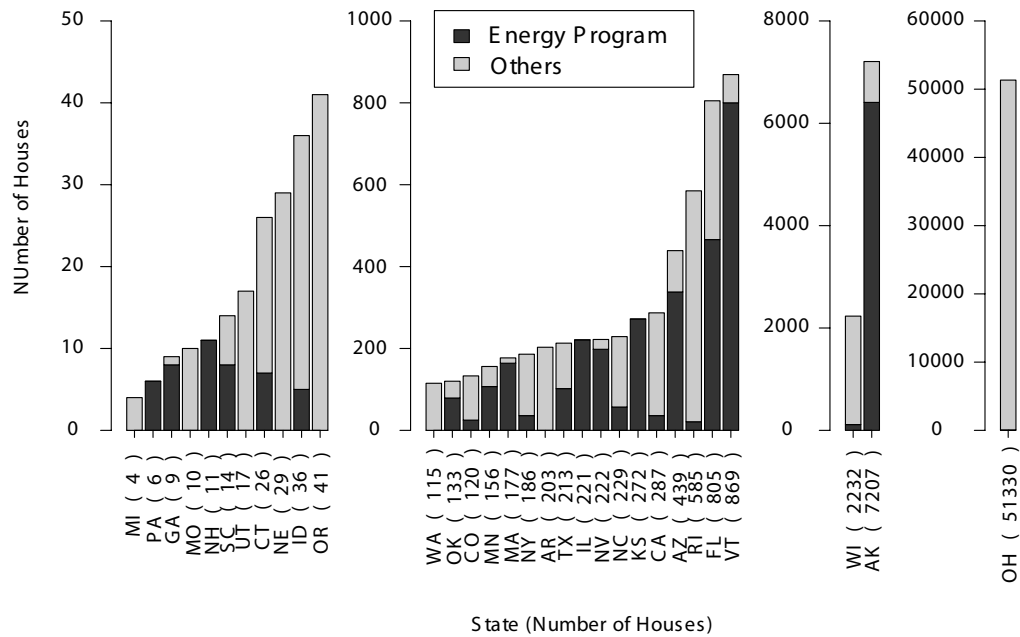
Figure 8: Histograms of the number of energy-efficiency program houses in the leakage database, sorted by the state where the measurement was taken. For ease of display, histograms are grouped according to the total number of samples in that state: 0–50 houses, 100–1,000 houses, 1,000—8,000 houses for Alaska and Wisconsin, and over 50,000 houses for Ohio. Data are grouped according to the total number of samples in each state for easy viewing. In parentheses are the number of data points from each state.

especially air tight to save thermal conditioning costs. These houses have very different leakage distributions and are therefore treated as a separate group in this analysis. Thirteen percent of the measurements are from energy-efficiency programs in 24 different states. All of the data from Alaska, Vermont and Kansas are from energy-efficiency programs, whereas only a small fraction of houses tested in Ohio and Wisconsin are from energy-efficiency programs (Fig. 8). The fraction of houses in an energy-efficiency program in the database is much higher than observed nationally. This is to be expected since blower-door measurements are often used for the energy analysis that is commonly performed on participating energy-efficient houses.

4   DATA ANALYSIS

Houses in the leakage database do not statistically represent the characteristics of the housing in US as a whole because of two main reasons. (1) Data were contributed voluntarily by home weatherization contractors (and others) from around the country, and some contractors contributed much more data than others. (2) Most of the data were gathered as part of programs to target particular classes of homes, primarily "low-income" homes that were tested as part of a weatherization program, and "energy-efficient" homes that were tested to check compliance with air infiltration targets of the energy programs. We seek to adjust for these effects so that an unbiased estimate of leakage distribution can be obtained for any given geographic region. Our approach is to determine the relationship between home characteristics and air leakage, so that we can adjust for different distributions of home types in different communities.

We divided the houses into three broad classes: low-income, energy-efficient, and conventional. A problem with the present data set is that all of the known low-income homes are from Ohio; data from the other states include some low-income homes, but these are not identified in the data. We will discuss this issue further, below.

4.1   Effect of house year-built and floor area

New homes tend to be much tighter than old homes because of improved materials (e.g. weatherstripped windows), better building and design techniques (e.g. insulation and vapor barriers), and lack of age-induced deterioration (e.g. setting of foundation). This trend has been reported by Sherman and Dickerhoff [1998] who observed substantial reduction in leakage in homes built after 1980. In the leakage dataset, houses were not tested all at once. This means not only are there differences in year built, but the age of the house when tested also differs. We tried to determine separately the contribution to leakage from improvements in building technologies, and the aging effect due to deterioration. However,

because almost all blower-door tests were carried out after 1990, we were unable to distinguish the effects of aging from that of construction year. When more leakage data are collected in the future, the effect of aging can perhaps be distinguished from the effect of improved building practices. For the rest of this paper, "year built" is used to describe both effects.

As an exploratory step, we use a locally weighted regression smoothing technique to relate the normalized leakage to year built and floor area (Fig. 9). As expected, normalized leakage is lower in homes built more recently; it also decreases with increasing house floor area. Analysis of the AHS 1999 revealed, unsurprisingly, that household income and house size are correlated: higher-income families tend to have larger homes. We would expect that larger homes are also more likely to be built with quality materials and superior workmanship, and that they tend to be better maintained. So it is not surprising that larger houses are tighter per unit of exterior wall area than smaller ones.

We also applied a classification tree analysis (Breiman et al. [1984]) to find the important building factors that describe the normalized leakage and minimize the variability within each category. We found that the normalized leakage of houses can best be grouped according to their year built into four categories: before 1950, 1950–1980, 1980–1995, and after 1995. These four categories were chosen because this partitioning resulted in groups of houses with the most homogeneous normalized leakage within each category. The exact years at which the splits occurred are affected by clustering of homes in the database by year built.

When categorizing normalized leakage with respect to house floor area, we used intervals employed by the American Housing Survey (AHS) and the US Census. We reduced the number of categories to five, namely: < 93 m^2, 93–140 m^2, 140–186 m^2, 186–232 m^2, and > 232 m^2. Figures 10 to 12 show normalized leakage and effective leakage area of houses sorted according to these categories. Geometric mean, geometric standard
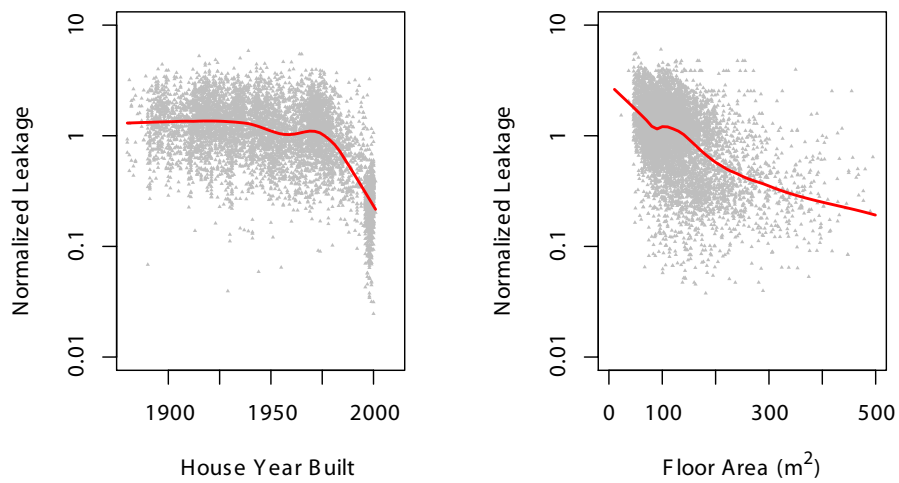
Figure 9: Locally weighted regression smoothing of log(normalized leakage) is displayed against year built and floor area. The smoothed line is influenced by 30% of the data points at each value. The method uses a tri-cube weight function to put more emphasis on values near the current position of the predicted value. Energy-efficiency program houses are excluded because they have especially low air leakage. Due to the large number of data points, only 10% chosen randomly are displayed. These graphs show that normalized leakage appears to have an inverse dependence on year built and floor area. Normalized leakage is presented in log scale because an approximate lognormal distribution is observed.

deviation, and 5th, 25th, 50th, 75th, 95th percentile of normalized leakage and effective leakage area are tabulated in Appendix C Table C1 to C6.

Larger houses are expected to have higher effective leakage areas as there are more surfaces for leaks to develop. Fig. 10 to 12 show such a trend across all year built groups regardless of house types (low-income, conventional or energy-efficient). In terms of normalized leakage, which is thought to be roughly insensitive to floor area, it appears that larger homes are relatively tighter. However, this difference diminishes in newer houses built after 1995.

Low-income houses have much higher leakage areas than conventional houses, regardless of year built and floor area. Energy-efficient houses are moderately tighter than conventional houses. Based on these exploratory analyses, we conclude that house type (low-income, conventional or energy-efficient), year built, and floor area all influence the leakage of a house. As these factors are not completely independent of one another, more detailed analysis is required to determine how each one is associated with leakage.

The geometric standard deviation (GSD) quantifies the variability among houses. When normalized leakage data are categorized according to house year built and floor area according to the classification tree method (Fig. 10 to 12), the observed GSDs in each category vary between 1.36 and 2.12, and average approximately 1.70. Despite the modest variability in *observed* GSDs, the actual variability in GSDs in this dataset is still likely to be overstated because it includes the effect of small-sample variability. House categories with large numbers of observations have observed GSDs very close to 1.70 (see "Prior GSD" in Fig. 13). The variation in observed GSDs does not seem to be related (upon examination) to year built, area, geometric mean of leakage, or house type.

