VOL. 6, NO. 1 HVAC&R RESEARCH JANUARY 2000

Uncertainty of "Measured" Energy Savings from Statistical Baseline Models

T. Agami Reddy, Ph.D., P.E.
Member ASHRAE

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

Baseline models are a crucial element in determining savings from energy conserving measures. The baseline model is obtained by regressing the energy consumption data for the period prior to the implementation of the energy conservation measures. The widely used criteria for determining adequacy of a particular baseline model is based on statistical cutoff criteria that do not give the user a knowledge of the error inherent in the savings determination. The absolute cutoff criteria may not be appropriate since baseline model development is not the desired end. It is proposed that models be evaluated in terms of the ratio of the expected uncertainty in the savings to the total savings ($\Delta E_{save}/E_{save}$). This physically and financially intuitive measure permits the user to vary the criteria according to factors most relevant for a particular energy conservation project. Simplified expressions for ($\Delta E_{save}/E_{save}$) appropriate for use by practitioners as applicable for cases with uncorrelated data and for use with correlated time series data are developed and discussed in the context of case-study data. The use of this concept to logically select the most appropriate measurement and verification protocol to verify savings is also described.

INTRODUCTION

The International Performance Measurement and Verification Protocol (IPMVP 1997) proposes three monitoring and verification methods or options for determining energy savings from energy conserving measures (ECM), either due to retrofits or operational and maintenance improvements in buildings. The first (Option A) deals with one-time measurements before and one-time measurements after the ECM. Options B and C, however, deal with continuous measurements before and after the ECM. Option B typically involves monitoring specific end uses for a period (e.g. weeks to months) before and after the retrofit, while Option C entails measuring whole building consumption for a baseline period of at least several months before the retrofit and continuously following the retrofit. The discussion in this paper is limited only to Options B and C wherein baseline model development is explicit. Such savings will be called "measured" savings to distinguish them from audit-estimated savings (or other savings based primarily on engineering analysis) and are determined from measured energy performance data.

With building energy use data, either utility bills or monitored data that is available, both prior to and after the ECM, the procedure used to determine savings is to normalize the data for any discrete one-time changes (e.g., in conditioned area or in occupancy), and to identify a pre-ECM model using known quantifiable variables that are likely to change after the ECM is implemented but which are not part of the ECM. These variables include climatic and building related variables such as internal loads or changes to building or HVAC operational schedules, (Katipamula et al. 1998). In most cases, however, a model with the outdoor dry-bulb temperature as the only

T. Agami Reddy is an associate professor in the Department of Civil and Architectural Engineering, Drexel University, Philadelphia, Pennsylvania and **David E. Claridge** is a professor in the Energy Systems Laboratory, Mechanical Engineering Department, Texas A&M University, College Station, Texas

regressor variable is adequate (Kissock et al. 1998. Reddy et al. 1997a). There are, subsequently, two variants used to determine savings depending on how the post-ECM data is used:

1. The "Normalized Annual Average Savings" approach, which involves developing an outdoor dry-bulb temperature-based post-ECM regression model. Long-term average dry-bulb temperature values over several years are used to drive this model along with the baseline model (Fels 1986, Ruch and Claridge 1993); and
2. The "Actual Savings" over a certain time period in which the baseline model is driven with actual monitored outdoor temperature under post-ECM conditions. The sum of the differences between these values and the observed post-ECM values over that time period yields the savings (Kissock et al. 1998, IPMVP 1997).

This study is concerned primarily with (2) above. In this case, the uncertainty in the baseline consumption as determined from the statistical model becomes the major determinant of the uncertainty in the resulting measured savings. The uncertainty is of major interest to building owners and financial institutions associated with energy service contracts that are intended to reduce energy cost of operating specific buildings. The uncertainty in the baseline consumption is in turn directly related to the goodness of fit of the baseline models.

Little work has been done to establish sound statistical limits for gauging the goodness-of-fit of the baseline model. The most widely used criterion is that proposed by Reynolds and Fels (1988). They proposed (1) that models with coefficient of determination, i.e., R^2 values >0.7 and coefficient of variation (CV values) < 7%, or (2) that models with low R^2 values and CV < 12%, be considered reliable models. Models that do not satisfy at least one of these criteria be considered poor and that data from buildings with poor baseline models should be discarded while performing demand-side management evaluations. These cutoffs appear arbitrary, since no clear rationale for their choice is provided. Further, the interpretation of R^2 and CV for change point models of energy use versus outdoor dry-bulb temperature (Kissock et al. 1998) is unclear since these indices have a clear physical/intuitive meaning only when applied to linear models.

Further, we contend that there is no basis for development of absolute statistical cutoff criteria since baseline model development is not an end in itself. We propose that the more relevant criterion for use in determining whether a baseline model is acceptable or not is the fractional uncertainty in savings measurement $\Delta E_{save}/E_{save}$. As an example, consider two cases. In one, an energy audit estimates that the retrofit is likely to reduce energy use by 30%, and for the other the anticipated energy reduction is only 5%. A much poorer baseline model in the first case would produce a particular choice of required uncertainty in savings (e.g., $\Delta E_{save}/E_{save} = 0.1$), but a very good model would be needed in the second case to produce the same uncertainty. A physically based criterion such as this would provide more meaningful evaluations of demand side management and ECM programs. This paper presents this physical concept in statistical and mathematical terms and illustrates the approach with energy use utility bill data from several Army bases nation-wide as described by Reddy et al. (1997b). The logical consequence of using this criterion for selection of the most appropriate measurement and verification method as described in the IPMVP (1997) to verify savings is also discussed.

