CS-BIGS 3(1): 37-47 http://www.bentley.edu/csbigs/documents/erinjeri2.pdf

Empirical Modeling of Air Tightness in Residential Homes in North Louisiana

Jinson J. Erinjeri
Louisiana Tech University, USA

Raja Nassar
Louisiana Tech University, USA

Myron Katz
Wisznia Associates, USA

Norman M. Witriol
Louisiana Tech University, USA

The measure of air-tightness in residential homes quantifies air leakage sites by measuring the airflow between the home and the outside in a range of pressures. Estimating air-tightness on the basis of certain physical characteristics of the home instead of actually measuring it is advantageous to homeowners as well as energy auditors. The objective of this study is to develop a region-specific empirical model to estimate air-tightness in residential homes. The air-tightness was measured for sixty-six homes in northern parts of Louisiana, USA. The three common air-tightness measures – CFM50, ELA, EqLA are used to develop three different multiple regression models based on the year of construction, floor area, number of bedrooms, and number of storeys of the homes. The predictive power for the three different models is also calculated. This case study is accessible to readers with a basic knowledge of statistics.

Keywords: *Air-tightness; CFM50; ELA; EqLA; Multiple Regression; Residential Buildings*

1. Background and Justification

Two key issues emerged post World War II in regards to construction of homes - healthy homes and tight homes. The healthy home refers to a building that is environmentally friendly, family safe, properly ventilated, and free from indoor pollutants. Tight construction refers to a home that is energy efficient, with an indoor environment well controlled through mechanical ventilation systems (Easley 2009). In reality, the construction of a good home involves a compromise between the above two issues. This article mainly deals with a study regarding the tightness of homes in northern Louisiana, assuming that homes have an efficient ventilation system. The purpose of this study is to develop an empirical model to estimate the air-tightness in residential houses in northern Louisiana without actually measuring air leakage rates.

Air-tightness quantifies the tendency of a home to allow air to flow through its pressure envelope in a range of pressures, typically between 4 and 50 Pascal (Energy Conservatory 2009). The air-tightness of buildings directly reflects air leakage sites, which include exterior doors, windows, foundations, electrical boxes and plumbing fixtures (Building Energy Codes Research Center 2009). Building air-tightness measurements are used for a variety of purposes such as (Energy Conservatory 2009):

1. Documenting the construction air-tightness of buildings
2. Estimating natural infiltration rates in houses
3. Measuring and documenting the effectiveness of air sealing activities
4. Measuring duct leakage in forced air distribution systems

There are a number of standardized formats for measuring air-tightness as described in the Minneapolis Blower Door Operation Manual. However, this study will focus on three of the commonly used formats namely Cubic Feet per Minute at 50 Pascal (CFM50), Effective Leakage Area (ELA) and Equivalent Leakage Area (EqLA). These variables are further described below (Energy Conservatory 2009).

CFM50: CFM50 is the airflow (in cubic feet per minute) through the blower door fan needed to create a change in building pressure of 50 Pa. It is the most common measure representing air-tightness.
ELA: ELA was developed by the Lawrence Berkeley Laboratory (LBL) and is defined as "the area of a special nozzle-shaped hole that would leak the same amount of air as a building does at a pressure of 4 Pa." ELA is most often expressed in square inches (sq. in.).
EqLA: The Equivalent Leakage Area (EqLA) is defined by Canadian researchers at the Canadian National Research Council as the area of a sharp edged orifice (a sharp round hole cut in a thin plate) that would leak the same amount of air as the building does at a pressure of 10 Pascal.

For air leakage to occur there must be both a hole or crack and a driving force (pressure difference) to push the air through the hole. The five most common driving forces, which operate in buildings, are stack effect, wind pressure, point source exhaust or supply devices, duct leakage to outside and door closure coupled with forced duct systems (Energy Conservatory 2009). Any of the above factors will lead to a pressure gradient between the home and the outside of a home. However, it is very difficult to quantify all of the above driving forces at the same time to come out with a fixed consistent air-tightness value. There are specialized devices which are used to measure the air-tightness of homes but these measurements are subject to change when the magnitude and direction of driving force change.

The Minneapolis Blower Door™, manufactured by the Energy Conservatory is a specialized tool used to measure air-tightness in residential buildings. The Blower Door fan blows air into or out of the building to create a pressure gradient between the inside and outside. This pressure gradient is used to measure the air-tightness in terms of volumetric units. Avoiding the use of the Blower Door to obtain air-tightness measurements is very beneficial to those who want a quick and reasonable estimate. This study develops an empirical model to estimate the air-tightness of residential buildings without actually performing a Blower Door test. This model estimates the air-tightness of a given house on the basis of physical information such as the year of construction, conditioned area, the number of storeys and the number of bedrooms. The estimate of the air-tightness obtained from the developed model is evaluated on houses of known air-tightness to check for its effectiveness. It is important to note that such a model will be applicable to northern Louisiana only, as we assume that the houses in this region have similar kind of building and environmental characteristics. We have not found such air-tightness models applied by any industry in a particular region. To our knowledge this region wise specific model is the first approach in this direction. The usefulness of the model developed lies in the fact that it reasonably estimates air-tightness in less time. Another advantage of this model is that the air-tightness can be estimated on the basis of physical information about the house without even visiting the house.