We applied the idea of hierarchical modeling from Bayesian statistics to estimate the posterior geometric standard deviation from the prior population, using a precision-weighted averaging method (Gelman et al. [1995]). We assume that the actual

Figure 10: Effect of house year built and floor area on normalized leakage (1st row) and effective leakage area (2nd row) of low-income houses in Ohio. Houses are first grouped into four categories according to year built: before 1950, 1950–1980, 1980–1995, after 1995. Within each year-built group, houses are further divided into five subgroups according to floor area (from left to right): < 93 m^2, 93–140 m^2, 140–186 m^2, 186–232 m^2, and > 232 m^2. The number below each boxplot shows the number of samples in that particular size bin.
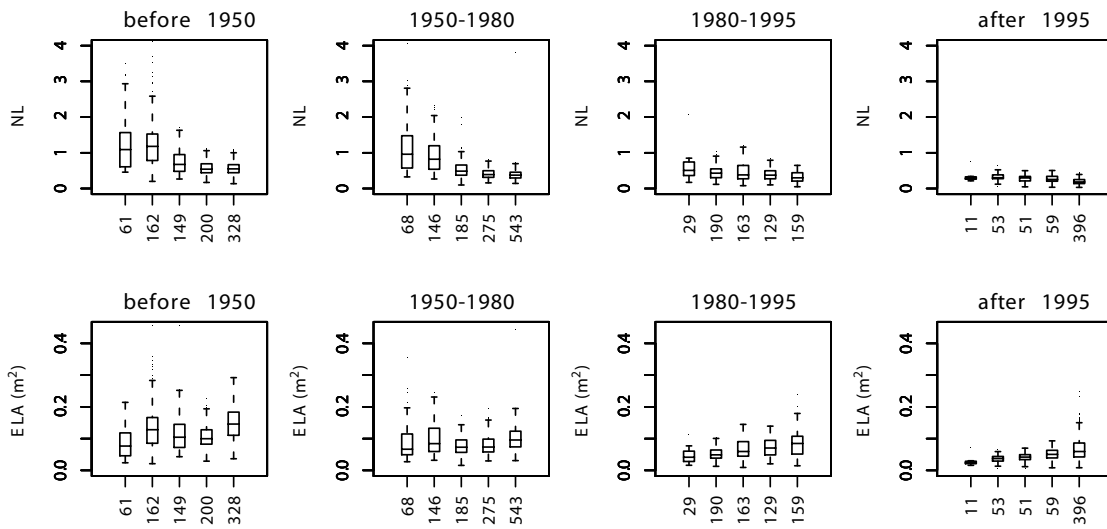
Figure 11: Effect of house year built and floor area on normalized leakage (1st row) and effective leakage area (2nd row) of conventional houses (known not to be participants of a low-income weatherization program nor an energy-efficiency program). Houses are first grouped into four categories according to year built: before 1950, 1950–1980, 1980–1995, after 1995. Within each year-built group, houses are further divided into five subgroups according to floor area (from left to right): < 93 m^2, 93–140 m^2, 140–186 m^2, 186–232 m^2, and > 232 m^2. The number below each boxplot shows the number of samples in that particular size bin.
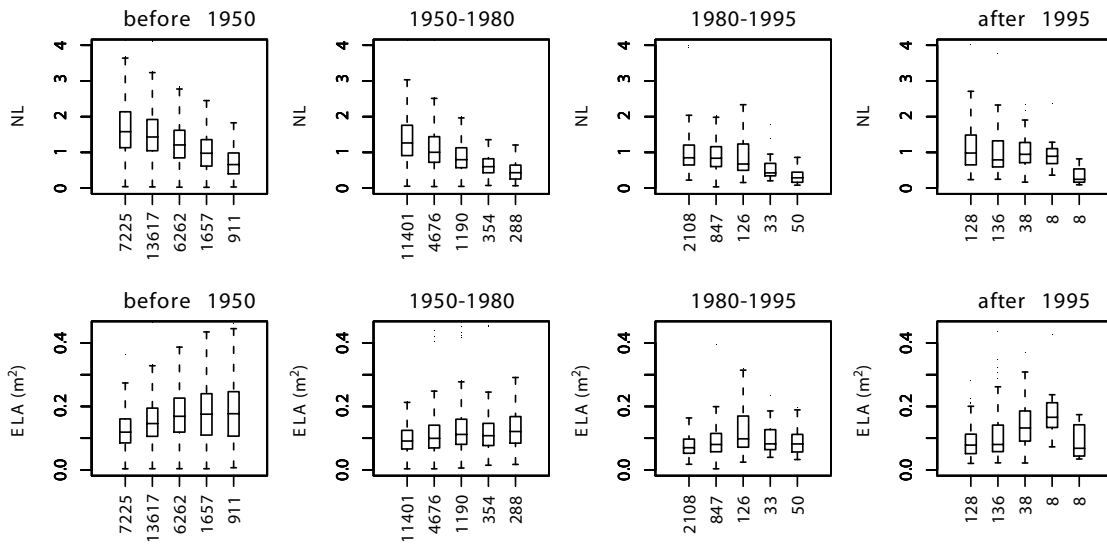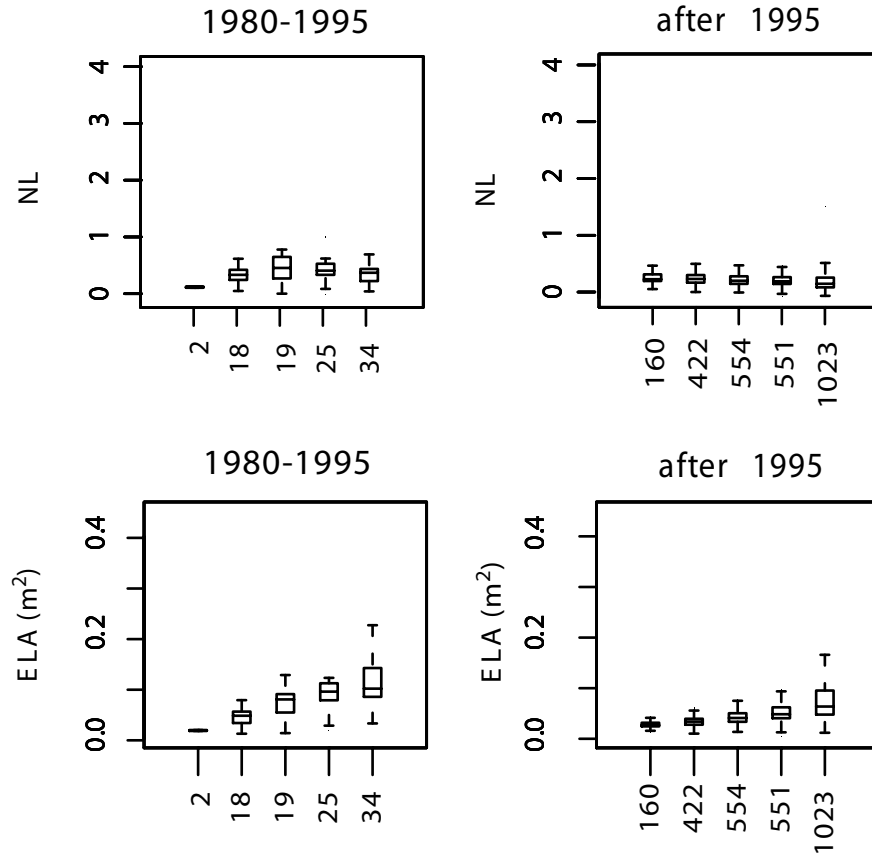
Figure 12: Effect of house year built and floor area on normalized leakage (1st row) and effective leakage area (2nd row) of houses that are known to be in an energy-efficiency program. Houses are first grouped into two categories only (there are no energy-efficiency program houses built before 1980 in the dataset) according to year built: 1980–1995 and after 1995. Within each year-built group, houses are further divided into five subgroups according to floor area (from left to right): < 93 m^2, 93–140 m^2, 140–186 m^2, 186–232 m^2, and > 232 m^2. The number below each boxplot shows the number of samples in that particular size bin.

GSDs are drawn from a Gaussian, or "normal", distribution, with a standard deviation τ; $\tau = 0$ would imply that all of the GSDs are identical, whereas a large value of τ would imply that the actual GSDs are highly variable. τ can be estimated by subtracting the within sample mean square (MS_W) from the between sample mean square (MS_B) and normalizing the difference by the number of sample groups (N). Assuming that the actual GSDs are normally distributed with mean μ and standard deviation τ, and given an observed GSD (GSD_i^{obs}), with standard error σ_i, the posterior estimate for the GSD is given by Eq. (7):

$$
\begin{aligned}
GSD_i &= \frac{\frac{1}{\sigma_i^2} \cdot GSD_i + \frac{1}{\tau^2} \cdot \mu}{\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}} \\
\tau^2 &= \frac{(MS_B - MS_W)}{N}
\end{aligned}
\tag{7}
$$

The result is shown in Fig. 13 (see "Posterior GSD"); the so-called "posterior estimates" of the GSDs range from 1.52 to 1.76, and average 1.69. Thus, although there is substantial variability in normalized leakage within each type of house, the variability appears to be independent of house types.

Analyzing the data in terms of distinct categories might make sense for some parameters, but this approach seems inappropriate for others. Specifically, we would not expect that continuous variables such as year built or floor area would be associated with discrete jumps in leakage (or normalized leakage), but would rather expect leakage to vary smoothly with year built or floor area.