STATISTICAL CRITERIA

The goodness-of-fit of statistical baseline models is customarily expressed in terms of statistical indices, the coefficient of determination R^2 and the coefficient of variation of the root mean square error (CVRMSE). In case of a mean model, the coefficient of variation of the standard deviation (CVSTD) is appropriate. In this study, y is the dependent variable of some function of the predictor variable(s), $\bar{y}$ the sample arithmetic mean, and $\hat{y}$ the regression model-predicted value of *y*. Then for linear models with an intercept term, these indices are defined as follows (Draper and Smith 1981):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{1}$$

and

$$\text{CVRMSE} = \frac{1}{\bar{y}}\left[\frac{\sum (y_i - \hat{y}_i)^2}{n-p}\right]^{0.5} \tag{2a}$$

$$\text{CVSTD} = \frac{1}{\bar{y}}\left[\frac{\sum (y_i - \bar{y})^2}{n-1}\right]^{0.5} \tag{2b}$$

where all summations are done from $i = 1$ to $i = n$, where n is the number of observations, and p is the number of parameters in the regression model. R^2 values are bounded in the range {0,1} while CV values (i.e., either CVRMSE or CVSTD) have a lower bound of zero and no upper bound. The closer to unity the R^2 values and the closer to zero the CV of a linear model, the better the model is deemed to be. As discussed later, these indices, though representative of the goodness-of-fit of the model, have slightly different interpretation, and situations often arise when the analyst is uncertain as to which measure to turn to in order to make absolute evaluations of models identified from different energy use data. For example, gas use often varies widely over the year, with a very low value in summer and a very high value in winter. As a result the models may have, for example, a R^2 value of 0.98 and a CV value of over 20% (Reddy et al., 1997b). Whether this should be considered a good model or not is ambiguous since the R^2 value taken by itself would suggest a good model, while the CV value considered alone would suggest a poor model. This paper will discuss this issue and provide an understanding of both these indices in terms of baseline energy models.

There are fundamental differences between the R^2 and CV statistics defined by Equations (1) and (2), respectively. Both the statistics are normalized indices, but the normalization is done differently and so their interpretation is different. Both have essentially the same numerator, but the denominators are different. R^2 is representative of the variation in the dependent variable data explained by the model as compared to the variation of the data about the mean, while CV is the mean variation in the data not explained by the model normalized by the mean value of the dependent variable. As illustrated conceptually in Figure 1a, if linear models are fit to two sets of data, both of which have the same mean and slope but one with greater data spread than the other (in Figure 1a the data spread of the right-hand set is twice that of the left-hand side), these models will have R^2 and CV values as shown. There is little ambiguity here because either index can be used to identify the preferred model. However, for data sets with the same mean and spread but different slopes (see Figure 1b), the CV values are unaltered while the R^2 values are widely different. A third set of models with one set of data having twice the mean value and twice the variability is shown in Figure 1c. Both models have the same CV values but widely different R^2 values.

Thus, in terms of baseline energy modeling, R^2 is of limited value when the slope of the linear model is too high or too low, while the CV is not sensitive to the slope of the model but it is sensitive to the data spread with respect to the mean value. Hence, if the objective is only to evaluate how well a baseline energy model has captured the variation in the data, and to assess how

