Air Leakage of U.S. Homes: Model Prediction

Max Sherman and Jennifer McWilliams

January 2007
Disclaimer

While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, nor any other sponsor of this work makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California or any other sponsor. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California or any other sponsor.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

This work was also supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program, of the U.S. Department of Energy under contract No. DE-AC03-76SF00098.
ABSTRACT

Air tightness is an important property of building envelopes. It is a key factor in determining infiltration and related wall-performance properties such as indoor air quality, maintainability and moisture balance. Air leakage in U.S. houses consumes roughly 1/3 of the HVAC energy but provides most of the ventilation used to control IAQ. The Lawrence Berkeley National Laboratory has been gathering residential air leakage data from many sources and now has a database of more than 100,000 raw measurements. This paper uses a model developed from that database in conjunction with US Census Bureau data for estimating air leakage as a function of location throughout the US.
Table of Contents

ABSTRACT .................................................................................................................. 1
List of Symbols ........................................................................................................... 2
Introduction ................................................................................................................. 3
Modeling Methods ..................................................................................................... 3
   Leakage Model ....................................................................................................... 3
   Table 1: Values of Model Parameters .................................................................. 4
   Figure 1: Climate Zones Defined by Building Science Corporation .................... 4
   Ventilation Model ................................................................................................ 5
   Stochastic Method ............................................................................................... 5
   Ventilation Effectiveness using the Sherman Wilson model ................................ 5
Data Sources .............................................................................................................. 6
   RECS data .......................................................................................................... 6
   Census data ......................................................................................................... 6
   Weather data ...................................................................................................... 6
Characteristics of the U. S. Housing Stock .............................................................. 6
   Figure 2: Average Year Houses Were Built in Each County .............................. 7
   Figure 3: Mean Dwelling Height in Meters by County ....................................... 8
   Figure 4: Probability of Floor Leakage by County .......................................... 8
   Figure 5: Fraction of Households Below 125% of the Poverty Level ................ 9
   Figure 6: Dwelling Floor Area by County [m$^2$] ............................................. 10
Discussion ................................................................................................................. 10
   Figure 7: Normalized Leakage by County ......................................................... 11
   Figure 8: Yearly Average Natural Air Exchange Rate by County .................... 12
   Figure 9: Ventilation Efficiency by County ....................................................... 12
   Figure 10: Effective Ventilation by County ....................................................... 13
Summary and Conclusions ....................................................................................... 13
References .................................................................................................................. 14
List of Symbols

Age  Age of house (yr)
Area  floor area inside the pressure boundary (m²) [ft²]
ELA  Effective Leakage Area as measured by ASTM E779 or equivalent (m²) [ft²]
N_{story}  height of the building above grade divided by the height of a single story (-)
NL  Normalized Leakage (-)
NL_{cz}  Normalized Leakage coefficient for each climate zone (-)
size  Floor area divided by the reference area of (100 m²) [1000 ft²]
\phi  Model coefficient (-) for property indicated by subscript
P  Probability (-) Is zero if it does not have property indicated by subscript; is unity if it does.

Subscripts:
Eff  Designates Energy-Efficient construction
Floor  Floor leakage possibility (e.g. vented crawlspace)
Height  Height of house above grade
LI  Designates Low-income

For equations 3-10:
\( A_C \) = leakage area in the ceiling plane [m²]
\( A_F \) = leakage area in the floor plane [m²]
\( A_o \) = total leakage area of the structure [m²]
\( C \) = wind shielding class parameter
\( Q \) is the airflow [m³/s]
\( A,B \) = terrain parameters
\( h \) = height of structure [m]
\( T_o \) = inside temperature [295 K]
\( g \) = acceleration of gravity [9.8 m/s²]
\( I_i \) = air change rate in the \( i^{th} \) time step
\( \Delta t \) = length of each discrete time period [s]
Introduction

Air leakage through the building envelope contributes to ventilation, heating and cooling costs and moisture migration. Understanding the magnitude of the leakage in an individual envelope is important in optimizing the HVAC system and in retrofitting. Understanding the magnitude of leakage in the building stock is important for prioritizing both research efforts and conservation measures for policy makers in both the public and private sector.