An alternative to an analysis based on categories is to use linear regression, which assumes that the relationship between the predicted variable and the explanatory variables is linear, and that the residuals are normally distributed. A quantile-quantile (Q-Q) plot (not shown) indicates that normalized leakage within each year-built and floor area
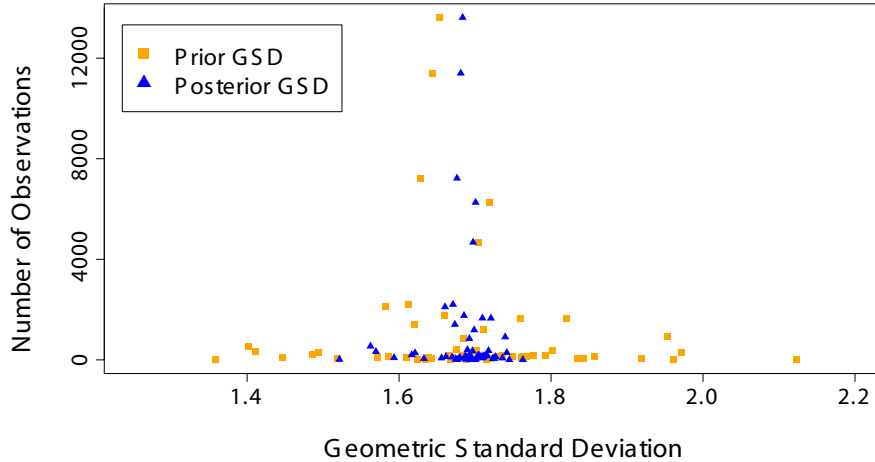
Figure 13: Scatter plot showing the prior and posterior geometric standard deviations of all 60 groups (4 year-built bins × 5 floor area bins × 3 house types, the same intervals used in Fig. 10 to 12). The hierarchical model narrows the range of GSDs by compensating samples of few observations with the mean GSD of all available groups.

subgroup is approximately lognormal. However, most of the plots show an excess of homes with extremely low normalized leakage, compared with what would be expected from a lognormal distribution. There are about 1,300 data points with reported air flow rates less than $170 \frac{\text{m}^3}{\text{h}}$ (or 100 cfm) at 50 Pa; all of these data are from the Ohio Weatherization Program. We have excluded those data from further analysis because blower-door tests showing less than $170 \frac{\text{m}^3}{\text{h}}$ at 50 Pa are mostly likely to be invalid entries (McWilliams, personal communication).

We used linear regression to investigate the relationship between the logarithm of normalized leakage and several explanatory variables. We first performed an analysis of the entire data set, using indicator variables (0=no, 1=yes) to identify "Energy Program" and "Low-Income" homes; after transforming back from log space, this corresponds to the following model:

$$NL = \exp[\beta_0 + \beta_1 \cdot \text{YearBuilt} + \beta_2 \cdot \text{Area} + \beta_3 \cdot \text{ind(EnergyProg)} + \beta_4 \cdot \text{ind(LowIncome)} + \varepsilon] \quad (8)$$

where ε is the residual term between the model prediction and observed value.

This approach yielded a reasonably good model fit ($R^2 = 0.56$); regression coefficients are shown in Table 2. Much of the predictive power of the model is due to the effects associated with energy-efficiency program homes, which are substantially less leaky than conventional homes, and low-income homes, which are much leakier.

As mentioned previously, all of the known low-income homes in the data set are from Ohio. We expect that the quantitative results for these homes should apply reasonably well to the midwest in general, but we are much less confident about other areas of the country. It does seem likely that low-income homes will everywhere be leakier than conventional homes of the same age and size, because low-income homes tend to be less well constructed, to use lower-quality materials, and to be less well maintained; however, the exact magnitude of this effect may vary with the region of the country.

The issue is further complicated by two facts. First, we know that some of the conventional and energy-efficient houses in the database are occupied by families with low incomes. Incomplete identification of houses with families having low incomes from the rest compromises our ability to analyze the effect of income on the leakage statistics of all house types. Second, we are aware of some recent effects to promote energy efficiency in houses especially designed for low-income families. As a result, some low-income homes are unlikely to be any leakier than houses occupied by higher income families. Collection of additional leakage data on low-income houses from other parts of the country will be required to fully justify the assumption that "low-income" houses tend to be leakier nationwide.

Using a single model for the entire data set, as discussed above, assumes that floor area and year built have the same influence on all homes, whether conventional, energy-efficient, or low-income, which investigation shows is not the case. We also performed separate analyses for each house type using this model:

Table 2: Multi-variable linear regression parameters for normalized leakage of all valid data points, using low-income and energy-efficiency program as indicator variables.

	Coefficients	Estimate	Std. Error	t-value	Multiple R^2
β_0	(Intercept)	9.63×10^0	0.16×10^0	59.8	
β_1	Year Built	-5.03×10^{-3}	0.08×10^{-3}	-62.0	
β_2	Floor Area	-2.69×10^{-3}	0.03×10^{-3}	-84.5	0.56
β_3	ind(Energy Program)	-4.80×10^{-1}	0.11×10^{-1}	-43.4	
β_4	ind(Low-Income)	6.10×10^{-1}	0.10×10^{-1}	59.4	

$$NL = \exp[\beta_0 + \beta_1 \cdot \text{YearBuilt} + \beta_2 \cdot \text{Area} + \varepsilon] \tag{9}$$

Table 3 shows the resulting coefficient estimates. The normalized leakage of energy-efficient houses is nearly independent of floor area, but that is not true for either low-income or conventional houses. Figure 14 shows the comparison of predicted leakage to observed leakage. The 95% confidence interval bands are computed from the standard error of the regression using the F distribution (S-Plus Guide, Ch.6). Owing to the large amount of data, the coefficient estimates are rather precise. However, the large residual term ε indicates that even for homes with a given age and area, the normalized leakage varies widely.

As shown in Fig. 15, the mean value of log(Normalized Leakage) is well predicted by the regression for each category of homes, even though individual homes in each category may have much higher or lower leakage. The vertical bars represent variances of the means calculated using the posterior standard deviation determined for each of the groups. The horizontal bars show the 95% confidence interval of the prediction. We conclude that the regression model predicts the geometric means of normalized leakage with high fidelity (R^2 ranges from 0.86 to 0.92). The largest discrepancies came from groups that have relatively few observations. Except for these outliers, the rest of the residuals appear to be well behaved.

The error terms ε for each of the regression models appear to follow an approximately
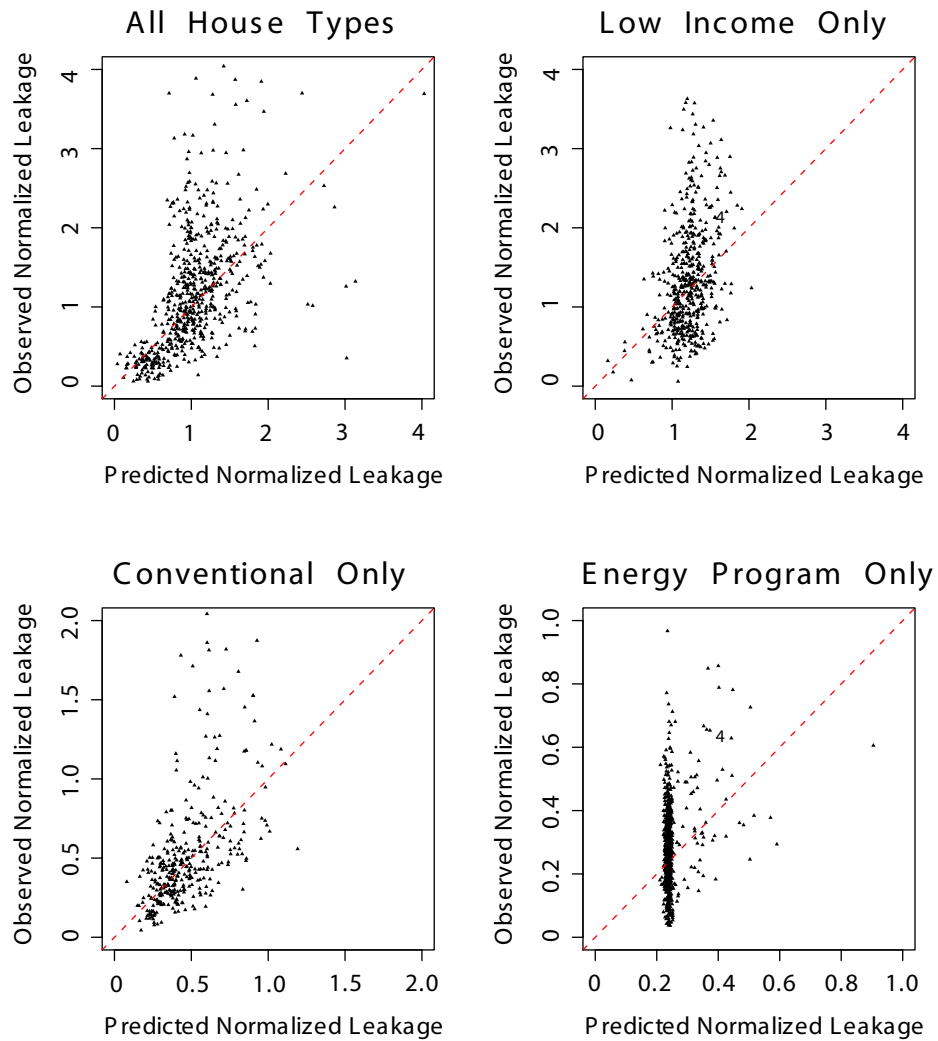
Figure 14: Regression model performance against observed values. Due to the large number of observations, only a few hundred houses, chosen at random, are displayed for ease of visualization.
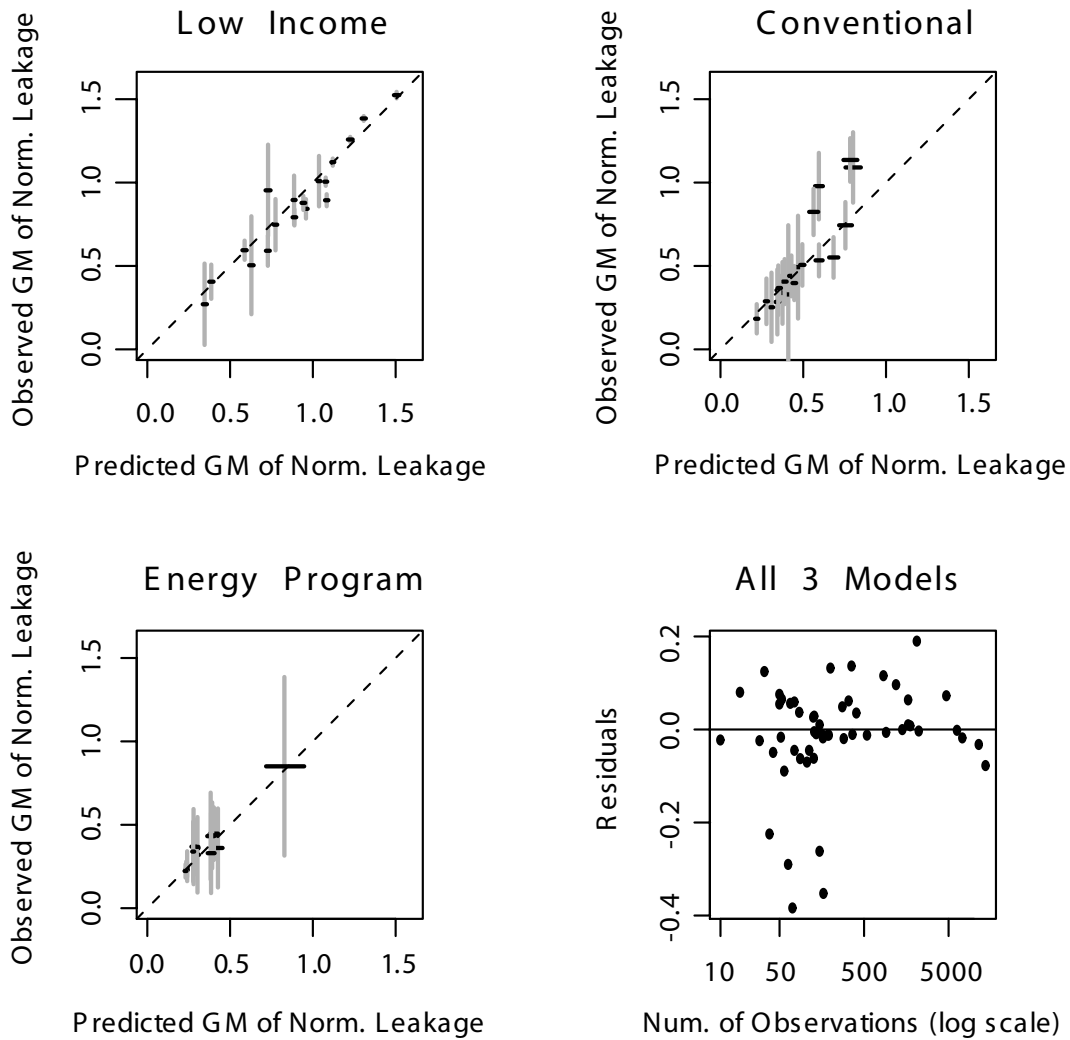
Figure 15: Scatter plot of observed geometric mean of normalized leakage as reported in Appendix C against predicted values based on linear regression models. The vertical error bar on each data point represents the sampling variance of the means calculated from the posterior geometric standard deviation. The horizontal bar shows the smoothed 95% confidence interval of the regression model predictions. The magnitude of residuals is seen to decrease when number of observations exceeds about 300.