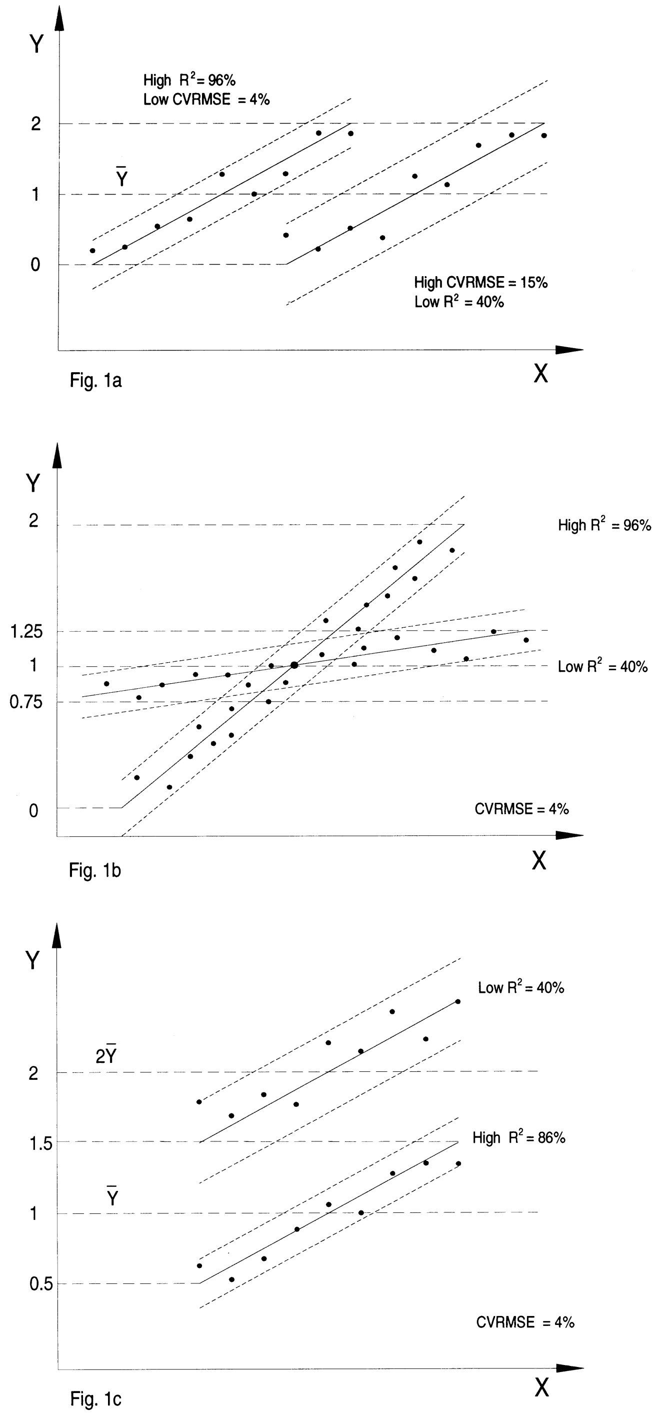

Figure 1. Conceptual plots to illustrate how different configurations of linear data affect model R^2 and CVRMSE

(a) Both models have the same mean values, but one has twice the data spread of the other
(b) Both models have the same mean and data spread
(c) One model has twice the mean value and the data spread as the other

different model functional forms have fared against one another when fit to the *same data set*, then R^2 is the right measure to use. However, if the uncertainty in the savings prediction is the quantity upon which we wish to base our choice of baseline model reliability and to do so with *different data sets*, then CV is the more relevant measure to use. The CV provides a more direct correspondence between the amount of random variation relative to the mean. If the user, were analyzing measured energy use data from several buildings and decided (based on the particular problem) that models with random components less than 10% of the mean were acceptable, then models with CV < 0.1 would be accepted for subsequent predictive purposes.

Several studies have pointed out that building energy use data E plotted against outdoor dry-bulb temperature T often exhibits a change point behavior, i.e., a segmented linear trend (Fels 1986, Kissock et al. 1992, IPMVP 1997). A four-parameter change point (4-P) cooling energy use model conceptually illustrated in Figure 2 would have the following functional form:

$$E = E_c + a(T - T_c)^- + b(T - T_c)^+ \tag{3}$$

If the base portion (i.e., Region 1 of Figure 2) is horizontal, i.e., of zero slope, a three-parameter (3-P) model is obtained. The lower slope in Region 1 and higher slope in Region 2 are representative of cooling energy models, while the trend is reversed for heating energy use models. The theoretical framework of linear segmented models, which are non-linear models, has been discussed by Beale (1960) and by Hinckley (1969) and has been subsequently adapted to energy use data by Goldberg (1982) and Ruch and Claridge (1993). Reddy et al. (1998) pointed out that the non-linear behavior can be simplified to a linear problem if the uncertainty associated with the change point is overlooked. In such a case, a 4-P model simplifies down to two simple linear models with a known hinge or change point value T_c. This assumption is implicit throughout this paper.

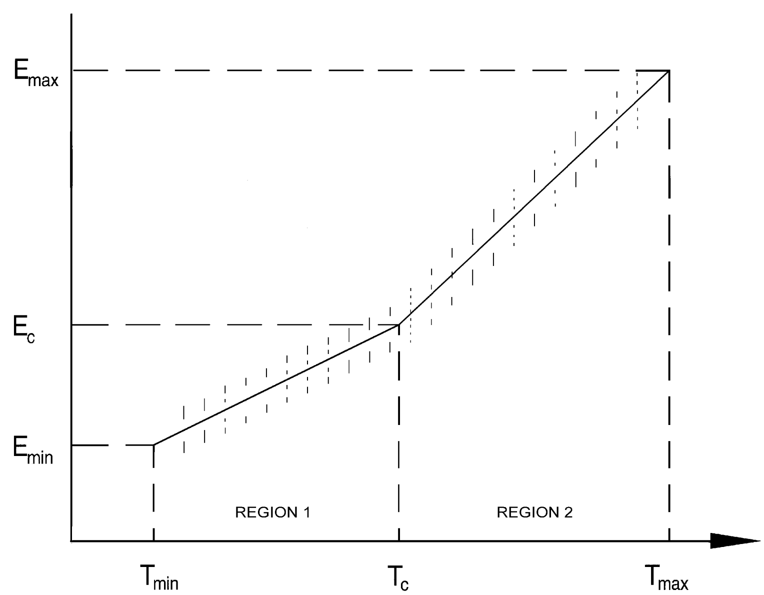

Figure 2. Illustration of data points fitted by a change-point four-parameter cooling model (4P-C)

