AIR LEAKAGE OF US HOMES: REGRESSION ANALYSIS AND IMPROVEMENTS FROM RETROFIT

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

Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R3058
Berkeley, CA 94720, USA
* *Corresponding author: wrchan@lbl.gov*

ABSTRACT

LBNL Residential Diagnostics Database (ResDB) contains blower door measurements and other diagnostic test results of homes in United States. Of these, approximately 134,000 single-family detached homes have sufficient information for the analysis of air leakage in relation to a number of housing characteristics. We performed regression analysis to consider the correlation between normalized leakage and a number of explanatory variables: IECC climate zone, floor area, height, year built, foundation type, duct location, and other housing characteristics. The regression model explains 68% of the observed variability in normalized leakage. ResDB also contains the before and after retrofit air leakage measurements of approximately 23,000 homes that participated in weatherization assistant programs (WAPs) or residential energy efficiency programs. The two types of programs achieve rather similar reductions in normalized leakage: 30% for WAPs and 20% for other energy programs.

KEYWORDS

Blower door, fan pressurization measurements, air infiltration, weatherization, retrofit

INTRODUCTION

Residential energy efficiency and weatherization assistance programs (WAPs) have led to many measurements of air leakage being made in recent years. Building envelope airtightness is important because heating and cooling accounts for about 50% of the total energy consumption by US households [1]. Therefore, knowledge on the current state of the US housing stock, and factors that are associated with excessive air leakage, can have substantial energy implications.

In 2011, we collected a large number of air leakage measurements and updated the LBNL Residential Diagnostics Database (ResDB). Our latest efforts not only increased the number of data points, but also improved the spatial representation of the dataset. It is the goal of this regression analysis to identify housing characteristics that explain the observed variability in air leakage of single-family detached homes. In addition, we compared the air leakage measurements of homes before and after retrofit. Insulation upgrades and air sealing are commonly performed in a retrofit. In the US, the expected energy saving in heating and cooling bills from tightening the building envelope and reducing air infiltration is 10% to 20% [2]. But many factors can influence the energy savings and cost-effectiveness of air sealing and other retrofit measures, such as the initial air leakage of the house, and the expected improvement in airtightness from retrofit. This analysis will characterize the airtightness of the current US housing stock, and provide some of the needed data to evaluate the energy saving potential from reducing air infiltration via retrofit.

DATA DESCRIPTION

Data Sources

The newly updated ResDB contains air leakage data from 134,000 single-family detached homes. However, many missing data are present. The handling of these missing data, including year built, foundation type, and duct location, will be explained in greater details below. Overall, forty-three states are represented. The median year built and floor area is 1969 and 140 m^2, respectively.

Income-qualified WAPs are the major sources of data, accounting for about half of the blower door measurements. In prior versions of ResDB [3][4], Ohio was the only WAP present. ResDB now contains WAP data from eleven other states, including Arkansas, California, Iowa, Idaho, Minnesota, Montana, Pennsylvania, Utah, Virginia, Washington, and Wisconsin.

Residential energy efficiency programs are another major sources of data. For example, the Home Performance with ENERGY STAR program[1] is implemented in over 30 US states to improve energy efficiency of homes. New Jersey and Minnesota are the two states with the most pre- and post-retrofit blower door measurements in ResDB. There are also many data from programs in Vermont, Indiana, California, and Georgia.

Other sources that contributed air leakage and other diagnostic measurements include new homes that were tested to obtain an energy efficiency rating, or to demonstrate that they met an airtightness guideline. Homes are identified as energy efficiency rated according to the programs that collected the data, so there are likely some differences in rating criteria between the energy efficient homes. Moreover, there are also data that were collected for research studies or other purposes. Sources voluntarily contributed data to ResDB. Therefore, even though ResDB contains a large volume of data, the self-selected samples are not representative of the homes in US.