Table 3: Multi-variable linear regression parameters for normalized leakage of low-income, energy-efficiency program and conventional houses.

Types		Coefficients	Estimate	Std. Error	t-value	Multiple R^2
	β_0	(Intercept)	$1.11 \times 10^{+1}$	$0.02 \times 10^{+1}$	67.0	
Low-Income	β_1	Year Built	-5.37×10^{-3}	0.08×10^{-3}	-63.7	0.17
	β_2	Floor Area	-4.18×10^{-3}	0.04×10^{-3}	-96.6	
	β_0	(Intercept)	$2.07 \times 10^{+1}$	$0.06 \times 10^{+1}$	35.7	
Conventional	β_1	Year Built	-1.07×10^{-2}	0.03×10^{-2}	-36.2	0.41
	β_2	Floor Area	-2.20×10^{-3}	0.07×10^{-3}	-29.4	
	β_0	(Intercept)	$3.43 \times 10^{+1}$	$0.21 \times 10^{+1}$	16.6	
Energy Program	β_1	Year Built	-1.79×10^{-2}	0.10×10^{-2}	-17.2	0.04
	β_2	Floor Area	-1.827×10^{-4}	0.57×10^{-4}	-3.24	

normal distribution $N(\mu, \sigma^2)$, and with the same statistical distributions in each regression (Fig. 16): low-income $N(0.00, 0.26)$, conventional $N(0.00, 0.27)$, energy program $N(0.00, 0.27)$. Regression analysis shows that the error terms are not a function of year built or floor area. Based on these results we conclude that the random errors are well-behaved and satisfy the least-squares estimation assumptions of the regression models.

We also tested to see if additional variables should be included when trying to predict the leakage distribution of a population of houses. We created four binary variables, each indicating whether the data point is collected from a corresponding US Census Region (Northeast, South, North Central, and West). None of the binary variables improved the fit of the model. By further dividing the data into nine US Census divisions (see Table 1 for distribution of data points in terms of regions and divisions) we also found that there is no relationship between location and normalized leakage. Another parameter of interest is the presence of ducts, since studies have shown that they can be a major source of leaks. The binary variable (1=ducts, 0=no ducts) introduced to indicate the presence of ducts again did not improve the fit of the model.

The only remaining available factor that improves the fit of the model other than year built and floor area is a binary variable indicating the presence of direct leaks through
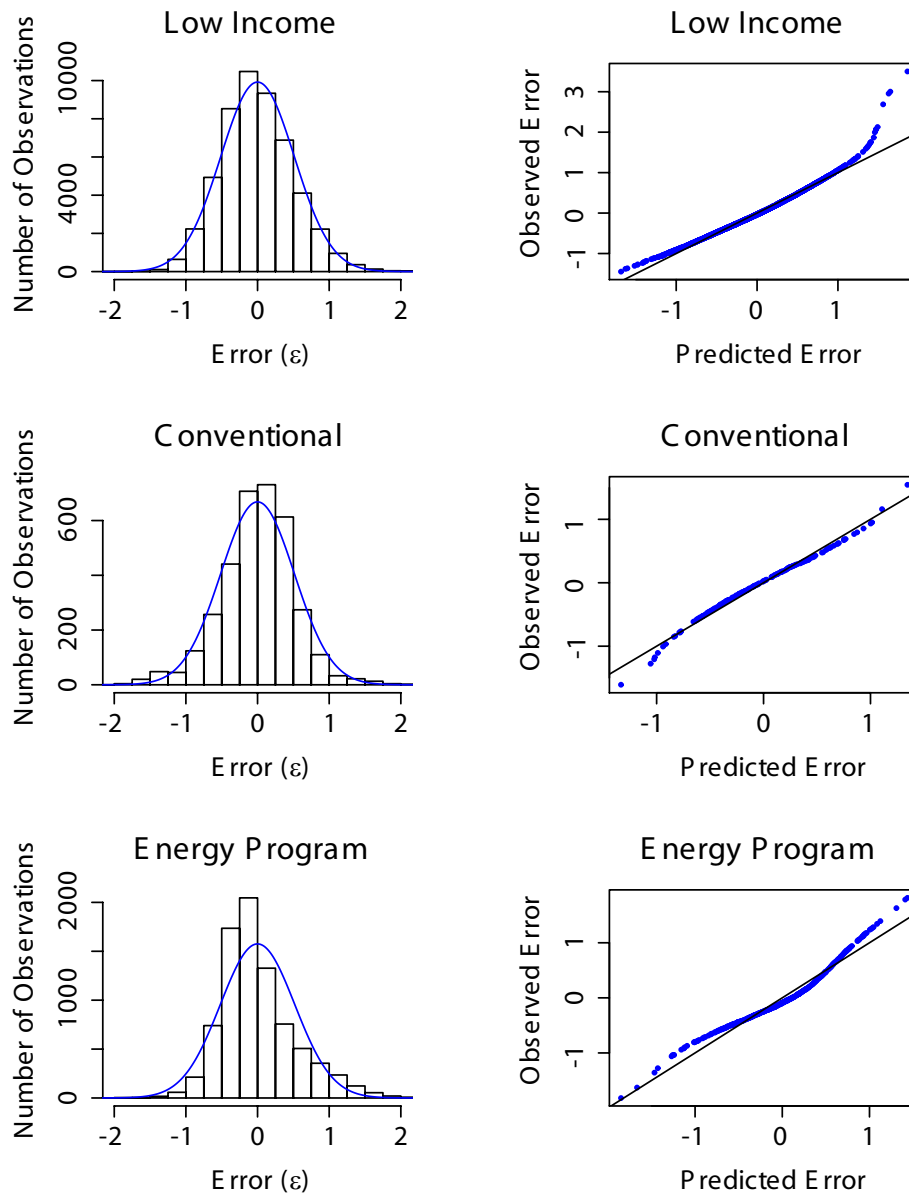
Figure 16: Residuals from the multiple linear regression model are shown for low-income, conventional, and energy-efficiency program houses. They are roughly normally distributed with means equal to zero and consistent variances.

crawlspace or unconditioned basement. The new regression model is as follows:

$$NL = \exp(\beta_0 + \beta_1 \cdot \text{YearBuilt} + \beta_2 \cdot \text{FloorArea} + \beta_3 \cdot \text{ind(Leaks)}) \qquad (10)$$

Houses with slab or conditioned basement are considered to have no direct floor leaks to the outside. Table 4 summaries the number of houses where the floor type is known[3]. Using only these data points, we obtained new linear regression models using Eq. (10). The difference in the estimated normalized leakage due to the presence of direct floor leaks can be calculated as follows:

$$
\begin{aligned}
\Delta\text{NL} &= \frac{NL_{\text{Leaks}} - NL_{\text{NoLeaks}}}{NL_{\text{NoLeaks}}} \\
&= \frac{\exp(\beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{Area} + \beta_3) - \exp(\beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{Area})}{\exp(\beta_0 + \beta_1 \cdot \text{Year} + \beta_2 \cdot \text{Area})} \\
&= \exp(\beta_3) - 1
\end{aligned}
$$

Using this approach, we found that conventional houses with direct floor leaks are 56% leakier than those without, but the difference is only 7% among energy-efficient houses. However, owing to the small number of homes known to have leaky floors, the fit of the model improves only slightly by incorporating this additional building characteristic (conventional: $\Delta R^2 = +0.025$; energy-efficiency program: $\Delta R^2 = +0.005$). A key complication is that for the rest of the houses, for which the floor construction types are not known, it is most likely that some have direct leaks to the outside, and some do not.