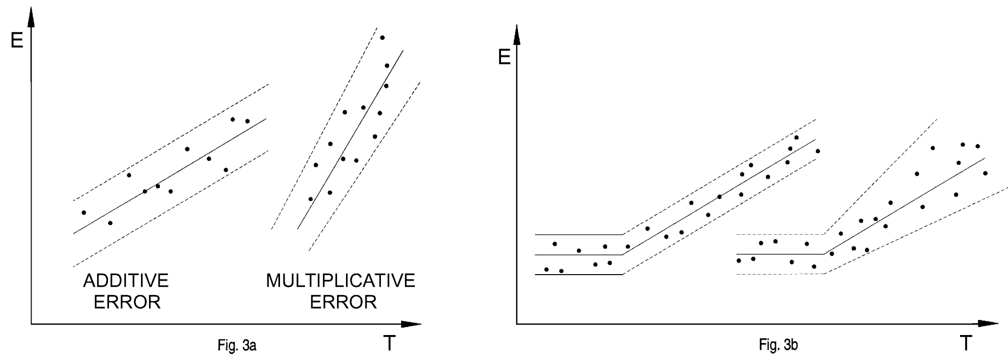

Figure 3. Conceptual plots to illustrate additive and multiplicative errors in linear and change-point three parameter-cooling models

Appendix A presents an unsuccessful attempt to renormalize the R^2 and CV indices, which though suitable for linear models, has been found to be misleading for the four-parameter change point models. Another intuitive approach is to calculate the CV (i.e., either the CVRMSE or the CVSTD as appropriate) of each of the two segments of the change point models separately from Equation (2) and weight them appropriately in order to obtain a weighted CV. This is addressed later on in this paper.

An underlying concept in model prediction uncertainty has to do with the type of noise present in the energy use observations, whether additive or multiplicative (Draper and Smith 1981). For a simple linear (2-P) model as shown in Figure 3 (a), the model form would be as follows:

$$\text{Additive noise:} E = (a + bT) + \alpha R \quad (4)$$

$$\text{Multiplicative noise:} E = (a + bT)(1 + \alpha R) \quad (5)$$

where R is a normally distributed random number with zero mean and unit standard deviation, and α is a multiplier, which can be used to inflate or deflate the “error” effect. Note that the two would be identical for a mean model (i.e., $b = 0$) with $a = 1$. Consequently for a three parameter (3-P) model, as shown in Figure 3b, we would have the same data spread about the mean line for the left-hand portion of the model line for both types of random errors. Prior studies indicate that there seems to be no consistency as to which type of error occurs in energy use data (Reddy et al. 1997a). Both these cases will be separately considered in the simplified expressions for $\Delta E_{save}/E_{save}$ developed in the next section.

RATIONAL METHODOLOGY FOR BASELINE MODEL EVALUATION

The primary objective of this paper is to develop a procedure for assessing the soundness of a baseline model for subsequent savings determination. The concept of a universal or absolute statistic is extended to view the issue in terms of some of the physical considerations of the problem. A major consideration is the relative impact of the retrofit on the baseline energy use. Both issues of goodness of fit of the baseline model and the expected fractional reduction in energy use due to the ECM are logical considerations. Though a certain amount of subjectivity is bound to be associated with the limits of model acceptability in such a paradigm, the interactions between the baseline model identification and the subsequent objective of determining energy retrofit savings needs to be recognized. The uncertainty arising from a model identified from

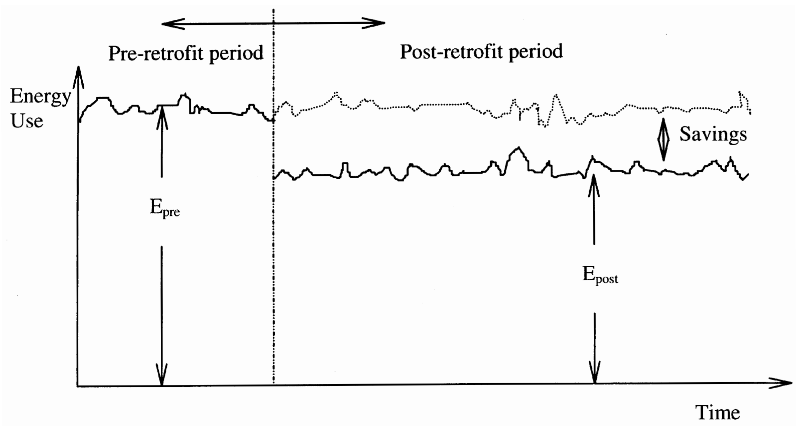

Figure 4. Conceptual plot showing how energy savings are determined

operational data that does not cover a whole annual cycle was not considered (Reddy et al. 1998). Hence the baseline model is assumed to be identified from yearlong data with no model extrapolation errors.

Conceptually, as shown in Figure 4, actual savings (as against "normalized" savings) E_{save} over m periods (hours, days or months, depending on the type of energy use data available) into the retrofit period are calculated as follows:

$$\sum_{j=1}^{m} E_{save,j} = \sum_{j=1}^{m} \hat{E}_{pre,j} - \sum_{j=1}^{m} E_{meas,j} \tag{6}$$

where

m = number of periods (hour, day, or month) in the post-retrofit period
$\hat{E}_{pre}$ = pre-retrofit energy use predicted by the baseline model per period, and
E_{meas} = measured post-retrofit energy use per period.