Normalized Leakage

Most of the air leakage data in ResDB are blower door measurements at 50 Pa pressure difference. Air leakage measurements are converted to normalized leakage (NL) for this analysis, as follows:

$$NL = 1000\left(\frac{ELA_{4\text{ Pa}}}{Area}\right)\left(\frac{H}{2.5\text{ m}}\right)^{0.3} \quad \text{where} \quad ELA_{4\text{ Pa}} = \sqrt{\frac{\rho}{2\times 4\text{ Pa}}}\left(Q_{50\text{ Pa}}\right)\left(\frac{4\text{ Pa}}{50\text{ Pa}}\right)^{n} \qquad (1)$$

$ELA_{4\text{ Pa}}$ (m^2) is the effective leakage area at 4 Pa, *Area* (m^2) is the dwelling floor area, H (m) is the dwelling height, ρ = 1.2 kg/m^3, and $Q_{50\text{ Pa}}$ (m^3/s) is the airflow rate at 50 Pa measured by the blower door. NL is roughly lognormal distributed, with a geometric mean of 0.61 and a geometric standard deviation of 2.5. ResDB contains 7,000 measurements of pressure exponent, n, which are used to compute NL when available. The distribution of n is roughly normal with a mean of 0.65, and a standard deviation of 0.06. n is assumed to be 0.65 for all other cases [5].

[1] http://www.energystar.gov/index.cfm?c=home_improvement.hm_improvement_index

If H is not provided in the data, we assumed 2.5 m for each story, and an additional 0.5 m for ground level and inter-floor framing. In some cases where both the number of story and house height are unknown, we assumed that houses <200 m^2 are single-story, and >200 m^2 are two-story. This simple allocation based on 200 m^2 as the reference point is the same as used in previous analyses of ResDB [3][4]. About 80% of single-story detached houses in US are <200 m^2, but only half of the multi-story detached houses are <200 m^2 [6]. Our method of using the house size to approximate number of story is reasonable, but it is a source of uncertain.

Multiple blower door measurements exist for some homes in ResDB. If additional tests were performed to verify a measurement, then the average value is used. If a house was tested under different configurations, then the one that best described the occupied condition is used, i.e., exclude attic, but include or exclude basement depending on the normal winter condition.

REGRESSION MODEL

The multivariate regression considers the relationship between NL and these housing characteristics:

- Floor area $Area$ (m^2)
- House height H (m)
- Year built $\overleftrightarrow{I_{year}}$: before 1960, 60–69, 70–79, 80–89, 90–99, 2000 and after
- IECC climate zones $\overleftrightarrow{I_{cz}}$: 12 categories
- Homes participated in WAP: $I_{LI} = 1$
- Homes rated for energy efficiency: $I_e = 1$
- Foundation type: I_{slab}, I_{floor1}, or I_{floor2}
- Duct location: I_{cond}, I_{duct1}, or I_{duct2}

Area and H are continuous variables, and all the remaining ones are indicator variables. Twelve of the 16 IECC climate zones[2] are represented: humid (5), dry (3), marine (2), and Alaska (2). The climate zone is determined by the house location, which is typically available by state and county, and the climate zone is identified correspondingly. For WAPs and other data with measurements before and after retrofit, the before values were used in the regression below. Homes are identified as energy efficiency rated by the programs that contributed the data.

Most of the data are missing foundation type and duct location. As a result, we first preformed the regression without these two parameters, as shown in Eq (2).

$$\ln(NL) = \beta_{area}Area + \beta_h H + \overleftrightarrow{\beta_{year}}\,\overleftrightarrow{I_{year}} + \beta_{LI}I_{LI} + \beta_e I_e + \overleftrightarrow{\beta_{cz}}\,\overleftrightarrow{I_{cz}} \qquad (2)$$

Using the coefficient estimates from Eq (2), the model residuals NL’ are computed as follows:

$$\ln(NL') = \ln(NL) - \left[\beta_{area}Area + \beta_h H + \overleftrightarrow{\beta_{year}}\,\overleftrightarrow{I_{year}} + \beta_{LI}I_{LI} + \beta_e I_e + \overleftrightarrow{\beta_{cz}}\,\overleftrightarrow{I_{cz}}\right] \qquad (3a)$$

We then considered the effects of foundation type and duct location on the model residuals to estimate their influence on NL. Only the data with known foundation type or duct location is considered in Eq (3b) and (3c) respectively, so the values of NL’ used for the regression are different in the two equations.