To summarize, Fig. 17 shows trends with respect to year built and floor area. Low-income houses appear to have higher leakage than conventional and energy-efficiency program houses, regardless of year built and floor area. Their normalized leakage areas vary more substantially with floor area than with year built. The opposite is true for

---

[3]No information is available regarding the floor construction type of low-income houses.

Table 4: Number of conventional and energy-efficient houses with known floor construction types. Houses that have crawlspace or unconditioned basement are categorized as having direct floor leaks to the outside; houses that have a conditioned basement or slab are considered to have no direct floor leaks.

	Have Floor Leaks	No Floor Leaks
Conventional Houses	390	35
Energy-Efficiency Program Houses	604	497
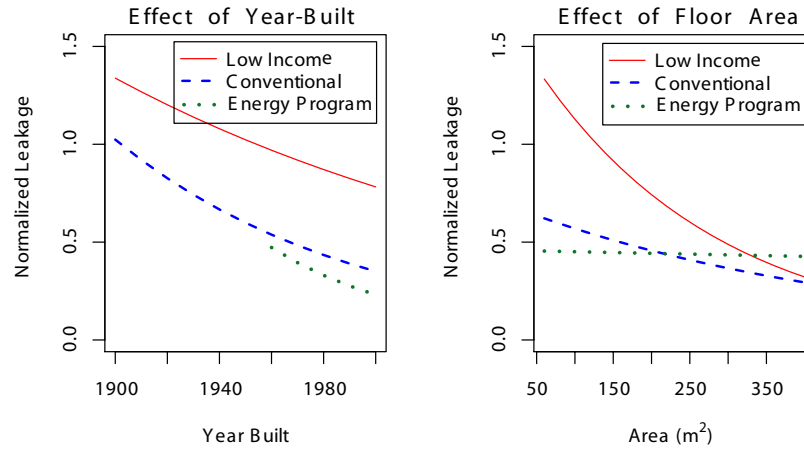
Figure 17: Effects of year built and floor area on normalized leakage based on multiple linear regression models described in Table 3. When calculating the effect of year built, floor area is set at 140 m^2 which is the median value reported by the American Housing Survey. Similarly, when calculating the floor area effect, the year built is chosen to be at the median value, which is 1963.

conventional and energy-efficiency program houses, where normalized leakage is almost independent of floor area, which is the intended effect of normalization. For energy-efficient houses, the regression is shown only from 1960 onwards because there are not enough data points to provide leakage information for houses built before 1960.

5 DATA AGGREGATION

The present report analyzes the available data on leakage area in the US housing stock. We find these main results:

1. At least in the midwest, low-income houses are much leakier than conventional

homes, by more than a factor of two in terms of normalized leakage.

2. Older homes are leakier than newer homes, to an extent that depends on whether the home is low-income or conventional (for 30-year-old versus new homes, this is about a 20% effect for low-income midwest homes, and a 35% effect for conventional homes).

3. Within a particular house type (midwest low-income, conventional or energy-efficient), age, and size of home, the distribution of normalized leakage is approximately lognormal and moderately variable, with a geometric standard deviation of about 1.7.

4. The results for conventional (non-low-income) houses seem to apply throughout the country, with little evidence of substantial regional differences, although our data lack the spatial coverage and representative sampling that would allow us to conclude this with certainty.

5. The results for low-income houses, and for the difference between low-income and conventional houses, are probably correct for the midwest but are much less certain for the rest of the country (for which we have no data on low-income homes). Based on the surprising lack of spatial variability in conventional homes, and the fact that the factors that lead to higher leakiness in low-income homes should apply throughout the country, it seems reasonable that the midwest results might apply elsewhere in the country, but this inference cannot be tested with the present data.

The normalized leakage for a particular house type, age, and size of home is moderately variable, so the normalized leakage for any single home cannot be predicted from these three factors. However, it is possible to predict the statistical distribution of normalized leakage over all homes of a given type, age, and size, and to combine these statistical distributions to predict the distribution of normalized leakage for the entire US, or for specific areas.

Since there are significant differences in the leakage behavior of low-income houses versus conventional ones, we make separate calculations for each type of home and then combine the predicted distributions. We ignore energy-efficiency program houses because their prevalence in the current US housing stock is small. For instance, as of July 2001, there were 34,642 EPA ENERGY STAR[4] rated homes in US, which constitute only a small fraction of the housing stock. EPA also suggests ways to make homes more energy efficient. Among those are methods that will reduce air infiltration such as sealing air leaks, installing air and vapor barriers, and sealing ducts. However, the number of houses with leakage areas significantly reduced is believed to be small, and therefore we can neglect it.

In the present analysis, low-income homes are those that qualified for Ohio's weatherization program; homes were eligible if their residents earned below 125% of the poverty guideline (which varies with the number of members of the household). In order to estimate the number of houses belonging to this category, we make use of data published by the American Housing Survey 1999, tabulating income of families by year built, floor area, and number of occupants. Out of 64,536,000 single detached housing units surveyed, about 15% of the households are "low-income." This percentage is estimated for single-family detached dwellings based on the income level of families adjusted by the number of occupants. This information is obtained from the American Housing Survey 1999 and the fractions of "low-income" families estimated by income brackets are shown in Table 5.

Weatherized homes usually have reduced air leakage, but the magnitude of improvement can vary. In addition, the number of houses that have been weatherized in the US remains small. In 2000, DOE's Weatherization Assistance Program estimated that only 16% of currently eligible households have received weatherization services[5]. Since the

---

[4]ENERGY STAR labeled homes are typically at least 30% more energy-efficient than standard homes. These technologies and building practices are used to achieve this improved performance: tight construction and ducts, improved insulation, high performance windows, and energy-efficient heating and cooling equipment.

[5]DOE Weatherization Eligible Households are defined as households with incomes at or below a certain fraction of the Federal poverty level; but a state may instead use a slight variation of this guideline.

Table 5: Estimation of number of single detached "low-income" occupied units for each income bracket based on statistics from 1999 American Housing Survey on income of families and primary individuals.

Income	less than $9,999	$10,000 to $14,999	$15,000 to $19,999	$20,000 to $29,999	$30,000 and over
Single-Family Detached Units (1000's)	3359	3916	3582	7810	43268
"Low-Income" Units	100%	57%	30%	3%	0%

number of weatherized houses is not large enough to affect the overall air leakage distribution of low-income houses, we will ignore this fine adjustment in the analysis to follow.

Data from the American Housing Survey 1999 show that low-income and conventional houses have slightly different distributions of year built and floor area. Not surprisingly, the fraction of low-income houses is highest among older and smaller houses. Table 6 shows that approximately 20% of houses built before 1950 are low-income, but only 3% of the newer homes are occupied by families with low incomes. Table 7 shows that more than 40% of the houses smaller than 500 ft^2 are occupied by low-income families, but this percentage drops to less than 10% for houses that are larger than 2500 ft^2. Figure 18 shows the differences between these distributions graphically.

Differences in the distribution of year built and floor area between low-income and conventional houses are particularly important when one is concerned with predicting the leakage of houses in a smaller spatial unit, such as a census tract, where the percentage of low-income families can be much higher. The US Census provides spatially resolved distributions for year built (e.g. Census 2000 Summary File 3) and floor area with which leakage distributions can be calculated as outlined here.

Based on the data obtained from the AHS 1999 described above, we divided all single-family detached housing units into 72 groups: 12 year-built categories (as outlined in Table 6) and 6 floor-area categories (as outlined in Table 7). To determine the number of

Table 6: Fraction of "low-income" single-family detached housing units as a function of year built based on statistics from the 1999 American Housing Survey on income of families and primary individuals.

Year Built	1919 and earlier	1920 to 1929	1930 to 1939	1940 to 1949	1950 to 1959	1960 to 1969
Single-Family Detached Units (1000's)	5609	3145	3400	4194	6396	8563
"Low-income" Units	21%	20%	22%	23%	13%	14%
Year Built	1970 to 1974	1975 to 1979	1980 to 1984	1985 to 1989	1990 to 1994	1995 to 1999
Single-Faily Detached Units (1000's)	7055	6827	4819	5564	4134	4830
"Low-inome" Units	21%	14%	15%	9%	2%	3%

Table 7: Fraction of "low-income" single-family detached housing units as a function of house floor area based on statistics from the 1999 American Housing Survey on square footage by household income.

