Inverse Modeling Toolkit: Numerical Algorithms

John Kelly Kissock, Ph.D., P.E.
Member ASHRAE

Jeff S. Haberl, Ph.D., P.E.
Member ASHRAE

David E. Claridge, Ph.D., P.E.
Member ASHRAE

ABSTRACT

In 1994, ASHRAE began developing a guideline for measuring retrofit savings (GPC-14P). In support of Guideline-14P, ASHRAE initiated RP-1050 to develop a toolkit for calculating linear, change-point linear, and multiple-linear inverse building energy models. The resulting Inverse Modeling Toolkit (IMT) can be used as part of a procedure to measure savings. This paper describes the numerical algorithms used to find general least squares regression, variable-base degree-day, change-point, and combination change-point multivariable regression models in the IMT, as well as the equations used to estimate the uncertainty of predicting energy use for the purpose of measuring savings using IMT models.

INTRODUCTION

Energy conservation retrofits are typically initiated based on predictions of how much energy and money a retrofit will save. However, predicted savings in early energy conservation programs often differed substantially from savings determined by measuring energy consumption before and after a retrofit (Nadel and Keating 1991; Greely et al. 1990; Jamieson and Qualmann 1990). These large discrepancies underscored the need to accurately measure energy savings. As the size and expense of energy conservation programs grew throughout the 1980s, so did the emphasis on evaluation, which became an important part of program management. Measured savings were used to verify the success of retrofits, guide the selection of future retrofits, and, in some cases, to identify and correct operational and maintenance problems (Claridge et al. 1994). The importance of measured savings increased further in the late 1980s when state regulatory agencies began granting shareholder incentives based on measured demand-side management (DSM) program results (Fels and Keating 1993).

In the 1990s, the move toward utility deregulation diminished the size and number of utility DSM programs. However, a new type of retrofit funding mechanism, called performance contracting, emerged in which payment for a retrofit is based on measured savings. The growing popularity of performance contracting created new incentives for developing protocols and standards for measuring savings. In response to this need, the National Association of Energy Service Contractors developed protocols for the measurement of retrofit savings in 1992. In 1994, the U.S. Department of Energy also initiated an effort that resulted in publication of the North American Energy Measurement and Verification Protocols (DOE 1996a) and, later, the International Performance, Measurement and Verification Protocols (DOE 1997, 2001). In addition, the U.S. Federal Energy Management Program developed their own set of Measurement and Verification Guidelines for Federal Energy Projects (DOE 1996b).

In 1994, ASHRAE began developing a guideline for measuring retrofit savings (GPC-14P). One method of measuring savings is described as follows (Kissock et al. 1992; Cowan and Schiller 1997).

1. Measure energy use and influential variables during the pre-retrofit period.

2. Develop a regression model of pre-retrofit energy use as function of influential variables.

3. Measure energy use and influential variables during post-retrofit period.

4. Use the values of the influential variables from the post-retrofit period (Step 3) in the pre-retrofit model (Step 2) to

John Kelly Kissock is an associate professor in the Department of Mechanical Engineering, University of Dayton, Dayton, Ohio. Jeff S. Haberl is an associate professor in the Department of Architecture and David E. Claridge is a professor in the Department of Mechanical Engineering at Texas A&M University, College Station, Tex.

predict how much energy the building would have used if it had not been retrofitted.

5. Subtract measured post-retrofit energy use (Step 3) from the predicted pre-retrofit energy use (Step 4) from to estimate savings.

In support of this method, ASHRAE initiated research project RP-1050 to develop a toolkit of regression models for modeling building energy use. The resulting Inverse Modeling Toolkit (IMT) includes several types of regression models designed to model a wide variety of energy use patterns (Kissock et al. 2002). These models include variable-base degree-day models, change-point models, and multivariable regression models. This paper describes the numerical algorithms used by the IMT regression models. It also describes the equations used to estimate the uncertainty of predicting energy use for the purpose of measuring savings using IMT models.

LEAST-SQUARES REGRESSION ALGORITHM

All regression models in the IMT use generalized least-squares regression to determine the model coefficients. The least-squares regression algorithm used by the IMT begins by filling the arrays X and Y, with values of the independent and dependent variables, respectively. To provide computational stability in cases where the values of the X observations are very large or very small, each X observation is then normalized by the mean value of each independent variable. If the mean value of an independent variable is zero, then no normalization is performed.[1]

Generalized least-squares regression estimates model coefficients that minimize the sum of the squared error between predicted and actual observations. IMT uses a matrix algebra approach to least-squares regression (Neter et al. 1989). In this approach, the matrix of dependent observations, Y, is equal to the product of the matrix of independent observations, X, and the matrix of estimated regression coefficients, β, plus an error term, E.

$$Y = X\beta + E \qquad (1)$$

Solving for β gives

$$\beta = (X^T X)^{-1} X^T Y . \qquad (2)$$

Equation 2 is solved for the matrix of estimated regression coefficients, β, using computational versions of standard matrix algebra (Miller 1981). This generalized algorithm

---

[1] It is possible that in the very rare case in which the mean value of an independent variable were near zero, while several values of the independent variable were very large, this method of normalization could inflate the large values and cause computational instability. However, in the thousands of regressions we have performed using this algorithm, we have never observed this situation.

calculates regression coefficients for both single- and multivariate regression equations.