---

[2] See http://energycode.pnl.gov/EnergyCodeReqs/ for IECC climate zone classification.

$$\ln(NL') = \beta_{slab} I_{slab} + \beta_{floor1} I_{floor1} + \beta_{floor2} I_{floor2} \quad (3b)$$
$$\ln(NL') = \beta_{cond} I_{cond} + \beta_{duct1} I_{duct1} + \beta_{duct2} I_{duct2} \quad (3c)$$

From our previous work, we expected NL to be strongly correlated with year built [3][4]. To maximize the number of data considered in the regression, we categorized year built by decades from 1960 and onwards. But even when year built is treated as a categorical variable, one-quarter of the data are still missing this information. For these data, we imputed a year built category as follows. We first performed a regression by using three-quarters of the data with no missing data (i.e., year built is known). From this regression, we determined that ln(NL) decreases at an average rate of 0.14 from one year built category to the next newer category. Using this result, we imputed a year built category such that the predicted ln(NL) would best fit the measurements that contain missing data. The results are shown in Figure 1.

The imputation does not change the portion of homes in the different year built categories (Figure 1(a)). Homes that are built before 1960 and after 2000 remain the most common. This imputation method allows more data to be included in the regression model. Otherwise, homes in dry climate zones B-4, 5, and 6, and in marine climate zones C-3 and 4, would not be sufficiently represented in the regression.

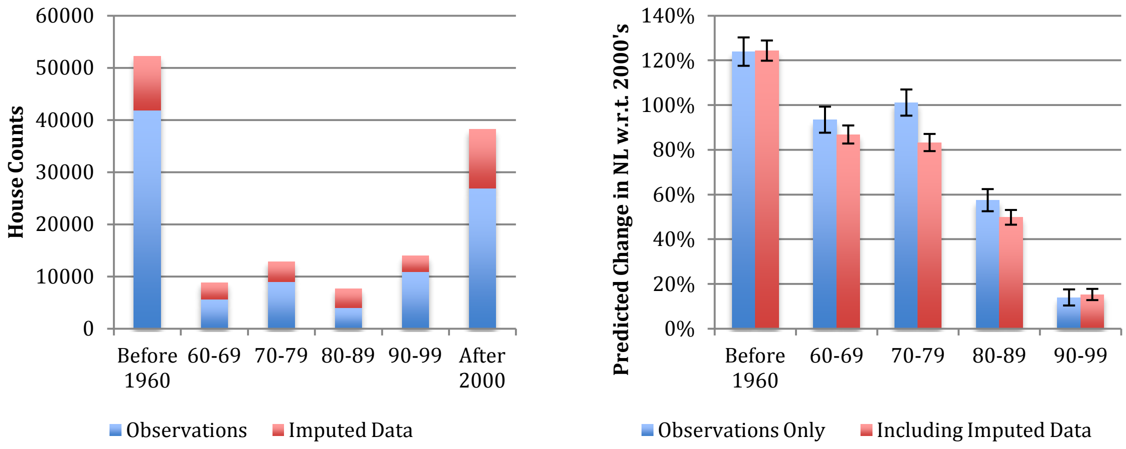


Figure 1 (a) Observed and imputed year built categories of single-family detached homes considered in the regression analysis. (b) Comparison of the predicted change in NL with respect to homes that are built in 2000's with and without imputation. Error bars show the 95% confidence interval.