In order to make the equations more readable, we shall rewrite Equation (6) as:

$$E_{save,m} = \hat{E}_{pre,m} - E_{meas,m} \tag{7}$$

where, E_m now denotes the energy use over m periods. For example, $\hat{E}_{pre,m}$ denotes the sum of m individual model predicted values of baseline energy use.

With the assumption that model prediction and measurement errors are independent, the total variance is the sum of both:

$$(\Delta E_{save,m})^2 = (\Delta \hat{E}_{pre,m})^2 + (\Delta E_{meas,m})^2 \tag{8}$$

The total prediction uncertainty increases with m, i.e., as the post-retrofit period gets longer. However, as the amount of energy savings also increases with m, a better indicator of the uncertainty is the fractional uncertainty defined as the energy savings uncertainty over m periods divided by the energy savings over m periods:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \left[\frac{(\Delta\hat{E}_{pre,m})^2 + (\Delta E_{meas,m})^2}{(\hat{E}_{pre,m})^2 F^2}\right]^{0.5} \tag{9}$$

where F is the ratio of energy savings to pre-retrofit energy use, i.e.,

$$F = (\hat{E}_{pre,m} - E_{meas,m}) / \hat{E}_{pre,m} \tag{10}$$

Equation (9) provides a means of calculating the fractional uncertainty in the "actual" savings, which consists of a term representative of the regression model prediction uncertainty and another term representative of the measurement error in the post-retrofit energy use. The measurement error in the pre-retrofit energy use is inherently contained in the model goodness-of-fit parameter (namely, the mean square error (MSE) statistic), and should not be introduced a second time.

In case the measurement uncertainty is small (for example, when electricity is the energy channel its associated error is of the order of 1 to 2%), the fractional uncertainty in our savings measurement can be simplified into

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{\Delta(m\bar{E}_{pre})^{0.5}}{m\bar{E}_{pre}F} \tag{11}$$

where $\bar{E}_{pre}$ is the mean pre-retrofit energy use during the selected period.

This expression can be cast into a more useful form by making certain simplifying assumptions. The effect of measurement errors in the post-retrofit data was neglected and it was assumed that model prediction errors are the only source of prediction uncertainty (the various sources of errors as applied to building energy analysis are described by Reddy et al. 1998). Two separate cases were considered: (a) models with uncorrelated residuals which one would encounter when analyzing utility bills, and (b) models with correlated residuals often encountered with models based on hourly or daily data (Ruch et al. 1999).

Models with Uncorrelated Residuals and Additive Errors

The prediction uncertainty of a simple linear model identified from random data is given in statistical textbooks (Draper and Smith 1981). The regression model prediction uncertainty for an individual observation during the post-retrofit period, for a model with constant variance and uncorrelated model residual behavior, is:

$$\Delta\hat{E}_j = \left[\mathrm{MSE}\left(1 + \frac{1}{n} + \frac{(T_j - \bar{T})^2}{\sum_{i=1}^{n}(T_i - \bar{T})^2}\right)\right]^{0.5} \tag{12}$$

where

MSE = variance of the model error = $\sum_{i=1}^{n}(E_i - \hat{E}_i)^2/(n-p)$ (13)

T_j = individual value of outdoor dry-bulb temperature during the prediction or post-retrofit period

$\bar{T}$ = mean value of T_i (i.e., mean value of the outdoor temperature) during model identification or pre-retrofit period

The retrofit savings methodology is not based on individual predictions of $\hat{E}_j$, but the sum over *m* days of the $\hat{E}_j$ values. The prediction error of a sum of m future observations (as is needed for determining energy savings) is given by Theil (1971):

$$\Delta\hat{E}_m = \left\{ \mathrm{MSE}[\mathbf{1}'(\boldsymbol{X}_{post}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'_{post} + \boldsymbol{I})\mathbf{1}] \right\}^{0.5} \tag{14}$$

where $\boldsymbol{X}$ is the matrix of regressor variables during the pre-retrofit period, $\boldsymbol{X}'$ denotes the transpose of $\boldsymbol{X}$, $\boldsymbol{I}$ is an identity matrix, and the subscript "post" indicates post-retrofit period. Note that pre and post multiplication of the matrix within the square brackets by unit matrices **1** is akin to summing all the elements of the matrix.

Equation (14) involves matrix algebra and was simplified using numerical trials to form the following equation that holds to within about 10% accuracy:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{1.26}{m\bar{E}_{pre}F}\left[\mathrm{MSE}\left(m + \frac{m}{n} + \frac{\sum_{j=1}^{m}(T_j - \bar{T}_j)^2}{\sum_{i=1}^{n}(T_i - \bar{T}_i)^2}\right)\right]^{0.5} \approx \frac{1.26}{m\bar{E}_{pre}F}\left[\mathrm{MSE}\left(m + \frac{m}{n} + \frac{m}{n}\right)\right]^{0.5} \tag{15a}$$

where

n, m = number of observations in the baseline and the post-ECM periods, respectively

$\bar{T}_j, \bar{T}_i$ = average outdoor dry-bulb temperature values during post-ECM and during model identification (i.e., pre-ECM) periods, respectively.