Floor Area	Less than 500ft^2	500ft^2 to 999ft^2	1000ft^2 to 1499ft^2	1500ft^2 to 1999ft^2	2000ft^2 to 2499ft^2	2500ft^2 and over
Single-Family Detached Units (1000's)	829	7350	16938	15718	10481	13220
"Low-income" Units	44%	31%	18%	11%	9%	8%
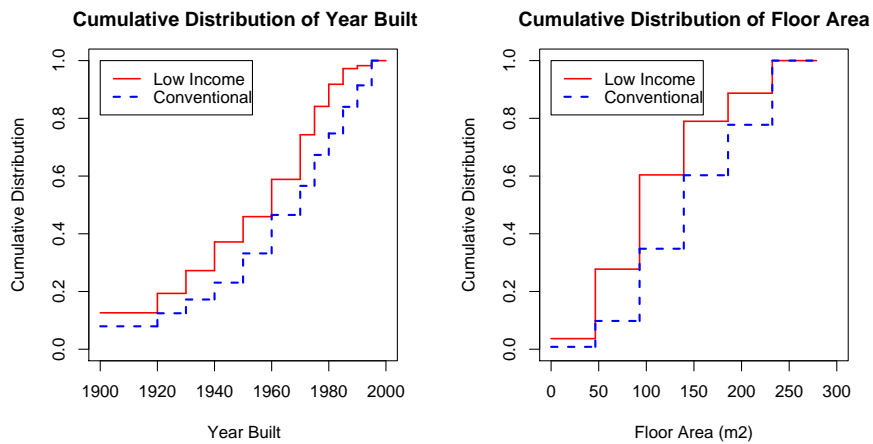
Figure 18: Distribution of year built and floor area of US housing stock showing low-income houses as a separate group. In 1999 when these distributions were obtained by the American Housing Survey, the fraction of houses eligible for the low-income program (at or below 125% of poverty guideline) was 15% of the total single detached housing unit. As expected, the fraction of older, smaller houses is relatively higher for the low-income category. This difference is important to capture when estimating the composite leakage distribution of each group, or their combined distribution.

Table 8: Statistics of estimated normalized leakage distribution weighted for all dwellings in US.

House Type	Conf. Interval	p05	p10	p25	p50	p75	p90	p95
Low-Income		0.32	0.42	0.64	0.98	1.48	2.09	2.60
	[5%,	[0.31,	[0.41,	[0.63,	[0.98,	[1.47,	[2.10,	[2.60,
	95%]	0.32]	0.42]	0.64]	1.00]	1.53]	2.17]	2.65]
Conventional		0.16	0.21	0.30	0.45	0.69	1.01	1.26
	[5%,	[0.16,	[0.20,	[0.29,	[0.44,	[0.66,	[0.96,	[1.20,
	95%]	0.17]	0.21]	0.31]	0.47]	0.72]	1.05]	1.32]
Whole US		0.17	0.22	0.32	0.50	0.79	1.24	1.59
	[5%,	[0.17,	[0.21,	[0.31,	[0.48,	[0.76,	[1.19,	[1.55,
	95%]	0.18]	0.22]	0.33]	0.52]	0.82]	1.27]	1.65]

houses within each of the 72 groups, we assume that the floor area distribution remains the same as in Table 7 regardless of the year built. The correlation between year built and floor area is ignored because the resulting difference is minor. The geometric mean of the normalized leakage area for each group of houses is computed using the linear regression model as shown in Eq. (9) [6]. Finally, an error term (ε) generated to conform to the residual distribution as shown in Fig. 16 is added to the linear model to give a prediction of the normalized leakage distribution of the housing stock.

The results are displayed in Fig. 19 for conventional houses, low-income houses and the composite of the two distributions. Most of the houses have normalized leakage areas below 1.5, with a mean of about 0.64. The statistics of the distributions are summarized in Table 8.

In 1997, the US Environmental Protection Agency published the Exposure Factors Handbook (1997) to provide a summary of the available statistical data on parameters commonly used in human exposure assessments. Chapter 17, entitled Residential Building Characteristics, included a section on air exchange rates and other factors that are also of

---

[6]Even within each group of houses, there are still some variations in year built (5 to 10 years) and floor area (500 ft^2). Depending on which year built and floor area values are used as input parameters for the regression model, slightly different normalized leakage can be obtained for the group. To address this, we simply take the average of normalized leakage predicted at an interval of every 2.5 years and 5 m^2 within the ranges considered for that group of houses.

interest, such as volumes and surface areas of rooms, mechanical ventilation system, foundation types, filtration, inter-zonal air flows, and so on. The exposure analysis community prefers measurements that directly report air exchange rates, which include not only infiltration, but also natural and mechanical ventilation. For this reason, there are several studies analyzing data using the perfluorocarbon tracer (PFT) technique, which measures air exchange rates. There is however a large difference between air exchange rate and normalized leakage; as discussed later, we were restricted by the availability of other blower door data from performing a more direct comparison. We were only able to compare our results to the air exchange rates reported in the EPA Exposure Factors Handbook (1997).

To compare our predictions against published values of air changes per hour, we converted normalized leakage to ACH as follows:

$$ACH \approx \frac{ACH50}{F} \tag{11}$$

where F is a factor used to relate typical air exchanges per hour with the air exchange rate at 50 Pa. Assuming that the volume of a house V [m^3] is approximately equal to the floor area A$_f$ [m^2] (*not* the footprint area) multiplied by an equivalent height H [m](typically about 3 m), Eq. (11) can be simplified to a function of normalized leakage, height of structure H and the parameter F only:

$$
\begin{aligned}
ACH &= \frac{Q_{50}}{V \cdot F} \\
&= \frac{NL\frac{A_f}{1000}\left(\frac{2.5m}{H}\right)^{0.3}\sqrt{\frac{2\cdot(4Pa)}{\rho}}\cdot\left(\frac{50Pa}{4Pa}\right)^{0.65}}{F \cdot V} \\
&= 48\left(\frac{2.5m}{H}\right)^{0.3}\frac{NL}{H \cdot F}[h^{-1}]
\end{aligned} \tag{12}
$$

We must emphasize that the notion of using a scaling factor to relate NL to ACH is a simplified treatment of a complex reality, and also might be inappropriate for detailed analyses of single houses. Normalized leakage describes only the tightness of a building, whereas the air exchange rate also includes operational contributions such as natural ventilation from opening windows and mechanical ventilation from exhaust fans as well as the variable influence of weather. The scaling factor F is an attempt to connect the two concepts. This factor typically varies from 10 to 30 for infiltration[7]. We find that the fit is best for the national data when 12 is used (Fig. 20). Although the main goal of this paper is to demonstrate a method of estimating the leakage distribution of a housing sample from year built and floor area data, we hope to convey the idea that the result is at least reasonable when compared with ACH datasets.

6 CONCLUSION

We have attempted to quantify the leakage characteristics of US houses as a function of household income level, house age, and house size. Although the data do not constitute a representative sample of houses, the consistency of the results across the different sources of data gives us some confidence in the basic results, at least with respect to conventional houses. The situation with low-income houses is more problematic, but we think the quantitative estimates are likely to be at least approximately correct.

The leakage characteristic of a community of houses depends on the distribution of year built and floor area. The US Census Bureau reports structure year built in intervals of 5 to 10 years at the census tract level, which can be used to predict air leakage and thus to estimate air exchange rates. We can obtain some information on the size of dwellings indirectly through related parameters such as the number of rooms. Additional information

---

[7] As an example, EPA ENERGY STAR Home Sealing Specification (2001) defined a "LBL Factor" based on climate region, number of storeys, and amount of sheltering from wind to convert ACH_{50} to ACH. The value of 'LBL Factor" ranges between 9.8 (3-storeys building with no sheilding in cold climate zone 1) to 29.4 (well-shielded, 1-storey building in warm climate zone 4).
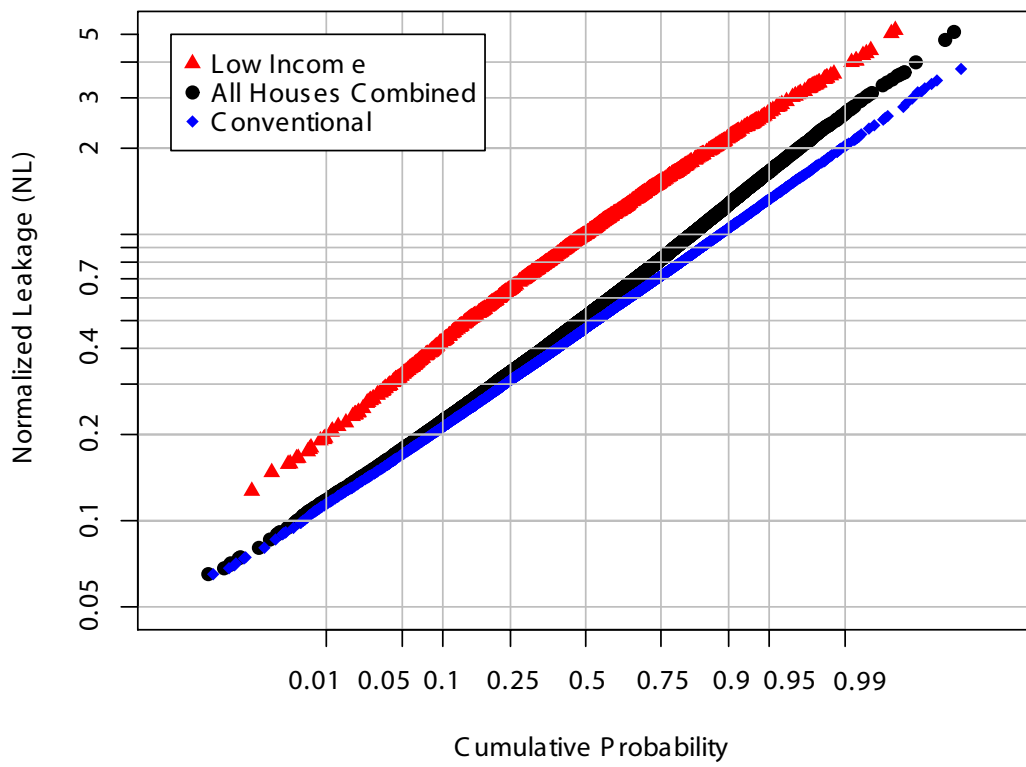
Figure 19: Distributions of normalized leakage for low-income and conventional houses, and their composite. Characteristics of US housing stocks are based on the 1999 American Housing Survey.
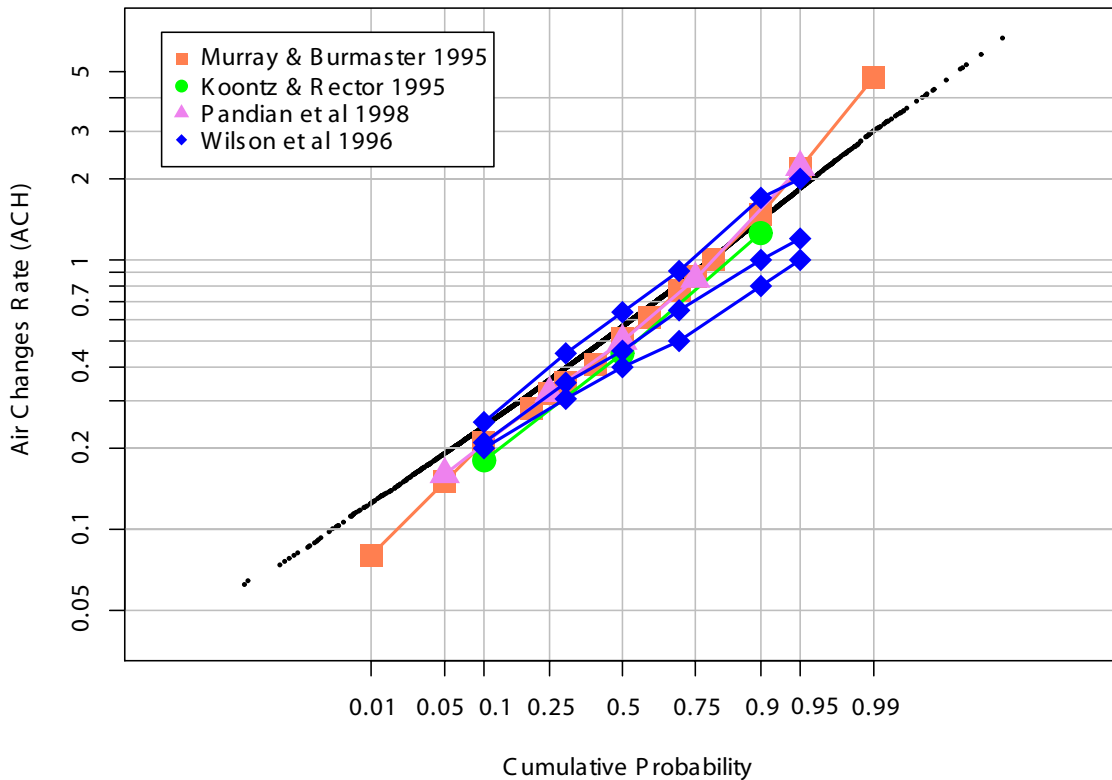
Figure 20: Comparison of air exchange rates estimated from linear regression models and values found in literature. The distribution reported by Pandian et al. [1998] are essentially inclusive of all the studies listed. With Wilson et al. [1996] being the exception in that all residences measured are located in California (3 distributions are shown: measurements from L.A. being most leaky, followed by San Diego and Northern California), the rest are analyses of collections of multiple projects. Koontz and Rector [1995] assigned weights to the results to compensate for geographic imbalance of measurements. Murray and Burmaster [1995] presented results as functions of weather using the degree-day metric.