Table 1 shows the regression results using the imputed data. Homes located in climate zone A-6,7 are selected as the reference, but other choices would give the same relative results. The model explains about 68% of the observed variability. The residuals ln(NL') are normally distributed: mean = 6.2e-17, and variance = 0.20.

One drawback of the imputation method used is that it can lead to underestimation of the differences between the observed and predicted values. In this case, however, the fit of model with (R^2=0.683) and without (R^2=0.682) the imputed data was essentially unchanged. With the imputed data, the predicted differences in NL for homes in different year built also remain roughly the same, as shown in Figure 1(b).

Figure 1(b) shows that homes built from more recent years have lower NL. This indicates that new homes were built with a more airtight building envelope compared to homes dated from earlier years. A recent study of new homes in California that are built between 2002 and 2004 also found that homes are built tighter compared to homes built in the 1980s and 1990s [7]. In addition to improvement in construction practices leading to tighter building envelope, it is also possible that there is a relationship between NL and house age. Older homes have higher NL not only because they were constructed that way, but also because the building envelope became more leaky over time. Both of these factors together likely explain a significant portion of the variability in NL among houses. Further analysis to isolate these two factors will be discussed in future analyses of ResDB.

Explanatory Variable	**Coefficient Estimates**	**Standard Error**	**Pr(>|t|)**	**95% Confidence Interval**
Area (m^2)	-0.00208	0.0000179	< 2e-16	-0.00211; -0.00204
Height (m)	0.064	0.00125	< 2e-16	0.061; 0.066
Year: Before 1960	-0.250	0.00705	< 2e-16	-0.264; -0.236
1960-69	-0.433	0.00811	< 2e-16	-0.449; -0.417
1970-79	-0.452	0.00762	< 2e-16	-0.467; -0.437
1980-89	-0.654	0.00836	< 2e-16	-0.670; -0.637
1990-99	-0.915	0.00816	< 2e-16	-0.931; -0.899
After 2000	-1.058	0.00748	< 2e-16	-1.073; -1.043
WAP Homes (pre-weatherization)	0.420	0.00428	< 2e-16	0.411; 0.428
Energy-Efficient Homes	-0.384	0.00453	< 2e-16	-0.393; -0.375
Humid A-1,2	0.473	0.01015	< 2e-16	0.453; 0.493
A-3	0.253	0.00653	< 2e-16	0.240; 0.266
A-4	0.326	0.00586	< 2e-16	0.315; 0.338
A-5	0.112	0.00551	< 2e-16	0.101; 0.123
A-6,7	0	--	--	--
Dry B-2,3	-0.038	0.00759	7.57e-07	-0.052; -0.023
B-4,5	-0.009	0.00684	2.00e-01	-0.022; 0.005
B-6	0.019	0.00988	4.91e-03	0.00008; 0.039
Marine C-3	0.048	0.01407	6.02e-04	0.021; 0.076
C-4	0.258	0.01133	< 2e-16	0.236; 0.281
Alaska AK-7	0.026	0.00589	1.42e-05	0.014; 0.037
AK-8	-0.512	0.00938	< 2e-16	-0.530; -0.439

Table 1 Results of regression model (β's in Eq. (2)) without considering foundation type and duct location.

All the coefficient estimates from the above regression are statistically significant at the 95% confidence interval, with the exception of climate zone B-4,5. This means that homes in climate zone B-4,5 tend to be less leaky than in the reference zone A-6,7, but the difference is small, and we cannot exclude the possibility that this apparent difference occurs only by chance in our data sample. We observed no effect on the overall model fit if homes B-4,5 and A-6,7 are grouped together or separately. Since these two climate areas are geographically far apart, for completeness we decided to keep all 12 climate zones in the model.