Finally, for baseline models with additive errors, Equation (15a) can be expressed in terms of the standard CV statistic (either the CVRMSE or the CVSTD as appropriate) as:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{1.26\mathrm{RMSE}\left(1 + \frac{2}{n}\right)^{0.5}}{m^{1/2}\bar{E}_{pre}F} = \frac{1.26\mathrm{CV}\left[\left(1 + \frac{2}{n}\right)\frac{1}{m}\right]^{0.5}}{F} \tag{15b}$$

$$\approx \frac{1.26\mathrm{CV}}{m^{1/2}F} \qquad \text{when } n \text{ is large (say, } n > 60) \tag{15c}$$

Note that the above expression yields the fractional energy savings uncertainty at one standard error (i.e., at 68% confidence level). For other confidence levels, say 95%, the bounds have to be multiplied by the student t-statistic evaluated at 0.025 significance level and (*n-p*) degrees of freedom (Draper and Smith 1981).

Models with Uncorrelated Errors and Multiplicative Errors

As illustrated in Figure 3, model residuals will provide the necessary indication as to whether the errors are additive or multiplicative. Reddy et al. (1997a) had proposed the following empirical modification to Equation (15) for the prediction uncertainty $\Delta E_{save,m}$ in the case of multiplicative errors:

$$\Delta E_{save,m} = 1.26\left[(\mathrm{MSE}_1 m_1 + \mathrm{MSE}_2 m_2) + \mathrm{MSE}\frac{2m}{n}\right]^{0.5} \tag{16}$$

where m_1 and m_2 are the number of data points on either side of the change point during the post-retrofit period, MSE_1 and MSE_2 are the mean square errors of the data points on either side of the change point as shown in Figure 2, and MSE is given by Equation (13). MSE_1 (and MSE_2 by analogy) can be calculated from pre-retrofit data:

$$\mathrm{MSE}_1 = \left[\sum_{i=1}^{n_1}(E_i - \hat{E}_i)^2_{pre} / n_1\right]^{0.5} \tag{17}$$

Finally, the fractional uncertainty in the annual savings is:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{1.26\mathrm{CV}\left[\left(\frac{\mathrm{MSE}_1 m_1 + \mathrm{MSE}_2 m_2}{\mathrm{MSE}m} + \frac{2}{n}\right)\frac{1}{m}\right]^{0.5}}{F} \tag{18}$$

which, to be consistent with Equation(15b), can be re-written as:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{1.26\mathrm{WCV}\left[\left(1 + \frac{2}{n}\right)\frac{1}{m}\right]^{0.5}}{F} \tag{19}$$

where WCV is termed the weighted CV value for a change point model with multiplicative errors, and is defined as:

$$\mathrm{WCV} = \mathrm{CV}\left[\frac{\left(\frac{\mathrm{MSE}_1 m_1 + \mathrm{MSE}_2 m_2}{\mathrm{MSE}m} + \frac{2}{n}\right)}{\left(1 + \frac{2}{n}\right)}\right]^{0.5} \tag{20}$$

Defining the WCV is this manner allows the uncertainty $\Delta E_{save,m}$ of change point baseline models with either additive or multiplicative errors to be treated in the same manner. In the remainder of the paper, the term CV to denote either the CV or the WCV is used as appropriate.

Models with Correlated Residuals

Equations (15) or (19) are appropriate for regression models without serial correlation in the residuals. This would apply to models identified from utility (i.e., monthly) data. When models are identified from hourly or daily data, previous studies (Ruch et al. 1999) have shown that serious autocorrelation often exists. These autocorrelations may be due to (1) “pseudo” patterned random behavior due to the strong autocorrelation in the regressor variables (for example, outdoor temperature from one day to the next is correlated), or (2) to seasonal operational changes in the building and HVAC system not captured by an annual model. Consequently, the uncertainty bands have to be widened appropriately. Accurate expressions for doing so have been proposed by Ruch et al. (1999). In the framework of this study, a number of numerical trials were performed and found that the following simplified treatment yields results accurate to within 20% of those proposed by Ruch et al. (1999).

From statistical sampling theory, the number of *independent* observations n' of n observations with constant variance but having a lag 1 autocorrelation ρ equals (Neter et al. 1989):

$$n' = n\frac{1-\rho}{1+\rho} \tag{21}$$

For example, if $n = 365$ (i.e., one year of daily data), and $\rho = 0.85$, then $n' = 365 \times 0.081 = 29.6$, which implies that there were effectively only about 30 *independent* observations. Equation (15) in the presence of serial autocorrelation can be expressed as:

$$\frac{\Delta E_{save,m}}{E_{save,m}} = \frac{1.26\mathrm{CV}\left[\frac{n}{n'}\left(1+\frac{2}{n'}\right)\frac{1}{m}\right]^{0.5}}{F} \tag{22}$$

The CV value calculated using n degrees of freedom has been renormalized by the new degrees of freedom n'.