The Lawrence Berkeley National Laboratory (LBNL) has gathered air leakage data from homes all over the United States. The database contains more than 100,000 individual measurements of residential envelope leakage. The database has been used to develop a descriptive model of residential air leakage as a function of house characteristics. The purpose of this study is to predict ventilation of houses in each county across the United States using publicly available data.

“Air Tightness” is the property of building envelopes most important to understanding ventilation. It is quantified in a variety of ways all of which typically go under the label of “air leakage”. Air tightness is important from a variety of perspectives, but most of them relate to the fact that air tightness is the fundamental building property that impacts infiltration. There are a variety of definitions of infiltration, but fundamentally infiltration is the movement of air through leaks, cracks, or other adventitious openings in the building envelope. The modeling of infiltration (and thus ventilation) requires a measure of air tightness as a starting point. More extensive information on air tightness can be found in Sherman and Chan (2003), who review the state of the art. This information is also part of a broader state of the art review on ventilation compiled by Santamouris and Wouters (2005).

Sherman and Chan (2003) also discuss the topic of metrics, reference pressures and one versus two parameter descriptions in some detail and will not be discussed here. We have chosen to use the metric of Normalized Leakage (NL) as defined by ASHRAE Standard 119 (1988, 2005) as our primary metric to describe air tightness of houses because it removes the influence of house size and height:

\[ NL = 1000 \cdot \frac{E_{LA}}{\text{Area}} \cdot \left( N_{\text{story}} \right)^{0.3} \]  

(1)

By such a normalization, this metric allows us to compare the leakiness of different house independent of size.

Modeling Methods

This study consisted of two parts: 1) developing a regression model from the leakage data, and 2) applying that model to existing data of housing stock characteristics to come up with leakage characteristics for the United States housing stock. The regression model is documented in other papers (e.g., McWilliams and Jung (2006)). This paper will focus on the application of that model to predict envelope leakage.

The first step in the prediction is to use the model to calculate the leakage area for each county using publicly available data. Once the leakage area is known, we calculate the airflow through the building envelope for every hour of the year using the LBL model (Sherman, 1980) with hourly temperature and wind speed data. The flow is converted to an air change rate by dividing by the volume of the house, and the air change rate can be related to ventilation effectiveness using the Sherman Wilson model (Sherman and Wilson, 1986)

Leakage Model

McWilliams and Jung (2006) used the data in the LBL air leakage database to create a predictive model that can be used to estimate the air tightness of a house based on certain physical characteristics. Their model is shown in equation (2) and the values of the parameters are shown in Table 1. The parameters, which are all dimensionless, were determined by regression analysis.

\[ NL = NL_{\text{age}} \cdot \phi_{\text{size}}^{-1} \cdot \frac{E_{\text{area}}}{\text{height}} \cdot \phi_{\text{age}} \cdot \phi_{\text{floor}} \cdot \left( N_{\text{story}} \right)^{0.3} \]  

(2)
Table 1: Values of Model Parameters

<table>
<thead>
<tr>
<th>NL_{cz}</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL-Alaska</td>
<td>0.36</td>
<td>$\phi$</td>
<td>1.156</td>
<td>$\phi$</td>
<td>1.08</td>
</tr>
<tr>
<td>NL-Cold</td>
<td>0.53</td>
<td>$\phi$</td>
<td>0.598</td>
<td>$\phi$</td>
<td>2.45</td>
</tr>
<tr>
<td>NL-Humid</td>
<td>0.35</td>
<td>$\phi$</td>
<td>1.0118</td>
<td>$\phi$</td>
<td>LI Age</td>
</tr>
<tr>
<td>NL-Dry</td>
<td>0.61</td>
<td>$\phi$</td>
<td>0.841</td>
<td>$\phi$</td>
<td>LI Area</td>
</tr>
</tbody>
</table>

All of the “P” parameters ($P_{\text{LI}}$, $P_{\text{Floor}}$ and $P_{\text{Eff}}$), can be treated in the model as either the probability of being true or as a fraction of the (large) sample for which it is true. $P_{\text{LI}}$ is unity for a low-income house and zero otherwise. $P_{\text{Eff}}$ is unity if the house has participated in an energy efficiency program and zero otherwise. $P_{\text{Floor}}$ is unity if the house has any air leakage through the floor plane and zero if there is no air leakage through that pathway.