Environmentally related problems such as poor indoor air quality (IAQ) can have a significant impact on a building's value. Lower market value or a lease rent reduction are two likely scenarios that can occur once an unsolved IAQ problem becomes known or a building is tagged with "sick building syndrome" (Green 1995). Presently, in the United States, energy experts can review one's plans and conduct a Home Energy Rating to assess the energy efficiency of a home. The Home Energy Rating System (Energy Star 2009) has now become a nationally recognized system used to evaluate all the features of a home. These features include structure and foundation type, insulation levels, heating and cooling systems, air-tightness, windows, water heating equipment, and appliances. A home is rated on a 0-100 point system; with 100 being the most energy-efficient (a 100-point home would use zero energy). To meet Energy Star Home guidelines, a score of 86 should be achieved. A wide variety of simple options without significant costs can help improve a home's energy performance. A simple model like the one developed in this study will enable us to obtain reasonable estimates of air-tightness of residential buildings. This estimate of air-tightness can account for the 0-100 point system rating of a home.

Building tightness limits (BTLs) have been developed in some states in the United States. BTLs are guidelines based on estimates of the minimum air exchange rate of a building necessary to provide enough fresh air to maintain satisfactory health of the occupants and durability of the structure (Tsongas 1993). BTLs usually specify a building's minimum air leakage rate in CFM50 for comparison with the measured value of CFM50. For acceptable IAQ, the BTL standard requires that 15 CFM per person (assuming a minimum of five people) or 0.35 air changes per hour (ACH), whichever is greater, must be supplied by natural air leakage and/or continuously operating ventilation. Therefore, estimating air-tightness becomes important to determine the IAQ limits of homes.

Blower doors measure building tightness, and the natural infiltration rate of a house on the basis of a number of parameters. A single BTL does not incorporate factors like climate, a building's wind exposure, building size, or the number of occupants. Air exchange rates can vary widely depending on such factors. Max Sherman of the Lawrence Berkeley Laboratory has developed tables for each of the four climate zones in the United States (Tsongas 1995). The tables include a United States map divided into four climate zones and account for the number of occupants, the number of storeys of the building, and wind shielding characteristics. Weatherization personnel can use the map to find their particular zone and then select the appropriate table with the correct CFM50 minimum values. However, the simple model presented in this paper does not take into consideration all of the factors discussed for determining air-tightness. The main objective of this empirical study is to give reasonable estimates of air-tightness with minimum inputs. The easily obtainable characteristics of a residential building are taken as the independent variables. The model developed is based on the assumption that the houses in a particular region have similar kind of building and environmental characteristics. This assumption is basically considered to suppress the influence of climatic factors and other related factors attributed to a particular region – in this case North Louisiana. Also, the variables considered in the model are obtainable without even visiting a home. However, if precise and accurate measurements are required, then this model may or may not be best suited. Details regarding the variables considered in this model are presented in Sections 3 and 4 of this article.

2. Sampling and Data Collection

The primary data were collected while performing Blower Door tests at 66 houses in and around Ruston, North Louisiana, USA. The dataset of 66 houses was split into two parts to perform cross-validation. Cross validation is a validation technique where the dataset is split into model building and prediction sets (Kutner et al. 2004). The first part comprising 46 homes formed the model building set or the estimation sample whereas the remaining 20 homes formed the prediction set or the validation sample. The complete set of data is presented in Table 9 (Appendix) of this article. The details of the validation process are described in Section 7 of this article.

The details of the testing procedure using a Blower Door are as follows:

Step 1: Calculate the floor area and the volume of the home.
Step 2: Set control on pilot for all combustion appliances.
Step 3: Turn off the air handler of the HVAC unit and remove the filter. Turn off attic fans, dryer and other exhaust fans.
Step 4: Attach the blower door to an exterior doorframe - selecting one which provides a clear airflow path to outside.
Step 5: Prepare the Automated Performance Testing System (APT) measuring equipment for testing in depressurized mode.
Step 6: Launch the TECTITE™ software and run the process.

In default mode of the TECTITE™ software, 100 data points are collected at the beginning and end of the test for each set pressure difference between the home and outside (50, 45, 40, 35, 30, 25, 20 and 15 Pascal). The output of this entire process gives air-tightness measures such as CFM50, ACH50, Effective Leakage Area (ELA), and Equivalent Leakage Area (EqLA). However, it is important to note that this study only considers the three most common measures of air-tightness - CFM50, ELA and EqLA. For more details regarding the testing procedure, see Energy Conservatory 2009, and Witriol et al. 2003.

3. Response Variables - CFM50, ELA and EqLA

In this study, three different models were fitted to the three measures of air-tightness CFM50, ELA, and EqLA. The three measures of air-tightness or dependent variables – CFM50, ELA and EqLA were compared against each other and the plot between these three variables are displayed in Figures 1, 2 and 3. From Figure 1, we can see that as ELA increases CFM50 also increases. The variability in the relationship between ELA and CFM50 widens beyond the 250 sq. in. (0.16 m^2) and 5000 CFM (2.36 m^3s^{-1}) mark.

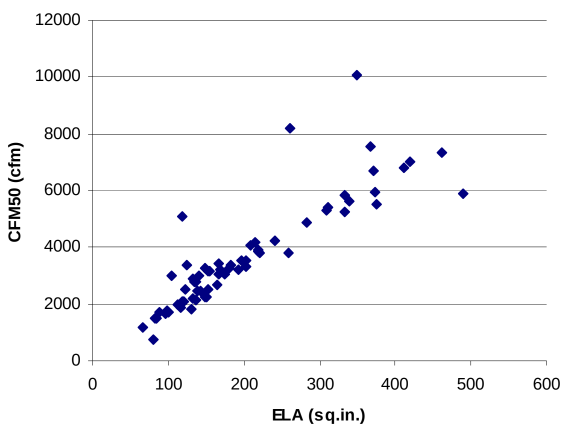

Figure 1. Plot of CFM50 vs. ELA.