Foundation Type and Duct Location

For foundation type and duct location, we performed the regression analyses using a subset of the data, and assumed that the coefficient estimates also apply to the larger dataset. There are 12,500 houses with known foundation types: $I_{slab} = 1$ means house is built on slab, $I_{floor1} = 1$ means conditioned basement or unvented crawlspace, and $I_{floor2} = 1$ means unconditioned basement or vented crawlspace. These categories are chosen because after adjusting for the other parameters using Eq. (3a), homes with slab have the lowest NL, followed by homes where $I_{floor1} = 1$, and homes with $I_{floor2} = 1$ have the highest NL (Figure 2(a)).

Figure 2(b) shows a similar comparison but for duct locations using another subset of the data where this information is available. Homes with ducts located inside the conditioned space have the lowest NL, followed by homes with ducts located in the unconditioned attic or basement, and homes with ducts located in the vented crawlspace have the highest NL. However, the comparison by duct location is uncertain because it is based on very few data (526 houses).

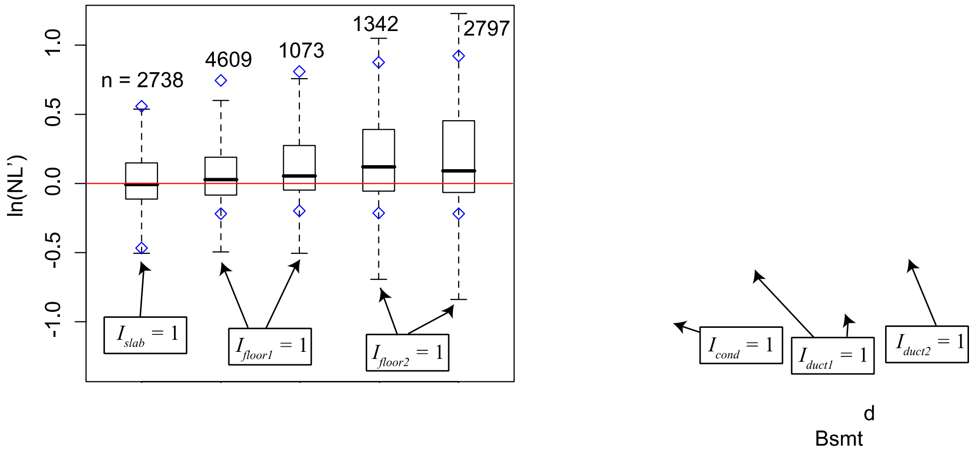


Figure 2 Model residuals of NL, computed using Eq. (3a), for homes with known (a) foundation type and (b) duct location (n = house counts). ln(NL') > 0 means that houses have NL higher than is predicted by Eq. (2).

Results of the regression (Table 2) show that the indicator variables considered are all statistically significant at 95% confidence interval. The coefficient estimates, β's, describe the influences of foundation type and duct location on NL as illustrated in the residual plots (Figure 2). Houses that are built on a slab and have ducts located inside the conditioned space tend to have the lowest NL. On the other hand, houses that have a vented crawlspace tend to have the highest NL, especially if the ducts are located in the crawlspace.

	Coefficient Estimates	**Standard Error**	**Pr(>|t|)**	**95% Confidence Interval**
(a) Foundation Type				
β_{slab}	-0.037	0.00709	1.85e-07	-0.051; -0.023
β_{floor1}	0.109	0.00492	< 2e-16	0.099; 0.118
β_{floor2}	0.180	0.00577	< 2e-16	0.169; 0.192
(b) Duct Location				
β_{cond}	-0.124	0.0255	1.53e-06	-0.174; -0.074
β_{duct1}	0.071	0.0339	3.59e-02	0.0047; 0.138
β_{duct2}	0.181	0.0383	2.98e-06	0.106; 0.256