The Normalized Leakage coefficients for the four climate zones, NL_{cz}, represent the average normalized leakage for a house in the reference condition. The reference condition is when all of the exponents are zero, which means a 100 m², single-story, non-energy efficient, unaged, slab-on-grade, non-low-income house.

The climate zones are based on combinations of those defined by Building Science Corporation as shown in Figure 1. The climates that were used in the model were Humid (made up of Mixed-Humid and Hot-Humid), Dry (made up of Marine, Mixed-Dry and Hot-Dry), Alaska (containing all counties in Alaska), and Cold (all counties in Cold, Very Cold and Subarctic that are not in Alaska). The climate coefficients for Humid and Alaska are similar and substantially lower than the coefficients for the Dry and Cold areas.

This model is based on a limited dataset, but should provide accurate leakage estimates when applied to broad enough spectrum of houses. Although the uncertainty of an individual prediction is estimated to be on the order of 50%, larger biases may be present when the narrow samples are used. For example, this model is expected to be biased high for conventional new construction, because of increasing improvements made to envelope air tightness in recent years.

Many areas of the country are under-represented in the database. It is not known whether this under-representation causes bias errors, but efforts should be made to fill in the gaps in the database to determine the sizes of such biases and to improve the model.

Figure 1: Climate Zones Defined by Building Science Corporation
Ventilation Model

The LBL ventilation model, shown in Equations (3) through (8), estimates flow through the building shell, $Q(t)$, based on the leakage area of the shell, wind and stack factors, and TMY weather conditions at the house site. The ventilation was calculated for each hour in a typical meteorological year.

$$Q(t) = (ELA) \cdot s(t)$$  \hspace{1cm} (3)

where

$$s(t) = \sqrt{f_s^2 \cdot \Delta T + f_w^2 \cdot V^2}$$  \hspace{1cm} (4)

$$f_s = \frac{1 + \frac{R}{2}}{3} \left(1 - \frac{x^2}{2 - R^2}\right)^{\frac{3}{2}} \left(\frac{g \cdot h}{T_o}\right)^{\frac{1}{2}}$$  \hspace{1cm} (5)

$$f_w = C \left(1 - R\right)^{\frac{3}{2}} A \left(\frac{h}{10}\right)^B$$  \hspace{1cm} (6)

$$X = \frac{A_c - A_F}{A_o}$$  \hspace{1cm} (7)

$$R = \frac{A_c + A_F}{A_o}$$  \hspace{1cm} (8)

Stochastic Method

A stochastic simulation method was used in the predictive calculation to produce a distribution of calculated building properties using input distributions of floor area, height, age, foundation type, energy efficiency improvements, and resident income for houses in each of the 3141 United States counties. The stochastic simulation allows the estimation of the mean and standard distribution for each of the calculated properties, namely equivalent leakage area (ELA), and the wind and stack parameters, $f_w$ and $f_s$ respectively.

For each county, we simulated data for 2000 houses by drawing each of the variables independently from a known distribution for that county. The sample size of 2000 was determined to be sufficient because the distribution did not become better defined when a sample twice as large was used. NL was calculated for each simulated house, and was then transformed into ELA since this is the input variable that is needed to calculate ventilation air flow, $Q$, in the ventilation model.

Ventilation Effectiveness using the Sherman Wilson model

Because ventilation removes pollutants from indoor air, a measure of indoor air quality could be the temporal average of the instantaneous ventilation rate. However, since pollutant concentration is non-linear with respect to ventilation rate a simple average cannot be used. Instead, the term effective ventilation is defined as the steady state ventilation that would yield the same average pollutant concentration over some time period as the actual time varying ventilation in that same time period. It is important to note that the contaminant source strength must be constant over the period of interest. This holds for many building contaminants where the source emission varies slowly with time or operates in a stepwise fashion, and is unaffected by ventilation rate. Some important exceptions are radon, formaldehyde, and carbon monoxide where the emission rate can be affected by the ventilation of the building. In such cases, more detailed techniques may be required.