The correlation coefficient between CFM50 and ELA was determined to be 0.88, which indicates a high correlation. However, this high correlation need not indicate the agreement between them. The relationship between ELA and CFM50 (State of California 2009) is given by:

$$ELA = 0.055 \times CFM50 \quad (1)$$

Eq.1 is an empirical relationship and the value of 0.055 is often questionable. The data obtained by our study were compared against the value of 0.055 by calculating the ratio of ELA to CFM50. This study on 66 homes obtained a mean of 0.056 with a standard deviation of 0.011. This suggests that there is some kind of relationship between CFM50 and ELA but it is to be noted that the factor of 0.055 varies from home to home.

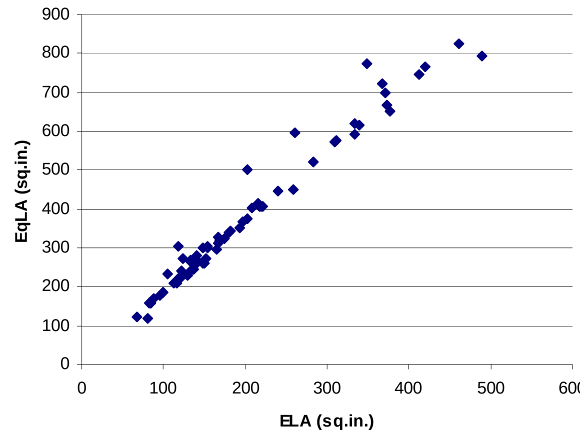

Figure 2. Plot of EqLA vs. ELA.

From the plot of EqLA vs. ELA in Figure 2, it can be seen that there is a linear relationship between the two variables with a high correlation of 0.98. The average ratio of EqLA to ELA was determined to be 1.896 with a standard deviation of 0.168. This ratio of 1.896 is in accordance with the Blower Door Manual (Energy Conservatory 2009), which states that the calculated EqLA will typically be about twice as large as the ELA.

Figure 3 displays the plot of EqLA vs. CFM50 and it is observed that as EqLA increases CFM50 also increases. Note that the variability in the relationship between ELA and CFM50 widens beyond the 450 sq. in. (0.29 m^2) and 4000 CFM mark (1.89 m^3s^{-1}).

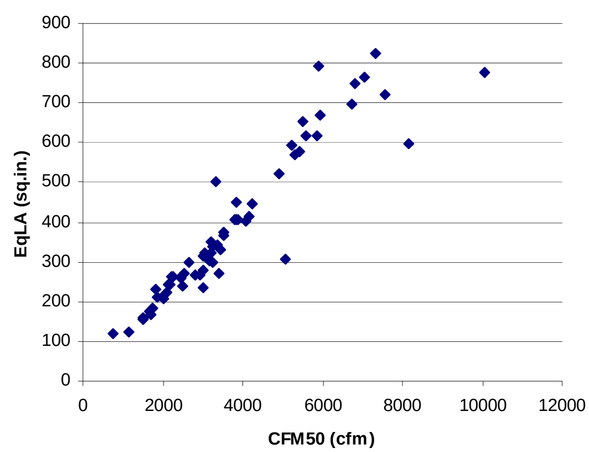

Figure 3. Plot of EqLA vs. CFM50.

The correlation coefficient between CFM50 and EqLA was determined to be 0.94, which indicates a high correlation. The average of ratio of EqLA to CFM50 was determined to be 0.105 with a standard deviation of

0.015. This value of 0.105 is approximately twice the average of ratio of ELA vs. CFM50; which corresponds to the fact that the measured value of EqLA is approximately twice that of ELA.

4. Model Building Process

A sample of 46 observations (estimation sample) from the total of 66 observations was used in the model building process. To build the empirical model for determining the air-tightness based on physical information, the multiple linear regression technique was applied. Three regression models of the form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_{p-1} x_{p-1} + e \qquad (2)$$

were developed, where
Y = air-tightness (CFM50 or ELA or EqLA)
X_i = home parameters, i = 1,2,……p-1.

In the above model, Y (CFM50/ELA/EqLA) is the dependent/response variable whereas house parameters constitute the independent/predictor variables. The house parameters considered in the model building process were Year of Construction/ Year Built (YB), Floor Area (FA), Number of Storeys (NOS), and Number of Bedrooms (NOB). These independent variables selected are easily available for a given home in a given region.

Table 1 shows the details of the predictor variables with their respective range and units for the total sample size of 66 homes. For example, Year Built indicates houses built between 1920 and 2004 with 1, 1.5, 2 storeys and a number of bedrooms ranging between 2 to 5. To allow for a more flexible specification of the relationship between the dependent variable and the variables Number of Storeys (NOS) and Number of Bedrooms (NOB), those variables were considered as categorical variables. Therefore, dummy variables (indicator variables) were introduced into the above model building process to categorize the values as 0 and 1. Tables 2 and 3 show the levels of the indicator variables for NOS and NOB.

The house parameter NOS represented by X_3 in Table 1 was categorized as X_{31} and X_{32} as shown in Table 2. The house parameter NOB (X_4) was similarly categorized as X_{41} and X_{42} as shown in Table 3. The response for the above model was the air-tightness measure of CFM50 or ELA or EqLA. The independent variables considered in Tables 1, 2 and 3 were used for regressing CFM50, ELA and EqLA to generate three different models.

Table 1. Predictor Details.

Predictors	Units	Range	House parameters
Year Built	Number	1920 - 2004	X_1
Area	Square Feet	1041 - 3866	X_2
Number of storeys	Number	1, 1.5 or 2	X_3
Number of bedrooms	Number	2, 3, 4 or 5	X_4

Table 2. Levels of Indicator Variables for the Number of Storeys.