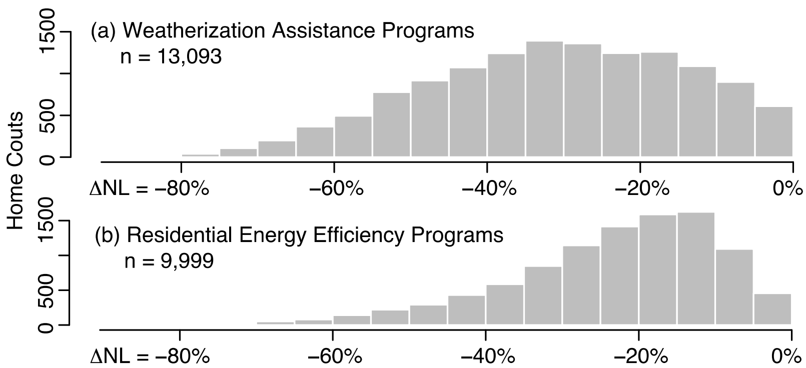
Table 2 Results of regression model considering the effects of (a) foundation type and (b) duct location.

RETROFIT IMPROVEMENTS

There are 23,000 houses with pre- and post-retrofit blower door measurements. Paired data that showed no improvements (462 homes) or increase in NL (449 homes) were excluded from this analysis. It is likely that those records reflect cases where retrofit did not include air sealing or other work that would reduce air leakage. In homes where NL increased, the percent change from the pre-retrofit measurement is <10% in half of the homes.

There are many differences in how WAPs and residential energy efficiency programs are implemented. WAPs use the minimum ventilation rate limit without mechanical ventilation (based on ASHRAE 62.2) as the target. The resulting savings-to-investment ratio must be greater than one for the work to qualified as allowable expenditures. On the other hand, energy efficiency programs, typically sponsored by utilities, tend to offer rebates and other financial incentives for homeowners to preform an energy audit, and to follow through with its recommendations.

Figure 3 shows a larger reduction in NL pre- and post-retrofit from WAPs overall, compared to the residential energy efficiency programs. When the two programs are considered together, the median ΔNL is -25%. Aside from differences in how the two types of programs are implemented, there are also other state-by-state differences in the kinds of retrofit measures performed, and how the air leakage measurements were collected and documented. As a result, there can be many explanations for the differences between the two programs.



Percent Change in NL:

$$\Delta NL = \left(\frac{NL_{post}}{NL_{pre}} - 1\right) \times 100\%$$

ΔNL	(a) WAPs	(b) E. Prog.
Median	-30%	-21%
25^{th} to 75^{th} Percentiles	-18 to -43%	-13 to -31%
5^{th} to 95^{th} Percentiles	-5 to -61%	-5 to -50%

Figure 3 Reduction in NL as a result of retrofit from (a) WAPs and (b) residential energy efficiency programs.

As shown by the regression model, WAP homes tend to have a higher NL pre-weatherization. This may be one of the reasons why WAPs appear to achieve a higher reduction in NL. It is easier to reduce obvious air leakage pathways that exist in leaky homes, than to make significant improvements in homes that are more airtight to begin with. To test this hypothesis, we considered the relationship between ΔNL and NL_{pre}, and also with other variables, including: climate zone, house dimensions, and year built. Regression analysis suggests that for WAPs, only NL_{pre}, floor area, and height are useful parameters in explaining ΔNL, but not climate zone or year built. However, this relationship does not hold for houses that participated in residential energy efficiency programs, where the regression analysis shows that none of the parameters considered are useful in explaining ΔNL.

RESULTS AND DISCUSSION

Figure 4 compares the potential influence of the various explanatory variables on NL predictions, including:

(a) Other climate zones compared with respect to A-6,7
(b) WAP homes versus non-WAP; homes rated for energy efficiency or not
(c) Floor area increased by 100 m^2; height increased by 2.5 m
(d) Other foundation types: conditioned basement or unvented crawlspace (*floor1*), or unconditioned basement or vented crawlspace (*floor2*), compared with respect to slab (*slab*)
(e) Duct located inside conditioned space (*cond*) or in unvented crawlspace (*duct2)* versus in unconditioned attic or basement (*duct1*)

The percent change in NL is computed using the coefficient estimates of the regression model, as shown in Table 1 and 2. For example, Figure 4(a) shows that houses in climate zone A-1,2 are 60% higher in NL than homes in A-6,7. This is computed by exp(0.473) – 1 = 0.6. The effects of year built are shown in Figure 1(b), and are not repeated here.