Effective ventilation is often calculated by first calculating the inverse, the characteristic time ($\tau_e$) for the pollutant concentration to reach steady state, which is given by Equation (9).
it
\tau_{ei} = \frac{1-e^{-I/\Delta t}}{I} + \tau_{ei-1} \cdot e^{-I/\Delta t} \quad (9)

The mean ventilation efficiency is a non-dimensional quantity which is defined as the ratio of the mean effective ventilation to the mean instantaneous ventilation. It is shown in terms of the characteristic time in Equation (10). The closer the actual ventilation rate is to steady state over the period of interest the higher the ventilation efficiency will be.

\varepsilon_m = \frac{1}{I \cdot \tau_c} \quad (10)

Data Sources

The data used for this project came from several different sources. The house characteristics gathered from publicly available data are: location, floor area, age, height of the structure, whether the house participated in an energy efficiency program, the existence of leakage at the floor level, and the income status of the residents. (See McWillimans and Jung (2006) for more details.) The houses in each county represent a distribution of each of the above variables. We assume in this analysis that each of the variables is independent, although this may not be exactly true. For example, it may be that taller houses tend to have a higher probability of floor leakage, but the data for correlation between most of the variables was unavailable and for simplicity we assumed that they are all independent.

RECS data

The Residential Energy Consumption Survey (U.S.D.O.E., 2001) was conducted by the Energy Information Administration for the U.S. Department of Energy, and is a statistically significant representation of the U.S. housing stock as it pertains to energy. The RECS data consists of approximately 4,800 single-family dwelling observations, each of which has approximately 900 reported survey values on topics of building characteristics, resident information, appliances and energy consumption. RECS data is summarized to the census division level so we use Census data which is summarized to the county level wherever possible.

The data of interest to this study that we obtain from RECS are “Total Floor space” (which can be linked to our size variable) and “Number of rooms” used to calculate the average floor space per room, “Number of Stories” used to calculate house height, “Foundation/Basement of Single-Family Homes and Apartments in Buildings” used to determine the existence of floor leaks, and “Income Relative to Poverty Line” used in the determination of the income status of residents. Data for two variables, house height and existence of floor leaks, were collected exclusively from RECS data. Two other variables, floor area and existence of low income residents, used a combination of RECS and Census 2000 data.

Census data

Census data is collected by the U.S. Census Bureau every 10 years. Census 2000 data was used to extract information for: floor area, age, the existence of low income residents for houses in each county across the U.S., and number of rooms given in nine bins of number of rooms per house.

Weather data

The weather data was derived from WYEC (Weather Year for Energy Calculations), TMY (Typical Meteorological Year), TRY (Typical Reference Year), and CTZ (California Climate Zones) weather files. For each county, the most representative weather location was chosen, based primarily on geography. Each weather file contains outside temperature and humidity, wind speed and direction and barometric pressure.

Characteristics of the U. S. Housing Stock

In the figures below we will examine the variation in specific stock characteristics on a regional or county-by-county basis. Each county will be represented by a single value, which is the best estimate of the mean for that county. There is, of course, a distribution around this mean in the stock, but we addressing the distribution of each variable in each country was beyond the scope of this study. Regional variation, however, can sometime be inferred by the county-to-county variation.
First it is useful to look at the spatial distribution of our input data since irregularities in the input data may be recognizable in the calculation output. Most of the data was given not as a single value for each county, but as a distribution with a percentage of houses falling into data bins predefined for each variable. What has been plotted on the following figures is the mean value for the variable in each county.

The Census data was the most detailed, giving data for each county in the US. The year houses were built in each county was given directly in the Census data and needed no additional manipulation for use in the model, but did not go beyond 1990. Figure 2 shows that houses in the Northeast and Midwest were built prior to 1950 for the most part, while houses in the Southeast and Southwest were mostly built after 1960.