X_{31}	X_{32}	
1	0	If the observation is from storey 1
0	1	If the observation is from storey 1.5
0	0	If the observation is from storey 2

The initial regression model was modeled using the 46 observations from the estimation sample including the following variables:

Dependent variable: ELA/CFM50/EqLA
Independent variables: YB, FA, NOS (X_{31} and X_{32} are the indicator variables), NOB (X_{41} and X_{42} are the indicator variables)

Table 3. Levels of Indicator Variables for the Number of Bedrooms.

X_{41}	X_{42}	
1	0	Home has 2 bedrooms
0	1	Home has 3 bedrooms
0	0	Home has 4 or more bedrooms

The model developed with CFM50 as the dependent variable will be termed CFM50 and the one developed with ELA will be termed ELA and so forth in the entire paper unless otherwise stated. The regression models for ELA, CFM50 and EqLA were constructed with the help of the SAS™ software. The presence of multicolinearity in each of the three models was checked by determining the variance inflation factors (VIF). The VIF's measure how much the variances of the estimated regression coefficients are inflated as compared to when the independent variables are not linearly related (Montgomery et al. 2001). A maximum VIF value in excess of 10 is often taken as an indication that multicolinearity may be unduly influencing the least squares estimates. However, in this study, the three models developed did not show any multicolinearity effects.

To know which variables will contribute significantly to the model and to minimize the effects of multicolinearity (if present), the Stepwise selection method was employed. Stepwise regression is a modification of forward selection in which at each step all regressors entered into the model previously are reassessed via their partial F-statistics (Montgomery et al. 2001). Stepwise regression requires two cutoff values p_{IN} and p_{OUT} as the p-values of the corresponding F statistics F_{IN} and F_{OUT} for adding and removing a predictor. In most applications, we choose $F_{IN} > F_{OUT}$ (so that $p_{IN} < p_{OUT}$) and in our case, we apply $p_{IN} = 0.05$ and $p_{OUT} = 0.10$. Applying the Stepwise regression technique to the three models (CFM50, ELA and EqLA), the results show that the variables Year Built, Area and X_{31} were significant for all three models. It is important to note the VIF for all the three models is much less than 10 indicating the absence of multicolinearity. The results of the stepwise selection process are summarized in Tables 4, 5 and 6.

From Table 4, we can see that the included predictor variables in the model for ELA are Area, YB and X_{31}. Therefore, incorporating the parameter estimates in the general regression Eq. 2, we obtain the estimated response ELA as

$$\textit{Estimated } Y_{ELA} = 5217.08 - 2.63x_1 + 0.10x_2 - 57.52x_{31} \quad (3)$$

Table 4. Summary of Stepwise Regression for ELA.

Step	Variable	Partial R^2	R^2	C(p)	F	Pr>F
1	Area	0.491	0.491	41.60	42.52	<.0001
2	YearBuilt	0.183	0.674	13.55	24.13	<.0001
3	X_{31}	0.054	0.729	6.61	8.41	0.0059

Table 5. Summary of Stepwise Regression for CFM50.

Step	Variable	Partial R^2	R^2	C(p)	F	Pr>F
1	Area	0.445	0.445	42.69	35.27	<.0001
2	YearBuilt	0.264	0.709	4.48	38.88	<.0001
3	X_{31}	0.033	0.742	1.38	5.44	0.0246

From Table 5, we can see that the included predictor variables in the CFM50 model are YB, Area and X_{31}. Therefore the regression equation for the CFM50 model becomes

$$\textit{Estimated } Y_{CFM50} = 122899 - 62.04x_1 + 1.95x_2 - 914.29x_{31} \quad (4)$$

The summary of stepwise regression for EqLA as the response variable is presented in Table 6. The significant predictor variables in the EqLA model are YB, Area and X_{31}.

Table 6. Summary of Stepwise Regression for EqLA.

Step	Variable	Partial R^2	R^2	C(p)	F	Pr>F
1	Area	0.503	0.503	49.79	44.52	<.0001
2	YearBuilt	0.223	0.726	10.66	34.92	<.0001
3	X_{31}	0.048	0.774	3.75	8.96	0.0046

Incorporating the parameter values, the regression equation for the response EqLA becomes

$$\textit{Estimated } Y_{EqLA} = 10732 - 5.40x_1 + 0.19x_2 - 102.40x_{31} \quad (5)$$

All three models have the same predictor variables significant as seen from Eqs. 3, 4, 5. In addition, the parameters also have the same sign indicating that the variables have the similar effect on the response variables in all models.

In general, all three models have a negative sign for the predictor variable –Year Built. This suggests that as the age of the house increases, the values of air-tightness decreases. The positive sign of the parameter - Area in all three models indicate that as the conditioned area increases the air-tightness value increases. It is important to note that the higher the value of air-tightness, the more leaky the house is. Due to sample limitations on the number of storeys/bedrooms and to avoid the negative consequences of including the independent variables number of storeys and number of bedrooms, regressions considering only year built and area of home as independent variables were considered. The revised air-tightness models without NOS and NOB are as follows:

$$\textit{Estimated } Y_{ELA} = 4726.77 - 2.41x_1 + 0.11x_2 \quad (6)$$

$$\textit{Estimated } Y_{CFM50} = 115105 - 58.64x_1 + 2.14x_2 \quad (7)$$

$$\textit{Estimated } Y_{EqLA} = 9859.40 - 5.02x_1 + 0.21x_2 \quad (8)$$

Eqs. (6), (7) and (8) are the three air-tightness models involving only area and the year as independent variables. All analyses from this point deal with these three equations (Eqs. 6, 7, 8) unless otherwise stated.