To calculate the model residuals, the predicted values of the dependent variable, $\hat{Y}$, are computed from

$$\hat{Y} = X\beta . \qquad (3)$$

The matrix of residuals, E, is then computed from

$$E = Y - \hat{Y} . \qquad (4)$$

The root mean squared error of the model, RMSE, is a measure of the scatter of the data around the model. RMSE is computed from

$$RMSE = \sqrt{\frac{\sum(Y - \hat{Y})^2}{(n - p)}} = \sqrt{\frac{Y^T Y - \beta^T X^T Y}{(n - p)}} , \qquad (5)$$

where n is the number of data observations and p is the number of regression coefficients.

The standard error of a regression coefficient is a measure of the uncertainty of the estimate of the regression coefficient. The matrix of the standard errors of the regression coefficients, S, is computed from

$$S = RMSE\sqrt{(X^T X)^{-1}} . \qquad (6)$$

The squared correlation coefficient, R^2, is a number between 0 and 1 that represents the fit of the data to the regression model compared to the fit of the data to the mean of the data. R^2 equal to 1.0 indicates a perfect fit between the data and the regression model. R^2 equal to 0.0 indicates that the regression model provides no better fit than the mean of the data. R^2 is computed from

$$R^2 = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} . \qquad (7)$$

R^2 increases whenever additional regression coefficients are added to a model. Adjusted R^2 compensates for this effect to give a better measure of model fit as additional variables are added. The adjusted R^2 is computed from

$$Adjusted\ R^2 = 1 - \frac{(n - 1)\sum(Y - \hat{Y})^2}{(n - p - 1)\sum(Y - \bar{Y})^2} . \qquad (8)$$

VARIABLE-BASE DEGREE-DAY MODELS

In many single-zone buildings, such as residences and small commercial buildings, space-heating energy use increases as outdoor air temperature decreases below some balance-point temperature. The heating balance-point temperature is defined as the temperature at which the heat gain from internal occupants and equipment balances heat loss through the building envelope. At outdoor air temperatures above the balance-point temperature, no thermal energy is needed for space heating; however, thermal energy may also be required
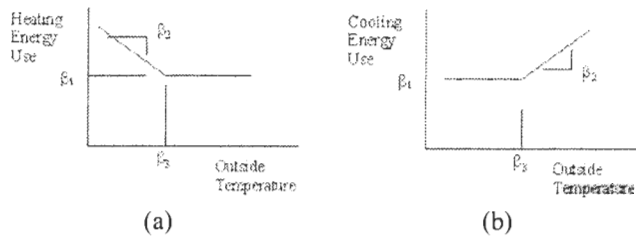
Figure 1 *Typical patterns of thermal (1a) and electrical (1b) energy use in single-zone buildings when the thermal energy is used for space heating and other applications and electricity is used for space cooling and other applications.*

for hot water or cooking. Thus, monthly gas or oil consumption in many single-zone residences forms a pattern such as that shown in Figure 1a. Similarly, cooling energy use frequently increases as outdoor air temperature increases above some cooling balance-point temperature, below which no space cooling is necessary. However, the type of energy used for cooling, such as electricity, may also be used for non-cooling applications. Thus, monthly electricity bills in single-zone buildings in which electricity is used for cooling frequently show an energy use pattern similar to that in Figure 1b.

The variable-base degree-day (VBDD) method was developed to model these types of energy use patterns. During the 1980s, Fels (1986) adapted the VBDD method for use in measuring savings as the PRInceton Scorekeeping Method (PRISM). The algorithm finds the base temperature that gives the best statistical fit between energy consumption and the number of variable-base degree-days in each energy use period. PRISM was one of the first methods to include an estimate of the standard error for all regression parameters (Goldberg 1982). The method had widespread use, especially in evaluation of residential energy conservation programs. Subsequently, PRISM was found to provide adequate fits with commercial building billing data (Eto 1988; Haberl and Vajda 1988; Haberl and Komer 1990; Kissock and Fels 1995; Sonderegger 1998); however, the physical interpretation of the variable-base degree-day method does not apply to commercial buildings with simultaneous heating and cooling (Rabl et al. 1992; Kissock 1993). The FASER (OmniComp 1984) and Metrix Utility (Silicon 2000) data analysis programs have also adapted the VBDD method for modeling monthly baseline energy use. Both programs use manual search procedures to identify the balance-point temperatures.