**Figure 2: Average Year Houses Were Built in Each County**

Data were also gathered from the Residential Energy Consumption Survey (RECS), which was not available on as fine a scale as the census data. The smallest subdivision for this data was the nine census divisions. The regional divisions are visible in Figure 3 and Figure 4, showing dwelling height and the probability of floor leaks respectively. The Northeast and Midwest have the tallest dwellings whereas the shortest dwellings are located in the south and west.

The probability of floor leakage is based on foundation type, which was given in the RECS data in three categories: basement, crawlspace or slab on grade. Slab on grade foundations are assigned a leakage probability of 0 because there are no leakage pathways through the slab. Crawlspace foundations were assigned a leakage probability of 1 because there are numerous leakage pathways through penetrations in the floor so we can be fairly sure that there will be some air leakage through these pathways. Conditioned basements, like slab on grade foundations, are assigned a leakage value of 0. Unconditioned basements, like crawlspace, are assigned a leakage value of 1. Conditioned basements were not separated from unconditioned basements in the RECS data so we assumed half the basements were conditioned and half were unconditioned yielding a leakage probability of 0.5 for the basement foundation category.

The mean probability was calculated in each county, and is shown in Figure 4. The highest probability of floor leakage is found in the region consisting of Mississippi, Alabama, Tennessee and Kentucky showing that crawlspace foundations are popular there. The lowest probability of floor leakage is found in the region just next door, containing Texas, Louisiana, Oklahoma and Arkansas where slab on grade foundations are common.
Figure 3: Mean Dwelling Height in Meters by County

Figure 4: Probability of Floor Leakage by County
For the last two variables, fraction of households below an income threshold and dwelling floor area, it was necessary to combine Census data and RECS data. The income threshold was chosen at 125% of the poverty level because that was the low-income threshold defined in the Normalized Leakage model developed by McWilliams and Jung (2006). This threshold is the qualification criteria used by the the Ohio Weatherization Program, which provided all of the low-income data used to develop the model.

RECS data provided the number of households below the poverty level, between poverty and one and a half times the poverty level, and above one and a half times the poverty level for each census region. These data points were used to define a gaussian distribution for income level. Census 2000 data provided the fraction of households below poverty level for each county, and this value was used to shift the mean of the divisional income distribution for each county. It was assumed that the standard deviation of the income distribution for each county was the same as that for the census division, thus the fraction of households below 125% of the poverty level was inferred. Most of the Northeast and Midwest have a low fraction of households below this poverty threshold, as seen in Figure 5. High poverty is concentrated in Appalachia and the South and Southeast, with a few other isolated areas such as central South Dakota, the southern tip of Texas, and one county in eastern Arizona.

**Figure 5: Fraction of Households Below 125% of the Poverty Level**

Dwelling floor area was also determined from a combination of Census and RECS data. Census data provided a distribution of the number of rooms in each county. Average floor area per room was calculated for each census division from “total floorspace” and “number of rooms per house” given in RECS data.

Figure 6 shows that larger dwellings are concentrated in the Northeast and Midwest while smaller homes are concentrated in the West and Southwest. The Southeast has larger homes in Mississippi, Alabama and Tennessee and smaller homes from Appalachia to Florida. The estimate of floor area is a bit crude because it is based on the number of rooms, but it was the best we could do with the data available. It is unfortunate for such a poorly defined/known
parameter that NL has a strong dependence (16% decrease in NL for every 100m² increase in floor area) on floor area in the model developed by McWilliams and Jung (2006).

No data was available in either Census or RECS data regarding participation of households in an energy efficiency program. The Energy Star Homes webpage states that currently 10% of newly constructed homes participate in the Energy Star program, and that there are approximately 500,000 Energy Star homes existing in the United States. The total number of single family dwellings in the United States is 73.7 million according to the RECS. Since the energy program variable is used to describe not only houses that participated in new construction energy programs, but retrofit programs as well, we assumed that 1% of houses nationwide had participated in an energy efficiency program. This assumption could be fine tuned as more data becomes available.