5. Aptness of the Regression Model

The regression models for CFM50, ELA and EqLA (Eqs. (6), (7) and (8)) were checked for aptness in order to verify the major assumptions behind the regression analysis. The major assumptions behind the regression analysis are (Montgomery et al. 2001):

1. The relationship between the response and regressors is linear, at least approximately.
2. The error terms have constant variance (homoskedasticity).
3. The errors are normally distributed.
4. The error term has zero mean.
5. The error terms are independent.

5.1 Linearity

To review the relationship between the dependent variable and each of the independent variables, plots were generated with each of the dependent variables (CFM50, ELA and EqLA) against each of the independent variables. The plots did not display any non-linear characteristics. The residual plots, i.e. the plot of residuals vs. predicted for the three dependent variables were also plotted and are presented in Figures 4, 5 and 6. Review of the plots showed that residuals are not displaying much of systematic tendencies or trends. Therefore, no transformations were performed on these data. The interpretation of outliers, if present, is discussed in detail in Section 6.

5.2 Non-constant Error Variance

The plots of the residuals against the predicted values as described above were again used to test for a non-constant error variance. The plots display some heteroskedasticity, which could be due to the presence of some outliers.

5.3 Normality of Error Terms

A significant departure from normality is a serious violation of the assumptions in regression. A simple method of checking the normality assumption is to construct a normal probability plot of the residuals (Montgomery et al. 2001). The points on Figures 7, 8 and 9 are expected to fall approximately along a straight line if the normality assumption is satisfied. The normal probability plots in Figures 7, 8 and 9 show that the resulting points lie approximately on a straight line, except possibly for some outliers.

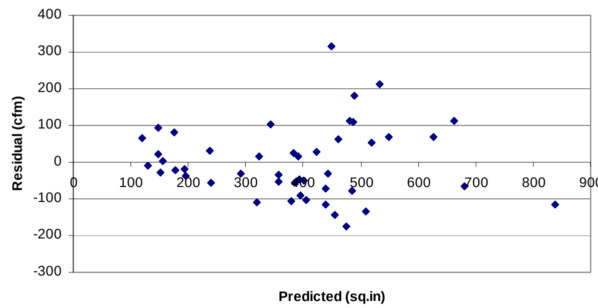

Figure 4. Plot of Residuals vs. Predicted for ELA.

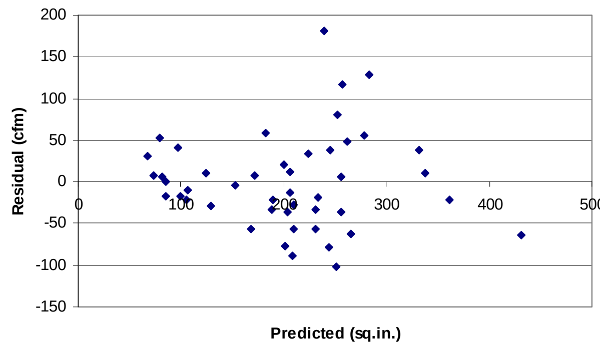

Figure 5. Plot of Residuals vs. Predicted for EqLA.

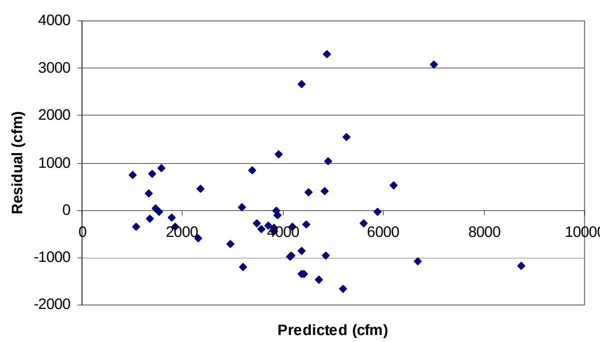

Figure 6. Plot of Residuals vs. Predicted for CFM50.

5.4 Independence of Error terms

A regression model requires independence of the error terms. Again, a residual plot can be used to check this assumption. The independence of errors is verified by plotting residuals against predicted values. A random, pattern-less lot implies independent errors. The plots of the residuals vs. predicted values for the three models did not display any dependence among error terms.

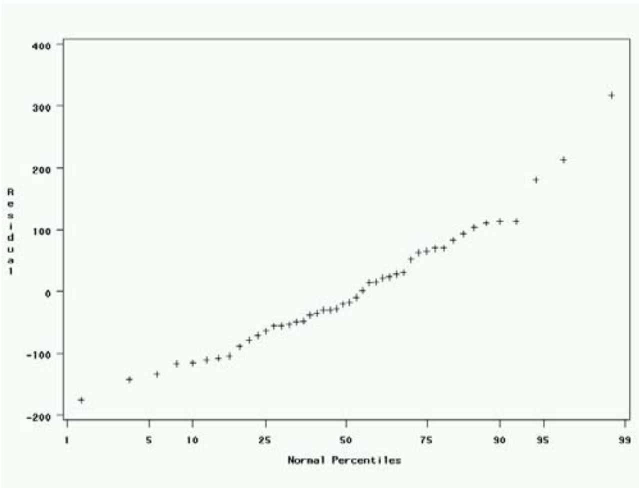

Figure 7. Normal plot for error terms of CFM50.