Discussion

Equations (15), (19) and (22) provide a more rational means of evaluating the accuracy of our baseline model to determine savings. Figure 5 that has been generated based on Equation (15b) allows easy determination of fractional savings uncertainty. Consider a change point baseline model based on daily energy use measurements with a CV (or WCV) of, for example, 10%, we wish to assess the uncertainty in savings for 6 months into the post-retrofit period for a retrofit measure that is supposed to save 10% of the pre-retrofit energy use (i.e., $F = 0.1$). Then for m = 6 months = 182.5 days, and CV = 10%, the y-ordinate value from Figure 5 is 0.01, i.e., $\Delta E_{save}/E_{save} = (0.01/0.1) \times 100 = 10\%$. On the other hand, if a baseline model is relatively poor, with, for example, a CV of 30%, and the retrofit is supposed to save 40% of the preretrofit energy use, then from Figure 5, $\Delta E_{save}/E_{save} = (0.028/0.4) \times 100 = 7.0\%$. Hence, the first model, despite having a CV value three times lower than that of the second model, leads to a larger fractional uncertainty in the savings than the second case. The above example serves to illustrate how viewing the retrofit savings problem in the perspective outlined in this paper is more relevant than merely looking at the baseline model goodness-of-fit. This example is based on 68% confidence level (i.e., t-statistic = 1). Suitable corrections need to be made for different confidence levels.

The variation of the fractional uncertainty $\Delta E_{save}/E_{save}$ with CV, n, m, and F is of interest to energy managers and energy service companies while negotiating energy conservation service contracts. If utility bills are the means of savings verification and if yearlong pre- and post-retrofit data are available, then Table 1 provides an indication of how $\Delta E_{save}/E_{save}$ varies with CV (or WCV) and F. For example, if both negotiating parties are comfortable with a fractional uncertainty in energy savings of 10% at the 68% confidence level, and if the retrofits are expected to save 20% (i.e., $F = 0.2$), then a baseline model with a CV of 5% or less will satisfy the expectations of savings verification when one year of pre-retrofit and one year of post-retrofit data are available. Table 2 provides the same information as Table 1 but assuming that daily monitored data is available for savings verification (i.e., we now have 365 data points instead of 12 data points only during either period) and that no residual autocorrelation is present (i.e., $\rho = 0$). For the above illustrative case, baseline models with up to 30% CV will still prove satisfactory.

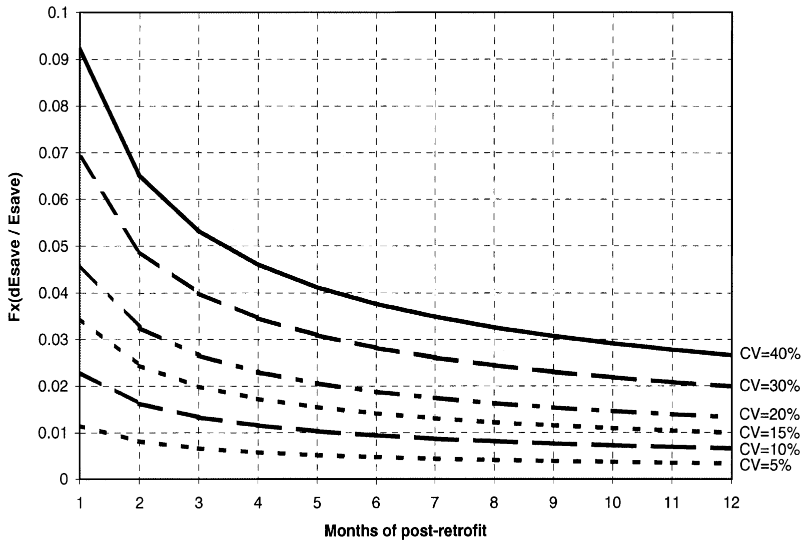

Figure 5. Variation of fractional uncertainty in savings
[Given by Equation (15b) at one standard error when daily measured energy values are available for post-retrofit period assuming no residual auto-correlation]

Table 1. Fractional Energy Savings Uncertainty
[At 68% confidence level after one year as given by Equation (15b) with baseline model identified from yearlong utility bills ($m = n = 12$) and no residual autocorrelation]

CV	***F***				
	0.05	**0.1**	**0.2**	**0.3**	**0.4**
0.05	0.393	0.196	0.098	0.065	0.049
0.1	0.786	0.393	0.196	0.131	0.098
0.15	1.179	0.589	0.295	0.196	0.147
0.2	1.571	0.786	0.393	0.262	0.196
0.25	1.964	0.982	0.491	0.327	0.246
0.3	2.357	1.179	0.589	0.393	0.295

Table 2. Fractional Energy Savings Uncertainty
[At 68% confidence level after one year as given by Equation (15b) with baseline model identified from yearlong daily monitored data ($m = n = 365$) and no residual autocorrelation]

CV	***F***				
	0.05	**0.1**	**0.2**	**0.3**	**0.4**
0.05	0.066	0.033	0.017	0.011	0.008
0.1	0.132	0.066	0.033	0.022	0.017
0.15	0.198	0.099	0.050	0.033	0.025
0.2	0.265	0.132	0.066	0.044	0.033
0.25	0.331	0.165	0.083	0.055	0.041
0.3	0.397	0.198	0.099	0.066	0.050