**Figure 6: Dwelling Floor Area by County [m²]**

![Dwelling Floor Area by County](image)

**Discussion**

The input parameters were used to predict the ventilation for houses in each county. First, the empirical model is used to predict normalized leakage in each county. Figure 7 shows the normalized leakage across the United States. Immediately visible is the lower normalized leakage in the south east of the country and in Alaska. The boundary of this lower leakage area almost exactly follows the boundary of the humid climate zone. The climate coefficient of 0.35 for the humid climate is similar to the climate coefficient of 0.36 for Alaska. In contrast, the climate coefficients for the cold climate (0.53) and the dry climate (0.61) are much higher.

Another feature of this map that is initially apparent are the two areas of higher normalized leakage, one in northern Louisiana and the other in mid-southern Georgia. These pockets of leaky houses can be attributed to the fact that the houses in these areas are smaller, slightly older, and have a higher percentage of low-income residents than the surrounding counties. Households in Mississippi and Alabama are similar in income, but they are on the order of 50 square meters larger, which decreases NL by 8% according to the model. The higher probability of floor leaks in
Mississippi and Alabama increases the NL by 2-3%, which counteracts, but does not negate, the effect of their larger size.

**Figure 7: Normalized Leakage by County**

![Normalized Leakage by County](image)

The air exchange rate is calculated using the LBL Ventilation model, which predicts flow through the building envelope based on the leakage area of the house and the weather conditions in that location. The shielding class and terrain parameters were both assumed to be moderate, or class III. The leakage distribution parameters, X and R were set according to the floor leakage parameter. Houses with floor leakage were assumed to have one quarter of the leakage in the ceiling, one quarter of the leakage in the floor, and one half in the walls, as in ASHRAE Fundamentals, Chapter 27. Houses with no floor leakage have one third of the leakage in the ceiling and two thirds in the walls. Model predictions are only weakly sensitive to the values of X and R (0-15%, Reinhold and Sonderegger, 1983) so precise determination of these variables is not necessary.

The simulation was performed for each hour over a year of typical weather conditions, and the mean air change rate is shown in Figure 8. The same pattern can be seen as in Figure 7 with lower air change rate in the tight houses of the Southeast.

Ventilation efficiency was calculated using the Sherman-Wilson model. It is clearly visible that the milder climates on the west coast and in the south east have ventilation efficiencies closer to 1, indicating that the infiltration is close to steady state over the course of the year. Climates with low ventilation efficiency—as those in the mountainous regions and in the northern part of the country—have increased infiltration variation over the course of the year.
Figure 8: Yearly Average Natural Air Exchange Rate by County

Figure 9: Ventilation Efficiency by County
Effective ventilation shows how much ventilation is going towards reducing exposure for human occupants to pollutants. Figure 10 presents this data in air change rates and shows that Alaska and the Southeast have the lowest effective ventilation.

When effective ventilation is calculated for each month of the year, we find most of the country experiences the highest ventilation rate during the summer months. Some parts of the Midwest and areas on the Gulf of Mexico experience the highest ventilation in the spring, contrastingly the east coast of Florida and Seattle experience the highest ventilation in October. Only the interior of Alaska experiences the highest ventilation rate in the winter, although the maximum effective ventilation values are very similar in Alaska in summer and winter.

**Figure 10: Effective Ventilation by County**

---

**Summary and Conclusions**

This report has presented and used a statistical model for predicting the air tightness of any U.S. home based on location, age, size and configuration. While this model is expected to have an uncertainty of approximately 50% for an individual prediction it can be used on larger populations to predict regional and other trends.

The housing stock in the U.S. contains a negligible number of houses with mechanical ventilation systems therefore infiltration provides the ventilation in these houses. Our results indicate that the vast majority of the residential building stock has effective air change rateas above 0.35 air changer per hour and therefore gets sufficient ventilation from infiltrations when looked at on an annual basis.

Our analysis can help to select which regions may be particularly good candidates for saving energy through air tightening, such as through weatherization programs. The data contained herein has been used to estimate leakage trends, but could, in the future, be used to estimate potential energy savings.

Because there can be a substantial variation between individual houses, many tighter homes—including most new construction—will likely not be sufficiently ventilated by infiltration alone. In these cases both energy and IAQ gains can be made through a combination of air tightening and designed ventilation systems.
References