The IMT VBDD model uses an automated search procedure to identify the balance-point temperature that produces the best fit to the data. The functional forms of the IMT heating and cooling VBDD models are shown below.

$$Y = \beta_1 + \beta_2 \, HDD(\beta_3) \qquad (9)$$

$$Y = \beta_1 + \beta_2 \, CDD(\beta_3) \qquad (10)$$

where β_1 is the constant term, β_2 is the slope term, and $HDD(\beta_3)$ and $CDD(\beta_3)$ are the number of heating and cooling degree-days, respectively, in each energy data period calculated with base temperature β_3. The number of heating and cooling degree-days in each energy data period of n days is

$$HDD(\beta_3) = \sum_{i=1}^{n} (\beta 3 - T_i)^+ \qquad (11)$$

$$CDD(\beta_3) = \sum_{i=1}^{n} (T_i - \beta 3)^+, \qquad (12)$$

where T_i is the average daily temperature.

To calculate VBDD models, IMT fills and returns the arrays HDD and CDD with the heating and cooling degree-days, respectively, for each energy period according to Equations 11 and 12. The arrays $HDD(i,j)$ and $CDD(i,j)$ contain the number of degree-days in each energy period (i) and for base temperatures from 41°F to 80°F (j). The best-fit VBDD model is identified using a search method by regressing Equation 9 or 10 using the HDDs or CDDs in each energy period for successive base temperatures, $\beta3$, from 41°F to 80°F. The base temperature that results in the model with the highest R^2 is recorded. Equation 9 or 10 is then regressed once more using the base temperature that results in the model with the highest R^2, and the results are reported.

CHANGE-POINT MODELS

In general, heating and cooling energy consumption in multizone buildings tends to vary with ambient temperature throughout the entire range of ambient temperatures encountered. Thus, the VBDD method, which specifies a constant energy usage below or above the balance-point temperature, is frequently not appropriate. In addition, linear two-parameter regression models fail to capture the nonlinear relationship between heating and cooling energy use and ambient temperature caused by system effects, such as VAV control or latent loads (Kissock et al. 1998a).

Change-point (CP) models, however, succeed at capturing both effects, and, as a consequence, have had widespread use as baseline models for measuring energy savings (Haberl et al. 1998; DOE 1997). The IMT includes four basic types of change-point models (Figure 2). The type of model is identified by the number of regression coefficients β.

Combination CP-MVR and VBDD-MVR Models

Change-point (CP) and variable-base degree-day (VBDD) models have been shown to provide good fits between building energy use and ambient temperature. However, other variables also influence building energy use. A simple multivariable regression (MVR) model could capture the effect of multiple independent variables; however, it could not model energy use that varies at ambient temperature change points and balance points. Combination CP-MVR and
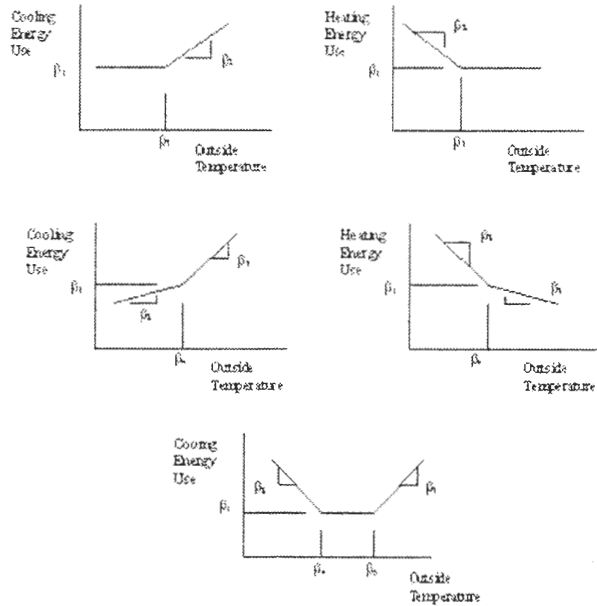
Figure 2 *IMT change-point models. Top row: 3P cooling and heating models. Second row from top: 4P cooling and heating models. Bottom row: 5P heating and cooling model.*

VBDD-MVR models retain the ability to model energy use with temperature change points and balance points, while including the effects of additional independent variables. One approach reported in the literature (Rabl and Riahle 1992; Ruch et al. 1993; Sonderegger 1997, 1998) is to sequentially identify the change-point or base temperature and then use this result in an MVR model. An alternative approach is to use indicator variables to produce separate CP or VBDD models for each operating or occupational mode (Austin 1997; Kissock et al. 1992).