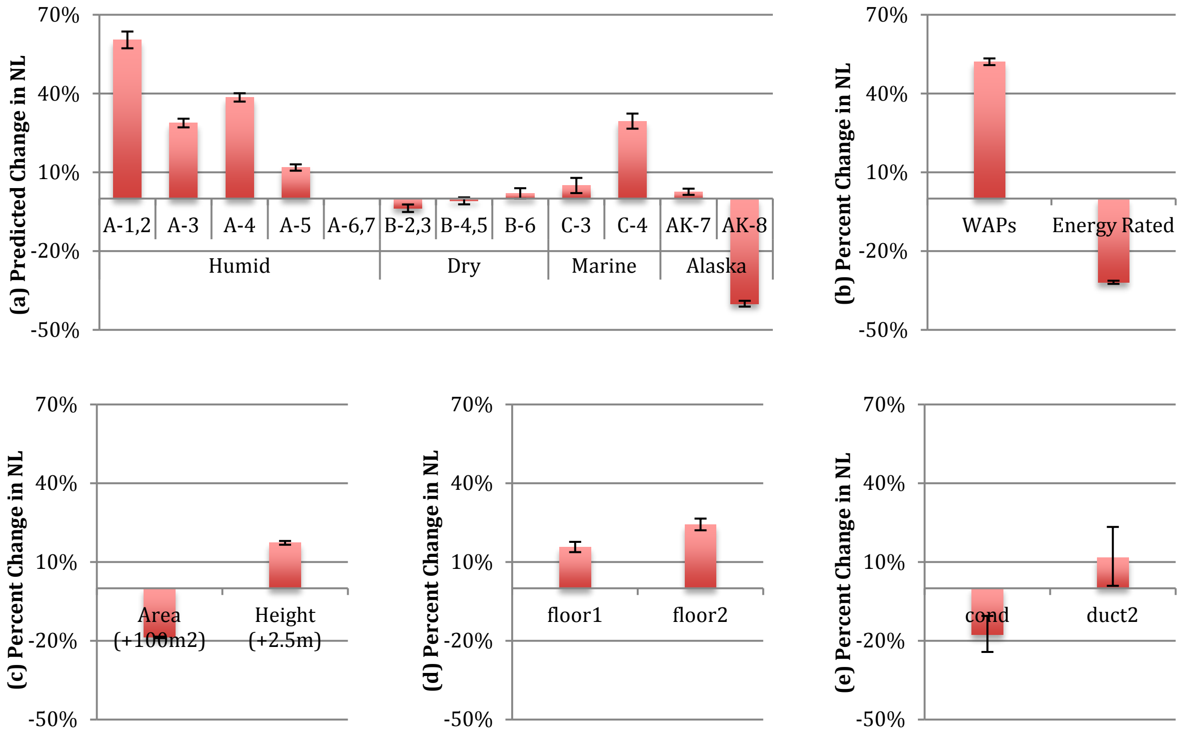


Figure 4 Predicted change in NL for homes in (a) difference climate zones with respect to climate zone A-6,7, and (b–e) other variables considered in the regression. Error bars show the 95% confidence interval.

Much of the variability observed in NL is associated with (a) climate zone, and (b) whether the houses are participants in WAPs or are energy efficiency rated homes. The difference in NL between the two extreme climate zones, A-1,2 and AK-8, is a factor of 2.7. The remaining factors, namely: (c) floor area and house height, (d) foundation type, and (e) duct location, each explain some differences in NL in the 10% to 20% range. In comparison, their importance is secondary for predicting NL. Overall, year built remains an important attribute to consider for predicting NL (see Figure 1(b)). The difference in NL between homes that are built before 1960 and after 2000 is a factor of 2.2.