is also available on the number of storeys, and on the presence of ducts. Owing to the limited leakage data, however, we are unable to quantitatively assess the influences of these factors on the distribution of normalized leakage area. From the analysis of the better reported attributes like year built and floor area, we find the leakage distribution to be roughly lognormal. This means not only do we have to consider the combinations of different house characteristics in a community of houses, but also the variations between individual houses owing to the wide range of leakage values present. We have demonstrated a simple method that makes use of a residential air leakage database as weighting factors in conjunction with the regression models developed to best estimate the normalized leakage distribution of the US residential housing stock.

ACKNOWLEDGMENTS

References

ASTM, Standard Test Method for Determining Air Leakage by Fan Pressurization. American Society of Testing and Materials, 1992. ASTM Book of Standards Volumn 04.07.

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees.* Wadsworth, 1984.

A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Anaylsis*, chapter 5, pages 134–141. Texts in Statistical Science. Chapman and Hall, London UK, first edition, 1995.

M.D. Koontz and H.E. Rector. Estimation of distributions for residential air exchange rates. Technical report, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, D.C., 1995. Contract No. 68-D9-0166, Work Assignment No. 3-19.

J. McWilliams, August 2002. Personal communication.

D.M. Murray and D.E. Burmaster. Residential air exchange rates in the united states: Emperical and estimated parametric distributions by season and climatic region. *Risk Analysis*, 15(4):459–465, 1995.

Office of Energy Efficiency and Renewable Energy. Buildings energy databook. buildingsdatabook.eren.doe.gov, 2002. Chapter 7.1: Building Profiles on Low-Income Housing.

M.D. Pandian, J.V. Behar, W.R. Ott, L.A. Wallace, A.L. Wilson, S.D. Colome, and M. Koontz. Correcting errors in the nationwide data base of residential air exchange rates. *Journal of Exposure Analysis and Environmental Epidemiology*, 8(4):577–586, 1998.

M.H. Sherman. *Air Infiltration in Buildings.* PhD thesis, University of California, Berkeley, CA 94720, October 1980. LBL-10712.

M.H. Sherman. The use of blower-door data. *Indoor Air*, 5:215–224, 1995.

M.H. Sherman and D.J. Dickerhoff. Air tightness of u.s. dwellings. In *Proceedings*, Buxton, UK, 1998. 15th Air Infiltration and Ventilation Centre Conference. LBL-48671.

M.H. Sherman and N.E. Matson. Residential ventilation and energy characteristics. *ASHRAE Transaction*, 103(1), 1997. LBNL-39036.

M.H. Sherman and N.E. Matson. Air tightness of new u.s. houses: A preliminary report. Technical report, Lawrence Berkeley National Laboratory, Berkeley, CA, March 2002. LBNL-48671.

US EPA ENERGY STAR. Home sealing specification. http://www.energystar.gov/ia/home_improvement/home_sealing/ES_HS_Spec_v1_0b.PDF, October 2001.

US EPA ENERGY STAR. Program snapshot. http://yosemite1.epa.gov/estar/homebuyers.nsf/content/NewsWhatsNewSnapshot.htm, January 2003.

S-Plus Guide to Statisical and Mathematical Analysis. StatSci, A Division of MathSoft, Inc., Seattle, Washington, 3.2 edition, Decemeber 1993.

American Housing Survey for the United States 1999–Current Housing Reports. U.S. Department of Housing and Urban Development and U.S. Census Bureau, issued october 2000 edition, 1999. Table 2-18, 2-20.

US EPA, Residential Building Characteristics - Exposure Factors Handbook. US EPA, volume iii, chapter 17 edition, August 1997.

A.L. Wilson, S.D. Colome, Y. Tian, E.W. Beck, P.E. Baker, D.W. Behrens, I.H. Billick, and C.A. Garrison. California residential air exchange rates and residence volumes. *Journal of Exposure Analysis and Environmental Epidemiology*, 6(3):311–326, 1996.

APPENDIX

A   List of Symbols

A_f	Building floor area	m^2
ACH	Air exchange rate	h^{-1}
$ACH50$	Air exchange at 50 Pa pressure difference	h^{-1}
ELA	Effective leakage area	m^2
f_s	Stack factor	$\frac{m}{s} \cdot K^{\frac{1}{2}}$
f_w	Wind factor	[-]
H	Building height	m
n	Power law flow exponent	[-]
NL	Normalized leakage area	[-]
P_f	Pressure	Pa
P_r	Reference pressure	Pa
Q_f	Air flow rate	$\frac{m^3}{s}$
s	Specific infiltration	$\frac{m}{s}$
ΔT	Indoor-outdoor temperature difference	K
v	Wind speed	$\frac{m}{s}$
κ	Leakage coefficient	$\frac{m^3}{s \cdot Pa^n}$
ρ	Density of air	$1.2\frac{kg}{m^3}$ at 1 Pa, 293K

B List of Contributors

The contributions of leakage and related data made by individuals and organizations are listed below:

Advanced Energy Corporation	Alaska Housing Finance Corporation
Arkansas Energy Office	Building Science Corporation
Building America	Building Industry Institute
Conservation Services Group	Davis Energy Group
Rob DeKieffer	E-Star Colorado
Geoff Reiler (Sitka, Ak)	Florida Solar Energy Center
Guaranteed Wattsavers	Kansas Energy-Star
Lawrence Berkeley National Laboratory	Ohio Home Energy Rating System
Ohio Weatherization Program	Vermont Energy Investment Corporation
Energy Rated Homes of Vermont	Daran Wastchak, L.L.C.
Wisconsin Energy Conservation Corporation	Wisconsin Energy Star Homes

C   Tables of Normalized Leakage and Effective Leakage Area

Statistics of the normalized leakage area and the effective leakage area in terms of house
year built and floor area for the three house types: low-income, conventional,
energy-efficient. Geometric mean and geometric standard deviation are not shown for
number of observation less than 8. Values at 5th percentile are not shown for number of
observations less than 20.