In the IMT, change-point algorithms were extended to include multiple independent variables. Using this approach, CP-MVR models can be identified in a single step, rather than sequentially, and without breaking up the data according to operational modes. The IMT can also produce VBDD-MVR models by first running the VBDD model and then running the MVR model on the VBDD residual file.

Three-Parameter Models

The functional forms for best-fit three-parameter change-point models for cooling (3PC) and heating (3PH), respectively, are

$$Y_c = \beta_1 + \beta_2 (X_1 - \beta_3)^+ \qquad (13)$$

$$Y_h = \beta_1 + \beta_2 (X_1 - \beta_3)^- , \qquad (14)$$

where β_1 is the constant term, β_2 is the slope term, and β_3 is the change point. The ()+ and ()- notations indicate that the values

of the parenthetic term shall be set to zero when they are negative and positive, respectively. The 3P models are appropriate for modeling building energy use that varies linearly with an independent variable over part of the range of the independent variable and remains constant over the other part. For example, 3PC models, using outside air temperature as the independent variable, are often appropriate for modeling whole building electricity use in residences electric air conditioning. Similarly, 3PH models, using outside air temperature as the independent variable, are often appropriate for modeling heating energy use in residences with gas or oil heating.

IMT can also find combination three-parameter multivariable regression models (3P-MVR), with up to four independent variables, of the type

$$Y_c = \beta_1 + \beta_2 (X_1 - \beta_3)^+ + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_4 \qquad (15)$$

$$Y_h = \beta_1 + \beta_2 (X_1 - \beta_3)^- + \beta_4 X_2 + \beta_5 X_3 + \beta_6 X_4 , \qquad (16)$$

where X_1 is typically temperature, and X_2, X_3, and X_4 are optional independent variables.

Four-Parameter Model

The functional form for best-fit four-parameter (4P) change-point model is

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_4)^+ \qquad (17)$$

where β_1 is the constant term, β_2 is the left slope, β_3 is the right slope, and β_4 is the change point. IMT can also find combination four-parameter multivariable regression models (4P-MVR), with up to three independent variables, of the type

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_4)^+ + \beta_5 X_2 + \beta_6 X_3 \qquad (18)$$

where X_1 is typically temperature, and X_2 and X_3 are optional independent variables.

Four-parameter models using outdoor air temperature as the independent variable are appropriate for modeling heating and cooling energy use in variable-air-volume systems and/or in buildings with high latent loads. In addition, these models are sometimes appropriate for describing nonlinear heating and cooling consumption associated with hot-deck reset schedules and economizer cycles (Kissock 1993).

Five-Parameter Model

The functional form for best-fit five-parameter (5P) change-point model is (Kissock 1996)

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_5)^+ , \qquad (19)$$

where β_1 is the constant term, β_2 is the left slope, β_3 is the right slope, β_4 is the left change point, and β_5 is the right change point.

IMT can also find combination five-parameter multivariable regression models (5P-MVR), with up to two independent variables, of the type

$$Y = \beta_1 + \beta_2 (X_1 - \beta_4)^- + \beta_3 (X_1 - \beta_5)^+ + \beta_6 X_2 , \qquad (20)$$

where X_1 is typically temperature and X_2 is an optional independent variable.

Five-parameter models using outdoor air temperature as the independent variable are appropriate for modeling energy consumption data that include both heating and cooling, such as whole-building electricity data from buildings with electric heat pumps or both electric chillers and electric resistance heating. They are also appropriate for modeling fan electricity consumption in variable-air-volume systems.

ALGORITHMS FOR FINDING BEST-FIT CHANGE POINT MODELS

In the statistical literature, change-point models are known as piece-wise linear regression models or spline fits. In these models, the data are divided into intervals and line segments fit to the data in each interval with the constraint that the line segments meet at a common point between each interval (Hudson 1966). Algorithms for piece-wise linear regression have been developed for cases in which the change point between linear sections is known in advance (Neter et al. 1989). When the change-point is not known in advance, it is sometimes estimated by inspection (Maidment et al. 1985; Schrock and Claridge 1989); however, this method does not guarantee a "best fit."

The literature review identified three algorithms that may be applicable for finding best-fit change-point models. The first algorithm was published by Crawford et al. (1991). The procedure begins by dividing the data into n bins along the x-axis. Developing simple linear regression models for each bin would result in discontinuities between the linear segments. To overcome this problem, the bin widths are varied until the lines intersect at the bin boundaries. Because of the uncertainty of obtaining convergence, the inability to specify the number of change points, and the reliance of the final result on the initial conditions, this method was not recommended for the Inverse Modeling Toolkit.