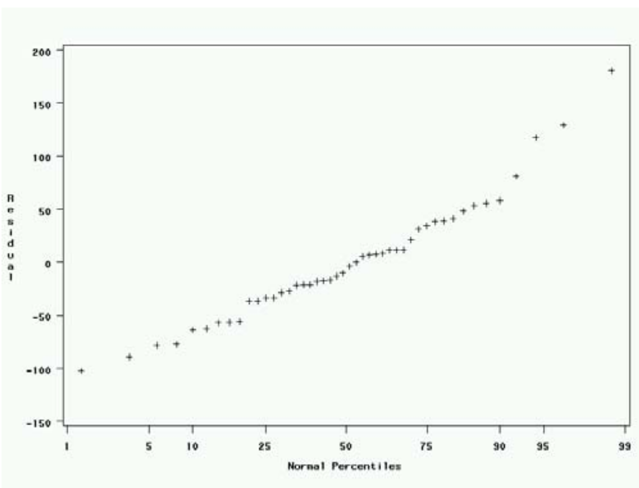

Figure 8. Normal plot for error terms of ELA.

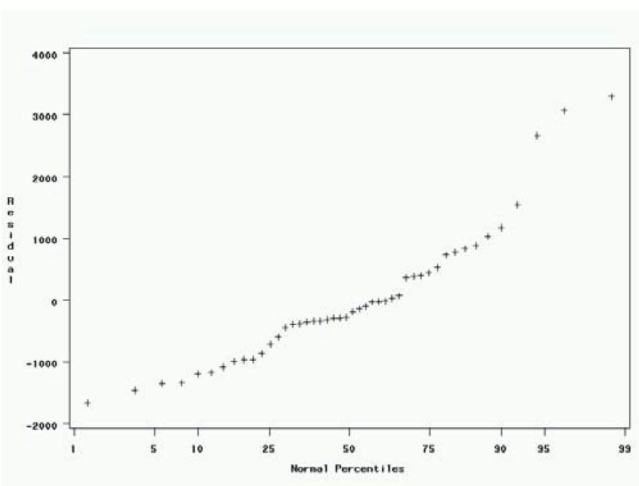

Figure 9. Normal plot for error terms of EqLA.

6. Outliers

6.1 X-Outliers

Leverage values (diagonal element h_{ii} in the hat matrix) greater than 2p/n are considered to be outlying cases. The target value is 2x(3)/46 = 0.13 for all three models - ELA, EqLA and CFM50. A comparison of h_{ii} to the target value detected observations 1, 4, 11, 14, 35 and 36 as X outliers in the case of CFM50, ELA and EqLA. These observations need to be examined to determine if they are really influential or not.

6.2 Y-Outliers

To identify outlying Y observations, an examination of the studentized deleted residuals (d_i) for large absolute values and the appropriate t distribution is necessary. Taking an alpha of 0.10, t_{tab} = (0.9998, 42) = 2.970. A comparison of d_i^* to t_{tab} for the case of EqLA and ELA, we find observation 22 as the outlying Y observation where as 37 as the outlying Y observation in the case of CFM50.

6.3 Influence of Outliers - DFFITS

To identify the influence of observations identified as X and Y outliers, the measures DFFITS were used. An observation is considered influential if the absolute value of DFFITS exceeds twice the square root of p/n for large datasets. The target value calculated was 0.51. Observations 11, 22, 35 and 37 are found to be influential in the case of CFM50 whereas 4, 7, and 22 were found to be influential in the case of ELA. In the case of EqLA, observations 4, 7, 35 and 22 were found to be influential.

An examination of the data associated with the observations listed in Sections 6.1, 6.2, 6.3 did not reveal any typographical errors or miscalculations and therefore all the observations were retained. The regression equations of CFM50, ELA and EqLA were maintained as obtained before and the predictive power of these three models was determined. A discussion of the predictive power of the three models is given in Section 7.

7. Model Validation and Predictive Power

The air-tightness model for EqLA, CFM50, and ELA were given in Eqs. (6), (7) and (8) respectively. The final step in a model-building process is the validation of the above selected regression models. Model validation involves checking a candidate model against independent data. For this study, we employed the preferred method of data splitting (Kutner et al. 2004). The first set called the model-building set was used to develop the model. We refer to this dataset as the estimation sample. The second dataset, called the validation set or the validation sample was used to determine the predictive ability of the selected model. Splits of the data are performed randomly (Kutner et al. 2004). However, it is important that the model-building data should be large enough to obtain a reliable model. In this case, we have 46 observations in the estimation sample and 20 observations in the prediction sample with a total of 66 observations. Predictive capability can be determined by calculating the mean of the squared prediction errors (MSPR) as follows:

$$MSPR = \frac{\sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2}{n^*} \quad (9)$$

where, Y_i is the value of the response variable in the i^{th} validation case, $\hat{Y}_i$ is the predicted value of the i^{th} validation case based on the model building data set, and n^* is the number of cases in the validation data set.

If the MSPR is fairly close to the error mean square (MSE) based on the regression fit to the model building data set, then the MSE for the selected regression model is not seriously biased and gives an appropriate indication of the predictive ability of the model. The validation results are presented in Table 7.

From Table 7, we can conclude that the MSPR of CFM50 is not very far from the MSE of the model building data set, whereas the MSPRs of ELA and EqLA are greater than twice those of the MSEs of their respective model building datasets. These results suggest that the predictive ability of these two models may not be high. Table 8 further examines the predictive power of the three models using Theil's statistic, the Root-Mean-Square Percent Error (RMSPE) and Mean Absolute Percent Error (MAPE) (SAS 2009, Cattin 1990, Makridakis et al. 1998).