The regression model presented here gives an estimate of NL based on a number of housing characteristics. For a housing stock, the model can explain 68% of the observed variability. However, the model is much more uncertain when it is applied to one house. This is because of the residual term. For example, the model predicts NL = 0.47 for a 150 m^2, single-story house built in 1990s that is located in climate zone C-3. The 95% confidence interval of this prediction is 0.44 to 0.50. However, the model residual ln(NL’) has a variance of 0.20. This means that there is only about 10% probability that a house with the exact characteristics will have NL between 0.44 and 0.50. For this one house, the model predicts there is a 95% probability that its NL is between 0.2 and 1.1. But, for many homes with the same characteristics, the regression model predicts with 95% confidence that the values of NL will likely center in between 0.44 and 0.50.

CONCLUSION

Many blower door measurements have been added to LBNL Residential Diagnostics Database from housing units across the US. Regression analyses were preformed on 134,00 single-family detached homes to describe the relationships between NL and house characteristics. By improving the spatial coverage of ResDB, more meaningful relationships were observed with climate zones. The predictive model explains about 68% of the observed variability, most of which are explained through year built, climate zone, and whether the houses are part of a WAP or energy efficiency rating program. Houses that are older, located in hot and humid areas of the US (climate zone A-1,2), and are occupied by households eligible for WAPs based on income are likely to have higher NL. Other characteristics that are associated with higher air leakage include houses with a vented crawlspace, and especially when ducts are located in the crawlspace as well. This information is useful for estimating the air leakage baseline of US homes, and can be used to target homes that would likely benefit the most from airtightness improvements to lower their energy costs.

Comparison of the before and after retrofit blower door measurements shows a reduction of NL in the 20% to 30% range. WAPs achieved somewhat higher reduction in NL than other residential energy efficiency programs, likely because WAP homes were more leaky pre-weatherization. The current data show comparably reduction in NL across all retrofit programs regardless of house location or year built. This is important because construction methods and practices vary greatly in the US. This analysis suggests that improvement in airtightness is possible across the US housing stock.

ACKNOWLEDGEMENTS

We greatly appreciate organizations and individuals who shared their blower door and other diagnostic data with us. This work was supported by the California Energy Commission Public Interest Energy Research Program award number CEC-500-07-006 and the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program, of the U.S. Department of Energy under contract No. DE-AC02-05CH11231.

REFERENCES

[1] RECS. 2005. Table US12. Total Consumption by Energy End Uses, 2005. Residential Energy Consumption Survey, U.S. Energy Information Administration. http://www.eia.gov/consumption/residential/data/2005/c&e/summary/pdf/tableus12.pdf

[2] Beachler M.C., Love, P.M. 2010. Retrofit Techniques and Technologies: Air Sealing, A Guide for Contractors to Share with Homeowners. Building America Best Practices Series, prepared for the U.S. Department of Energy. PNNL-19284. http://apps1.eere.energy.gov/buildings/publications/pdfs/building_america/ba_airsealing_report.pdf

[3] McWilliams, J., Jung, E. 2006. *Development of a Mathematical Air-Leakage Model from Measured Data*, LBNL-59041, Lawrence Berkeley National Lab, Berkeley, CA.

[4] Chan, W.R., Nazaroff, W.W., Price, P.N., Sohn, M.D., Gadgil, A.J., 2005. Analyzing a database of residential air leakage in the United States. *Atmospheric Environment* 39, 3445-3455.

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

[6] U.S. Census Bureau, Current Housing Reports, Series H150/09, American Housing Survey for the United States : 2009, U.S. Government Printing Office, Washington, D.C.

[7] Offermann, F.J. 2009. *Ventilation and Indoor Air Quality in New Homes*. California Air Resources Board and California Energy Commission, PIER Energy-Related Environmental Research Program. Collaborative Report. CEC-500-2009-085.