The second method was published by Ruch and Claridge (1992). This method develops a four-parameter change-point model of energy consumption, typically as a function of dry-bulb temperature, along with accompanying error diagnostics for the model's parameters. The algorithm finds the optimal change point by searching within an interval known to contain the change point. The first step is to split the data into two temperature regimes, fit ordinary least-squared lines in each regime, and calculate the intersection of the lines. This is repeated for numerous temperature regions. In the second stage, the change point is assumed and the model is fit using linear regression. From the collection of fits in the two stages, the algorithm chooses the one with the lowest RMSE. The reliability of the parameter estimated is then calculated. The algorithm was coded into a computer program called 4P in the early stages of the Texas LoanSTAR program. Unfortunately, the method did not prove to be robust when used on actual measured energy data. In addition, the prescription of defining an acceptable region for the change point (1) required that the data be preinspected and (2) created the possibility that the true best-fit change point might lie outside of that region. For these reasons, this algorithm was not recommended for the Inverse Modeling Toolkit.

The third algorithm uses a two-stage grid search to identify the best-fit change point (Kissock et al. 1994). In this method, the minimum x value is selected as the initial change point in a standard piece-wise linear regression equation. The change point is then incremented and the regression is repeated across the range of x-values. The change point that results in the lowest RMSE is selected as the best-fit change-point temperature. This method is then repeated with a finer grid centered about the initial best-fit change point. The uncertainty with which the change-point temperature is known can be approximated as the width of the finest grid. The method is easily adaptable to three-parameter heating, three-parameter cooling, and four-parameter models. A similar algorithm for five-parameter models was developed by Kissock (1996). Models based on this algorithm have been used extensively with building energy data and have proven to be extremely robust (Haberl et al. 1998). Because of the simplicity, robustness, and accuracy of this algorithm, it was selected for use in the Inverse Modeling Toolkit.

IMT uses the same algorithm for finding all change-point models, including both single and multivariable change-point models. The algorithm is demonstrated for the 3P models in the following description.

The best-fit change-point temperature β_3 is identified using a two-part grid-search method (Figure 3). The first step is to identify minimum and maximum values of $X1$ and to divide the interval defined by these values into ten increments of width dx. Next, the minimum value of $X1$ is selected as the initial value of β_3 and the model is regressed against the data to find β_1, β_2, and RMSE. The value of β_3 is then incremented by dx and the regression is repeated until β_3 has traversed the entire range of possible X values. The value of β_3 that results in the lowest RMSE is selected as the initial best-fit change-point. This method is then repeated using a finer grid of width $2\,dx$, centered about the initial best-fit value of β_3. The uncertainty with which the final change-point temperature is known is reported as twice the width of the finest grid since, as a result of the search method, the change-point temperature that produced the best fit lies within this interval.[2]

The algorithms for finding change-point and change-point multivariable regression models are identical, the only difference being that the regression model is in the form of Equations 15, 16, 18, or 20 instead of Equations 13, 14, 17, or 19.

[2.] The uncertainty of savings, described in the next section, depends on the overall fit of the model with the data as quantified by the model's RMSE. It is not an explicit function of the uncertainty of the individual regression parameters or the change-point temperature.
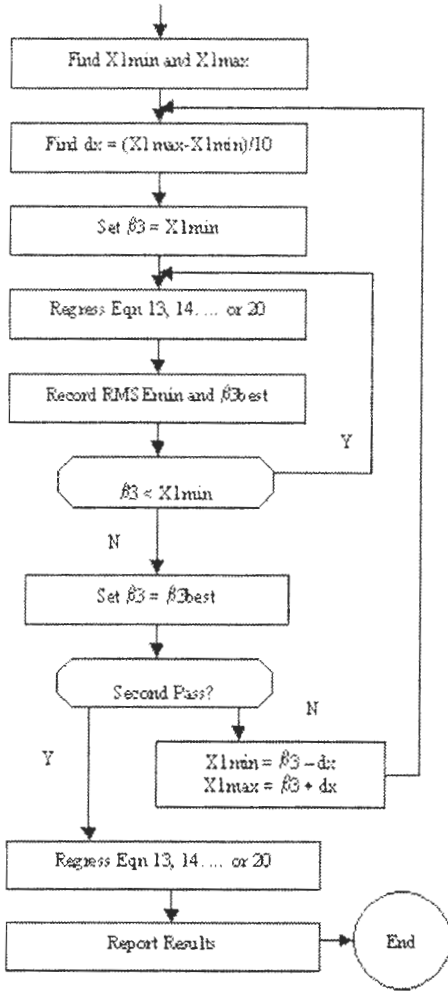
Figure 3 *Flow diagram of algorithm for finding the best-fit change-point model.*