Table C1: Statistics of the normalized leakage area of low-income houses tested by the
Ohio Weatherization Program. Tabulated are the number of observation, geometric mean,
geometric standard deviation, and the 5th, 25th, 50th, 75th, 95th precentiles.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m²	before 1950	7225	1.52	1.63	0.67	1.13	1.58	2.14	3.14
	1950-1979	11401	1.26	1.64	0.56	0.91	1.26	1.76	2.83
	1980-1995	2108	0.89	1.58	0.45	0.65	0.85	1.20	2.03
	after 1995	128	1.01	1.77	0.39	0.65	0.98	1.47	2.84
93-139 m²	before 1950	13617	1.38	1.65	0.59	1.05	1.43	1.92	2.99
	1950-1979	4676	1.01	1.70	0.42	0.72	1.00	1.44	2.38
	1980-1995	847	0.84	1.68	0.37	0.60	0.84	1.16	1.98
	after 1995	136	0.90	1.75	0.45	0.59	0.79	1.31	2.33
140-185 m²	before 1950	6262	1.12	1.72	0.41	0.85	1.21	1.62	2.39
	1950-1979	1190	0.79	1.71	0.33	0.57	0.79	1.13	1.89
	1980-1995	126	0.75	1.86	0.27	0.50	0.67	1.23	2.33
	after 1995	38	0.95	1.67	0.43	0.71	0.92	1.26	2.18
186-232 m²	before 1950	1657	0.88	1.82	0.30	0.62	0.97	1.36	1.96
	1950-1979	354	0.59	1.70	0.24	0.43	0.60	0.82	1.28
	1980-1995	33	0.50	1.67	0.21	0.34	0.43	0.69	0.95
	after 1995	8	0.89	1.72					
>232 m²	before 1950	911	0.59	1.95	0.17	0.40	0.66	0.98	1.49
	1950-1979	288	0.41	1.97	0.12	0.25	0.43	0.64	1.21
	1980-1995	50	0.27	1.84	0.10	0.16	0.29	0.45	0.68
	after 1995	8	0.28	2.12					

Table C2: Statistics of the normalized leakage area of conventional houses that are not participants of a low-income or an energy-efficiency program.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m^2	before 1950	63	1.09	1.84	0.47	0.61	1.09	1.57	3.20
	1950-1979	71	0.98	1.92	0.37	0.60	0.96	1.55	2.88
	1980-1995	29	0.49	1.62	0.17	0.34	0.48	0.74	0.85
	after 1995	17	0.33	1.36		0.28	0.32	0.34	0.48
93-139 m^2	before 1950	166	1.14	1.78	0.37	0.79	1.19	1.53	2.93
	1950-1979	149	0.82	1.73	0.35	0.54	0.82	1.26	2.21
	1980-1995	191	0.44	1.74	0.22	0.30	0.43	0.55	1.11
	after 1995	85	0.34	1.45	0.20	0.28	0.33	0.41	0.61
140-185 m^2	before 1950	149	0.74	1.76	0.33	0.48	0.67	0.95	2.24
	1950-1979	186	0.51	1.68	0.23	0.37	0.49	0.66	1.78
	1980-1995	163	0.41	1.79	0.16	0.27	0.38	0.64	1.05
	after 1995	75	0.28	1.57	0.11	0.23	0.31	0.36	0.47
186-232 m^2	before 1950	200	0.55	1.49	0.31	0.44	0.54	0.69	1.09
	1950-1979	275	0.40	1.49	0.21	0.31	0.39	0.50	0.75
	1980-1995	129	0.35	1.59	0.15	0.27	0.37	0.50	0.74
	after 1995	67	0.25	1.64	0.06	0.21	0.27	0.35	0.43
>232 m^2	before 1950	328	0.53	1.41	0.29	0.44	0.55	0.67	0.89
	1950-1979	543	0.37	1.40	0.21	0.30	0.37	0.46	0.61
	1980-1995	159	0.29	1.67	0.13	0.20	0.30	0.44	0.58
	after 1995	403	0.18	1.68	0.07	0.14	0.19	0.25	0.39

Table C3: Statistics of the normalized leakage area of energy-efficiency program houses.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m^2	before 1950	10	0.85	1.69	0.00	0.50	0.67	1.47	1.81
	1950-1979	53	0.36	1.83	0.22	0.25	0.31	0.52	0.90
	1980-1995	52	0.32	1.52	0.17	0.24	0.30	0.39	0.65
	after 1995	366	0.25	1.80	0.08	0.19	0.30	0.36	0.52
93-139 m^2	before 1950	9	0.86	1.63					
	1950-1979	112	0.45	1.61	0.22	0.32	0.43	0.57	0.99
	1980-1995	127	0.36	1.69	0.14	0.26	0.37	0.50	0.82
	after 1995	1661	0.23	1.76	0.08	0.17	0.26	0.34	0.47
140-185 m^2	before 1950	9	0.86	1.96					
	1950-1979	75	0.44	1.76	0.19	0.30	0.40	0.64	1.09
	1980-1995	106	0.37	1.67	0.14	0.28	0.35	0.51	0.83
	after 1995	2205	0.24	1.61	0.09	0.20	0.27	0.33	0.42
186-232 m^2	before 1950	2							
	1950-1979	42	0.43	1.68	0.19	0.27	0.42	0.67	0.87
	1980-1995	57	0.37	1.72	0.16	0.26	0.40	0.51	1.05
	after 1995	1409	0.24	1.62	0.09	0.20	0.26	0.32	0.43
>232 m^2	before 1950	2							
	1950-1979	50	0.33	1.69	0.12	0.22	0.34	0.48	0.76
	1980-1995	88	0.34	1.64	0.12	0.25	0.39	0.49	0.56
	after 1995	1762	0.22	1.66	0.09	0.16	0.24	0.32	0.46

Table C4: Statistics of the effective leakage area (m^2) of low-income houses tested by the Ohio Weatherization Program.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m^2	before 1950	7225	0.11	1.61	0.05	0.09	0.12	0.16	0.23
	1950-1979	11401	0.09	1.61	0.04	0.07	0.09	0.12	0.20
	1980-1995	2108	0.07	1.56	0.04	0.05	0.07	0.10	0.15
	after 1995	128	0.08	1.72	0.03	0.05	0.08	0.11	0.19
93-139 m^2	before 1950	13617	0.14	1.65	0.06	0.11	0.15	0.19	0.30
	1950-1979	4676	0.10	1.70	0.04	0.07	0.10	0.14	0.23
	1980-1995	847	0.08	1.69	0.04	0.06	0.08	0.11	0.19
	after 1995	136	0.09	1.76	0.04	0.06	0.08	0.14	0.24
140-185 m^2	before 1950	6262	0.16	1.71	0.06	0.12	0.17	0.23	0.34
	1950-1979	1190	0.11	1.70	0.05	0.08	0.11	0.16	0.26
	1980-1995	126	0.10	1.84	0.04	0.07	0.10	0.17	0.31
	after 1995	38	0.13	1.69	0.07	0.09	0.13	0.19	0.30
186-232 m^2	before 1950	1657	0.16	1.81	0.05	0.11	0.18	0.24	0.35
	1950-1979	354	0.11	1.71	0.04	0.08	0.11	0.15	0.23
	1980-1995	33	0.09	1.65	0.04	0.06	0.08	0.13	0.19
	after 1995	8	0.17	1.66					
>232 m^2	before 1950	911	0.16	1.83	0.05	0.11	0.18	0.25	0.35
	1950-1979	288	0.12	1.75	0.05	0.08	0.12	0.17	0.32
	1980-1995	50	0.08	1.58	0.03	0.05	0.08	0.11	0.18
	after 1995	8	0.08	1.87					

Table C5: Statistics of the effective leakage area (m^2) of conventional houses that are not participants of a a low-income or an energy-efficiency program.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m^2	before 1950	61	0.08	1.81	0.04	0.04	0.08	0.12	0.21
	1950-1979	68	0.07	1.89	0.03	0.05	0.07	0.11	0.21
	1980-1995	29	0.04	1.56	0.02	0.03	0.04	0.06	0.08
	after 1995	11	0.03	1.48					
93-139 m^2	before 1950	162	0.12	1.76	0.04	0.08	0.13	0.17	0.29
	1950-1979	146	0.09	1.69	0.03	0.06	0.08	0.13	0.21
	1980-1995	190	0.05	1.65	0.02	0.04	0.05	0.06	0.13
	after 1995	53	0.04	1.46	0.02	0.03	0.04	0.04	0.05
140-185 m^2	before 1950	149	0.11	1.68	0.05	0.07	0.10	0.15	0.28
	1950-1979	185	0.08	1.62	0.04	0.06	0.07	0.10	0.24
	1980-1995	163	0.06	1.73	0.02	0.04	0.06	0.09	0.14
	after 1995	51	0.06	6.38	0.01	0.03	0.04	0.05	0.07
186-232 m^2	before 1950	200	0.10	1.49	0.05	0.08	0.10	0.13	0.19
	1950-1979	275	0.07	1.48	0.04	0.06	0.07	0.10	0.14
	1980-1995	129	0.07	1.56	0.03	0.05	0.07	0.09	0.13
	after 1995	59	0.06	4.07	0.01	0.04	0.05	0.06	0.08
>232 m^2	before 1950	328	0.14	1.48	0.07	0.11	0.15	0.18	0.28
	1950-1979	543	0.10	1.47	0.05	0.07	0.10	0.12	0.17
	1980-1995	159	0.07	1.70	0.03	0.05	0.08	0.11	0.16
	after 1995	396	0.07	3.66	0.02	0.04	0.06	0.09	0.16

Table C6: Statistics of the effective leakage area (m^2) of energy-efficiency program houses.

Floor Area	Year Built	# Data	GM	GSD	p05	p25	p50	p75	p95
<93 m^2	1980-1995	2							
	after 1995	160	0.03	1.36	0.02	0.02	0.03	0.03	0.05
93-139 m^2	1980-1995	18	0.04	1.61		0.03	0.04	0.05	0.06
	after 1995	422	0.03	1.39	0.02	0.03	0.03	0.04	0.06
140-185 m^2	1980-1995	19	0.06	1.83		0.05	0.08	0.09	0.09
	after 1995	554	0.04	1.40	0.02	0.03	0.04	0.05	0.07
186-232 m^2	1980-1995	25	0.08	1.79	0.02	0.07	0.09	0.11	0.18
	after 1995	551	0.05	1.41	0.03	0.04	0.05	0.06	0.09
>232 m^2	1980-1995	159	0.07	1.70	0.03	0.05	0.08	0.11	0.16
	after 1995	1023	0.07	1.66	0.03	0.05	0.06	0.10	0.15