Theil's Statistic is given by:

$$U = \sqrt{\sum_{i=1}^{n^*} (Y_i - \hat{Y}_i)^2) / \sum_i Y_i^2} \quad (10)$$

RMSPE is defined as:

$$RMSPE = \frac{1}{n^*} \sum_{i=1}^{n^*} (\frac{Y_i - \hat{Y}_i}{\hat{Y}_i})^2 \quad (11)$$

MAPE is defined as:

$$MAPE = \frac{1}{n^*} \sum_{i=1}^{n^*} (\frac{|Y_i - \hat{Y}_i|}{Y_i}) \quad (12)$$

In Equations 10, 11, and 12, Y_i and $\hat{Y}_i$ represent the actual and the predicted response values respectively for the i^{th} observation.

Table 7. Validation Results.

	MSPR	MSE	β_1 (95% confidence limits)	β_2 (95% confidence limits)
Estimation sample				
CFM50	-	1252981	-77.61 -39.67	1.61 2.67
ELA	-	3405.49	-3.40 -1.42	0.08 0.14
EqLA	-	10240	-6.74 -3.31	0.16 0.26
Validation sample				
CFM50	1218783	1433863	-113.41 -32.28	-0.37 2.95
ELA	8535.04	10041	-8.20 -1.41	-0.09 0.19
EqLA	23294.5	27405	-13.91 -2.70	-0.12 0.34

Table 8. Predictive Power of Air-tightness Models.

Statistic	CFM50	EqLA	ELA
Theil's U	0.3055	0.3720	0.4077
RMSPE	0.0042	0.0060	0.0068
MAPE	0.2570	0.3002	0.3107

From Table 8, we can conclude that the CFM50 based model has a better predictive power than ELA and EqLA since all the three measures show a lower value.

8. Conclusions and Recommendations

Three models (CFM50, ELA and EqLA) to determine air-tightness based on the age of the home and conditioned area were developed using multiple regression analysis (Eqs. (6), (7) and (8)). The parameter estimates from these three models shows that as the age of the house increases, the air leakage increases in Louisiana homes. The conditioned area also shows a similar trend with respect to air leakage. The CFM50 model tends to have a better predictive power than ELA and EqLA based on the predictive-power measures.

The proposed model will be very beneficial to those who are involved with building science, especially those who want quick and reasonable estimates of air-tightness. In addition, this region-specific model can be used when homes are remodeled to quickly estimate the air-tightness of buildings - mainly for energy related calculations.

The model will be advantageous to energy raters as well as those involved in real estate. In the real estate business, this model will serve the buyer/seller as an initial tool to approximately estimate the air-tightness of a given home. This measure can lead to further evaluations or price adjustments on a given home.

The main leakage sites in buildings are exterior doors, windows, foundations, electrical boxes and plumbing fixtures (Building Energy Codes Resource Center 2009). Therefore, to enhance the model we could include variables such as the number of windows and number of exterior doors. However, this inclusion might limit the usefulness of the model, as physical presence at a home might be needed to collect data. Such a model could be developed in this regard to make a comparison with the model developed in this article.

Future models should include a variable to account for homes which are rehabilitated. Air-tightness estimates for rehabilitated homes based on the current model may not be reasonable since we consider the age of the house as a significant factor.

Acknowledgments

We gratefully acknowledge the sponsorship of the Louisiana Department of Natural Resources under DNR Interagency agreement No. PVE 29-01-12. During the course of this research project, the research team at Louisiana Tech University received valuable support and guidance from the DNR energy section staff, Harvey Landry, Buddy Justice, Paula Ridgeway and especially Wade Byrd. The authors are grateful to Ray Sterling for his support throughout the research. The assistance of the TTC-Physics staff at Louisiana Tech University Sandra Perry and Lory Gray is greatly appreciated.

REFERENCES

Areas of Air leakage, Building Energy Codes Resource Center. 2009. http://resourcecenter.pnl.gov /cocoon/morf/ResourceCenter/article/1385 (Accessed July 2009).

California. 1993. Residential indoor air quality ventilation, available at http://www.energy.ca.gov/title24 /2008standards/prerulemaking/documents/2006-07-12_workshop/reviewdocs/RESIDENTIAL_VENTILATION_STUDY.PDF (Accessed July 2009).

Cattin P. 1980. Estimation of the Predictive Power of a Regression Model, *Journal of Applied Psychology*, 65(4): 407-414.

Easley S. 2009. Tips for managing moisture and air in your home. http://www.moldknowledge.com/house-wrap.htm (Accessed July 2009).

The Energy Conservatory. 1993. Minneapolis Blower Door Operation Manual, Minneapolis, MN. www.EnergyConservatory.com (Accessed July 2009).

Energy Star. 2009. What is a HERS Rating? Available at http://www.energystar.gov/index.cfm?c=bldrs_lenders_raters.nh_HERS (Accessed July 2009).

Green RR. 1995. "Clearing the air", *Commercial Investment Real Estate Journal*, 14(4): 3-15.

Kutner MH, Nachtsheim CJ, Neter J, and Li W. 2004. *Applied Linear Statistical Models (5th edition)*, McGraw-Hill.

Makridakis S., Wheelwright S.C., and Hyndman R.J. 1998. *Forecasting Methods and Applications, (3rd edition)*, John Wiley and Sons.

Montgomery D.C., Peck E.A., and Vining G.G. 2001. *Introduction to Linear Regression Analysis, (3rd edition)*, John Wiley and Sons.

SAS Institute. 1999. OnlineDoc, version 8, SAS Institute, NC. http://v8doc.sas.com/sashtml/ (Accessed July 2009).