When regressing change-point models, the parenthetic + and − terms are computed with the use of an indicator variable, I. For example, in Equation 13, the regression equation passed to the regression subroutine is

$$Y = a + b\chi , \qquad (21)$$

where χ represents $(X_1 - \beta_3)^+$. The numerical value of χ is computed as

$$\chi = I(X_1 - \beta_3) \text{ where } I = 0 \text{ when } X_1 \le \beta_3$$
$$\text{and } I = 1 \text{ when } X_1 > \beta_3 . \qquad (22)$$

UNCERTAINTY OF SAVINGS

Goldberg (1982) estimated the uncertainty of VBDD parameters in the PRISM method. Cowan and Schiller (1997), among others, discuss the uncertainty of the estimated savings in terms of the money, time, and equipment required to reduce the uncertainty. Kissock et al. (1993) and Katapamula et al. (1995) investigate how the length and timing and data time periods of baseline periods affect the prediction accuracy of the baseline regression models. Kissock et al. (1998b) discuss the error in retrofit savings calculations due to varying indoor air temperature or internal gains. A complicated algorithm for estimating error associated with linear models was proposed by Ruch et al. (1999). The algorithm was translated into a computer code by Ruch and Kissock and tested in development versions of EModel (Kissock et al. 1994). Unfortunately, the uncertainty routines were sometimes unstable.

A simplified method of estimating the uncertainty associated has been described by Reddy et al. (1998) and Kissock et al. (1998a). In this method, the energy savings, Sav_i, and the associated uncertainty, e_{si}, for any data time interval i in the post-retrofit period can be written as

$$Sav_i \pm \varepsilon_{si} = (\hat{Y}_i \pm \varepsilon_{pi}) - (Y_{mi} \pm \varepsilon_{mi}) , \qquad (23)$$

where $\hat{Y}_i$ is the energy consumption predicted by a pre-retrofit model, Y_{mi} is the measured energy consumption during the post-retrofit period, and ε is the random error associated with each parameter. The uncertainty ε_{pi} of a predicted value of energy use, $\hat{Y}_i$, is (Neter et al. 1989)

$$\varepsilon_{pi} = t(1 - \alpha/2, n - 2) \text{RMSE} \left[1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum\limits_{i=1}^{n} (X_i - \bar{X})^2} \right]^{\frac{1}{2}} , \qquad (24)$$

where the t-statistic, $t(1-\alpha/2, n-p)$, is a function of the level of significance, α, the number of days in the pre-retrofit period, n, and the number of parameters in the model, p. The level of significance, α, indicates the fraction of predictions that are likely to fall outside of the prediction uncertainty bands. In practice, the value of the t-statistic is close to 1.96 for a reasonable number of pre-retrofit data points and a 5% significance (95% confidence) level. In addition, the value of the parenthetic term is usually very close to unity. Thus, ε_{pi} can be closely approximated as

$$\varepsilon_{pi} \approx 1.96 \text{ RMSE } (1 + 2/n)^{1/2}. \qquad (25)$$

Assuming that the prediction and measurement errors are independent, the uncertainty of a value of savings, ε_{si}, is

$$\varepsilon_{si} = [(\varepsilon_{pi})^2 + (\varepsilon_{mi})^2]^{\frac{1}{2}}. \qquad (26)$$

The uncertainty associated with the total savings over m periods in the post-retrofit period is the root sum of squares of the uncertainty associated with each daily value of savings.

$$\varepsilon_{st} = \left[\sum_{i=1}^{m} (\varepsilon_{si})^2 \right]^{\frac{1}{2}} \qquad (27)$$

When ε_{pi} and ε_{mi} are constants, Equation 27 reduces to

$$\varepsilon_{st} = \varepsilon_{si}\sqrt{m} . \qquad (28)$$

SUMMARY AND CONCLUSIONS

This paper describes numerical algorithms used by the ASHRAE inverse modeling toolkit to derive regression models of building energy use. These algorithms include the generalized least-squares algorithm; the algorithms for calculating goodness-of-fit parameters; the algorithms for finding the best-fit variable-base heating and cooling degree-day models; the algorithms for finding the best-fit three, four, and five-parameter change-point models; the algorithms for finding the best-fit combination change-point multivariable-regression models; and the algorithms for finding the uncertainty of model predictions and savings.

The change-point model algorithms use a two-stage grid search to identify the best-fit change point(s). The grid search method has proven to be simple to program, robust, and able to handle the problem of multiple local maximums. Perhaps the greatest potential drawback associated with grid-search optimization methods is increased computational time. However, due to the efficiency of the matrix algebra least-squares algorithm and the computational speed of modern computers, the time required to identify change-point models, even with up to 9,000 observations, is acceptably short.

ACKNOWLEDGMENTS

NOMENCLATURE

α	=	level of significance, fraction of predictions likely to fall outside prediction uncertainty bands
β	=	matrix of estimated regression coefficients
$CDD(\beta_3)$	=	number of cooling degree-days calculated with base temperature β_3
CP	=	change-point model
CV-RMSE	=	coefficient of variation of root mean square error
dx	=	width of search grid
E	=	matrix of residuals
ε_{pd}	=	uncertainty associated with predicting energy use
$HDD(\beta_3)$	=	number of heating degree-days calculated with base temperature β_3
I	=	indicator variable (0 or 1)
m	=	number of periods in the post-retrofit period
MVR	=	multiple variable regression model
n	=	number of data observations
p	=	number of regression coefficients
P	=	number of regression coefficients
R^2	=	squared correlation coefficient
RMSE	=	root mean square error
S	=	matrix of the standard errors of the regression coefficients
T	=	air temperature
$t()$	=	t-statistic
VBDD	=	variable-base degree day model
X	=	array of independent variables
Y	=	array of dependent variables
$\hat{Y}$	=	array of predicted values of the dependent variable

REFERENCES

Austin, S. 1997. Regression analysis for savings verification. *ASHRAE Journal,* October, pp. 57-63.

Claridge, D., J. Haberl, M. Liu, J. Houcek, and A. Aamer. 1994. Can you achieve 150% of predicted retrofit savings? Is it time for recommissioning? *Proceedings of the ACEEE 1994 Summer Study on Energy Efficient Buildings, Pacific Grove, CA*, August, pp. 5.73-5.87.

Cowan, J., and S. Schiller. 1997. Measuring' Energy Savings for Modernization Projects. *ASHRAE Journal*, August, pp. 60-62.

Crawford, R., R. Dykowski, and S. Czajkowski. 1991. A segmented linear least-squares modeling procedure for nonlinear HVAC components. *ASHRAE Transactions* 97(2): 11-18.

DOE. 1996a. North American Energy Measurement and Verification Protocol. DOE/EE-0081, U.S. Department of Energy, Washington, D.C.

DOE. 1996b. Measurement and Verification Guidelines for Federal Energy Projects. DOE/GO-10096-248, U.S. Department of Energy, Washington, D.C.

DOE. 1997. International Performance Measurement and Verification Protocol. U.S. Department of Energy, Washington, D.C.

Eto, J. 1988. On using degree-days to account for the effects of weather on annual energy use in office buildings. *Energy and Buildings* 12(2): 113-127.

Fels, M. 1986. PRISM: An introduction. *Energy and Buildings*, Vol. 9, pp. 5-18.

Fels, M., and K. Keating. 1993. Measurement of energy savings from demand-side management programs in US electric utilities. *Annual Review of Energy and Environment*, 18:57-88.

Goldberg, M. 1982. A geometrical approach to nondifferentiable regression models as related to methods for assessing residential energy conservation. Ph.D. disser-

tation, Department of Statistics. Princeton University, Princeton, N.J.

Greely, K., J. Harris, and A. Hatcher. 1990. Measured savings and cost-effectiveness of conservation retrofits in commercial buildings. Lawrence Berkeley Laboratory Report - 27586, Berkeley, CA.

Haberl, Thamilseran, Reddy, Claridge, O'Neal, and Turner. 1998. Baseline calculations for measurement and verification of energy and demand savings in a revolving loan program in Texas. *ASHRAE Transactions* 104(2): 841-858.

Haberl, J., and E. Vajda. 1988. Use of metered data analysis to improve building operation and maintenance: Early results from two federal complexes. *Proceedings of the ACEEE 1988 Summer Study on Energy Efficient Buildings, Pacific Grove, CA,* August, pp. 3.98 - 3.111.

Haberl, J., and P. Komer. 1990. Improving commercial building energy audits: How annual and monthly consumption data can help. *ASHRAE Journal* 32(8): 26-33.

Hudson, D. 1966. Fitting segmented curves whose join points have to be estimated. *Journal of American Statistical Association.* Vol. 61, pp. 1097-1125.

Jamieson, D., and R. Qualmann. 1990. Computer simulation energy use metering or can we count on energy savings estimates in designing demand side programs. *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA,* August, pp. 10.105-10.114.

Katipamula, S., T. Reddy, and D. Claridge. 1995. Effect of time resolution on statistical modeling of cooling energy use in large commercial buildings. *ASHRAE Transactions* 101(2): 172-185.