Tsongas G. 1993. "Building Tightness Guidelines: When is a house too tight?", *Home Energy*, 10(2), available at http://homeenergy.org/archive/hem.dis.anl.gov/eehem/93/930309.html (Accessed July 2009).

Witriol N, Erinjeri J, Katz M, McKim R, and Nassar R. 2003. Testing HVAC Duct Leakage in Existing Residential Buildings in North Louisiana, Trenchless Technology Center, Louisiana Tech University, Ruston, LA.

Appendix

Table 9. Air-tightness data for 66 homes.

Test #	Year Built	Area	Volume	ELA	CFM50	NOS	NOB	Eq.la
		sq.ft.	cu.ft.	sq. in.	CFM			sq. in.
1	1920	1445.0	13005.0	310	5318.0	1.0	3	571
2	1980	2100.0	17208.0	168	3212.1	2.0	3	321
3	1972	2296.0	18368.0	197	3512.2	1.0	3	367
4	1930	3190.0	25968.0	367	7579.7	1.0	2	722
5	1985	1230.0	9984.0	88	1697.6	1.0	2	169
6	1990	1216.0	9968.0	99	1744.6	1.0	3	184
7	1982	2985.0	28130.0	412	6812.1	2.0	4	747
8	1975	1920.0	15360.0	240	4233.3	1.0	3	447
9	1990	1370.0	12380.0	67	1156.2	1.5	3	124
10	1964	1847.0	14766.0	221	3790.0	1.0	4	407
11	1990	3866.0	34794.0	340	5602.1	1.0	4	616
12	1987	1500.0	12000.0	96	1651.4	1.0	2	177
13	1970	2486.0	19468.0	374	5927.8	2.0	4	669
14	1984	3474.0	31748.0	371	6723.2	1.0	4	697
15	1970	2276.0	18540.0	215	4160.8	2.0	3	414
16	1970	1296.0	10368.0	136	2798.4	1.0	3	267
17	1981	1360.0	11212.0	84	1506.3	1.0	3	157
18	1977	1041.6	8332.8	133	2177.2	1.0	3	241
19	1979	1526.0	12759.0	100	1723.2	1.0	3	184
20	1975	2118.0	18474.0	167	3443.9	1.0	3	329
21	1950	1595.0	12760.0	193	3201.6	1.0	3	351
22	1987	2703.0	20934.0	420	7033.8	2.0	5	766
23	1975	2143.0	17144.0	218	3855.6	1.0	3	406
24	1970	1688.0	13864.0	112	2003.0	1.0	3	209
25	1976	1850.5	14804.0	180	3263.8	1.0	3	338
26	1971	1888.0	15664.0	154	3168.0	1.0	3	303
27	1958	1806.5	14452.0	153	3163.6	1.0	3	301
28	1970	2254.8	18038.0	174	3066.0	1.0	4	324
29	1989	1458.0	12474.0	138	2468.4	1.0	3	258
30	1999	1702.0	13999.9	83	1495.0	1.0	4	156
31	1980	1154.0	9232.0	85	1496.4	1.0	2	158
32	1973	2544.8	22866.0	219	3885.9	1.0	2	407
33	1977	2592.0	20736.0	149	3258.2	2.0	4	300
34	1961	2373.0	18984.0	203	3524.8	1.0	3	376
35	1927	2284.8	22848.0	349	10057.4	1.0	3	775
36	1925	1706.5	17065.0	334	5846.7	1.0	3	619
37	1970	2477.0	26094.0	261	8173.4	1.5	3	597
38	1968	2402.0	20366.0	333	5237.4	2.0	3	594
39	1965	1439.0	11514.0	149	2242.5	1.0	3	262
40	1984	2686.0	25466.0	283	4891.2	2.0	4	523
41	1977	2343.0	18746.0	258	3827.9	1.5	4	451
42	1994	2899.0	28985.0	166	3029.2	1.0	3	313
43	1985	2388.0	19104.0	182	3372.0	1.0	4	344
44	1980	2205.0	17640.0	124	3396.0	1.0	3	271
45	1993	1333.0	10664.0	81	747.0	1.0	3	119
46	1975	2160.0	17280.0	119	5087.0	1.0	3	305
47	2004	1648.0	16480.0	121	2110.0	1.0	2	224
48	1970	2250.0	18000.0	461	7332.0	1.0	3	825
49	1997	1789.0	14585.0	117	1876.0	1.0	3	210
50	1995	2300.0	18400.0	134	2808.0	1.0	3	265
51	2001	2458.0	24580.0	133	2915.0	1.0	3	267
52	1994	2100.0	16800.0	152	2526.0	1.0	3	271
53	1995	2275.0	21000.0	150	2226.0	1.5	2	262
54	1955	2200.0	19800.0	312	5432.0	1.0	2	577
55	1990	2143.0	17144.0	202	3339.0	1.0	2	503
56	1985	1600.0	12800.0	119	2072.0	1.0	3	220
57	1957	1550.0	12400.0	165	2664.0	1.0	2	297
58	1988	2500.0	21875.0	208	4087.0	1.0	3	402
59	2000	2800.0	25200.0	130	1803.0	1.0	3	230
60	1975	2070.0	16560.0	376	5499.0	1.0	3	653
61	1970	2200.0	18700.0	105	3015.0	1.0	3	233
62	2004	2200.0	23100.0	122	2502.0	1.0	3	239
63	1972	1700.0	14025.0	489	5889.0	1.0	3	792
64	1983	1350.0	11070.0	136	2155.0	1.0	2	244
65	1995	2200.0	20240.0	143	2447.0	1.0	4	263
66	1991	1950.0	17050.0	140	2997.0	1.0	2	279

Correspondence: jinson_je@yahoo.com