Kissock, K, D.E. Claridge, J.S. Haberl, and T.A. Reddy. 1992. Measuring retrofit savings for the Texas Loan-STAR Program: Preliminary methodology and results. *Solar Engineering, 1992: Proceedings of the ASME-JSES-SSME International Solar Energy Conference, Lahaina, HI,* April.

Kissock, K., T. Agami, D. Fletcher, and D. Claridge. 1993. The effect of short data periods on the annual prediction accuracy of temperature-dependent regression. *International Solar Engineering Conference,* pp. 455-463.

Kissock, K. 1993. A methodology to measure energy savings in commercial buildings. Ph.D. dissertation, Mechanical Engineering Department, Texas A&M University, College Station, TX, December.

Kissock, K, W. Xun, R. Sparks, D. Claridge, J. Mahoney and J. Haberl. 1994. EModel Version 1.4de. Copyright Texas A&M University, Energy Systems Laboratory, Department of Mechanical Engineering, Texas A&M University, College Station, TX, December.

Kissock, K., and M. Fels. 1995. An assessment of PRISM's reliability for commercial buildings. *Proceedings of the National Energy Program Evaluation Conference,* Chicago, IL, August.

Kissock, K. 1996. Development of analysis tools in support of the Texas LoanSTAR Program. University of Dayton, Department of Mechanical and Aerospace Engineering, Dayton, OH, August.

Kissock, K., A. Reddy, and D. Claridge. 1998a. Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. *ASME Journal of Solar Energy Engineering* 120(3): 168-176.

Kissock, K., H. Joseph, and J. McBride. 1998b. The effects of varying indoor air temperature and heat gain on the measurement of retrofit savings. *ASHRAE Transactions* 104(2): 895-900.

Kissock, K., J. Haberl, and D. Claridge. 2002. Development of a toolkit for calculating linear, change-point linear and multiple-linear inverse building energy analysis models. Final Report, ASHRAE 1050-RP, November.

Maidment, D., S. Miaou, D. Nvule, S. Buchberger. 1985. Analysis of daily water use in nine cities. Center for Research in Water Resources, Bureau of Engineering Research, University of Texas, Austin, TX, CRWR 201.

Miller, A. 1981. BASIC Programs for Scientists and Engineers. Sybex, Inc., Berkeley, Calif.

Nadel, S., and K. Keating. 1991. Engineering estimates vs. impact evaluation results: How do they compare and why? *Energy Program Evaluation Conference,* Chicago, pp. 24-33.

Neter, J., W. Wasserman, and M. Kutner. 1989. *Applied linear regression models.* Boston, Mass.: Irwin Press.

Omnicorp. 1984. FASER.

Rabl, A., L. Norford, and J. Spadaro. 1992. Steady state models for analysis of commercial building energy data. *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA,* August, pp. 9.239-9.261.

Rabl, A., and A. Rialhe. 1992. Energy signature models for commercial buildings: Test with measured data and interpretation. *Energy and Buildings* 19(2): 143-154.

Reddy, T., J. Kissock, and D. Ruch. 1998. Uncertainty in baseline regression modeling and in determination of retrofit savings. *ASME Journal of Solar Energy Engineering* 120(3): 185-192.

Ruch, D., and D. Claridge. 1992. A four-parameter change-point model for predicting energy consumption in commercial buildings. *ASME Journal of Solar Energy Engineering* 114(2): 77-83.

Ruch, D., Chen, J. Haberl, and D. Claridge. 1993. A change-point principal component analysis (CP/PCA) method for predicting energy usage in commercial buildings: The PCA model. *ASME Journal of Solar Energy Engineering* 115(2): 77-84.

Ruch, D.K., J.K. Kissock, and T.A. Reddy. 1999. Model identification and prediction uncertainty of linear building energy use models with autocorrelated residuals. *ASME Journal of Solar Energy Engineering* 121(1): 63-68.

Schrock. D., and D. Claridge. 1989. Predicting energy usage in a supermarket. *Proceedings of the Sixth Symposium on Improving Building Systems in Hot and Humid Climates.* Mechanical Engineering Department, Texas A&M University, Dallas, TX, October, pp. 44 - 54.

Silicon Energy Corp. 2000. Metrix Utility Accounting System. 1010 Atlantic, Alameda, Calif.

Sonderregger, R. 1997. Energy retrofits in performance contracts: Linking modeling and tracking. *Cool Sense National Forum on Integrated Chiller Retrofits,* San Francisco, September 23-24.

Sondcregger, R.A. 1998. Baseline model for utility bill analysis using both weather and non-weather-related variables. *ASHRAE Transactions* 104(2): 859-870.