Table 3. Product of Fractional Energy Savings Uncertainty and Ratio of Energy Savings to Baseline Energy Use

[At 68% confidence level and savings fraction after one year as given by Equation (22) with baseline model identified from year-long daily monitored data ($m = n = 365$)]

CV	ρ				
	0	**0.5**	**0.75**	**0.85**	**0.95**
0.05	0.003	0.006	0.009	0.012	0.023
0.1	0.007	0.012	0.018	0.024	0.045
0.15	0.010	0.017	0.027	0.036	0.068
0.2	0.013	0.023	0.036	0.048	0.091
0.25	0.017	0.029	0.044	0.060	0.113
0.3	0.020	0.035	0.053	0.072	0.136

If the model residuals exhibit auto-correlated behavior, then Equation (22) should be used. Since an extra variable, namely ρ is present, it is better to look at how the variable $[(\Delta E_{save,m}/E_{save,m})\cdot F]$ varies with ρ (Table 3). The numbers below ρ = 0 have a direct correspondence to those given in Table 2. For example, if CV = 0.2, from Table 3, $(\Delta E_{save,m}/E_{save,m})\cdot F = 0.013$. Further, if $F = 0.2$, then $\Delta E_{save,m}/E_{save,m} = 0.013/0.2 = 0.065$, which is almost identical to the value of 0.066 shown in Table 2 for CV = 0.2 and $F = 0.2$.

As shown in Table 3 the term $[(\Delta E_{save,m}/E_{save,m})\cdot F]$ increases as ρ increases. For example, for the same case of CV = 0.2 and $F = 0.2$, and for ρ = 0.85, $\Delta E_{save,m}/E_{save,m} = 0.048/0.20 = 0.24$, which is almost four times the value found when ρ = 0. The previous discussion illustrates how the time scale of data and the presence of serial autocorrelation affect fractional savings uncertainty and have direct bearing on energy savings verification.

APPLICATION TO DATA

In order to illustrate the above concepts, six U.S. Army bases (AB) whose monthly utility data in the form of electricity use and gas use consumed by the entire base were available for a complete year, were selected. The data for these Army bases, which are geographically dispersed in the continental U.S., were provided by the U.S. Army Construction Engineering Research Laboratories in Champaign, IL. Reddy et al. (1997b) provides specific details concerning the Army bases, the years during which data was provided, how the data was screened, and the year chosen as the baseline.

Table 4 presents the salient results of the analysis. Various linear and change-point models were fit to the utility bill data for all six bases. The models that yielded the best fits are shown in Table 4. Other than electricity use for AB1 which is best fit by a 4-P model and two other bases which are best modeled by mean models, all other models are 3-P. Whether the model is a cooling model or a heating model is also indicated in column 2 by C and H respectively. The R^2 and CV as calculated by Equations (1) and (2) are shown in columns 3 and 4. The number of data points for each Army base that fall on either side of the change point are shown under n_1 and n_2. The corresponding mean square error (MSE) calculated from Equation (17) as well as the associated CVs are also tabulated. Other than gas use for AB3, which seems poor, all other models seem satisfactory. As discussed earlier, there are certain models which have similar CVs and different R^2 values, like electricity use for AB3 which has a low R^2 (of 0.66) and a CV of 10.7%, while gas use for AB4 has a similar CV but a R^2 of 0.98. From the point of view of model prediction uncertainty in energy savings estimate, both models are as good provided F values for both cases are the same.

Differences between the CV of the two segments fitted with two individual models (i.e., CV_1 and CV_2) yield additional insight into whether the errors are additive or multiplicative. The CVs

Table 4. Application to Utility Bill Energy Use Data from Six Army Bases in the United States
(Adapted from Reddy et al. 1997a)

Army Base	Model Type	R^2	CV, %	n_1	n_2	MSE_1 *, +	MSE_2 *,+	CV_1 %	CV_2 %	WCV %
					Electricity					
AB1	4P-HC	0.95	4.7	6	6	3.754	1.568	5.37	2.82	5.80
AB2	Mean	—	8.0	—	—	—	—	—	—	8.00
AB3	3P-H	0.66	10.7	7	5	3.595	3.861	8.85	11.30	9.90
AB4	3P-C	0.98	5.2	5	7	0.363	7.165	1.93	5.07	4.55
AB5	3P-C	0.66	7.1	6	6	2.503	4.623	5.83	6.91	6.71
AB6	Mean	—	8.3	—	—	—	—	—	—	8.30
					Gas					
AB1	3P-H	0.87	14.4	7	5	293.0	200.9	11.66	15.40	13.27
AB2	3P-H	0.94	13.7	9	3	4871	430.6	10.71	10.48	11.56
AB3	3P-H	0.69	36.5	7	5	77.17	201.1	19.79	72.08	33.77
AB4	3P-H	0.98	10.1	7	5	551.8	63.00	8.24	9.11	9.34
AB5	3P-H	0.93	18.4	8	4	1428	15.32	13.02	5.30	13.92
AB6	3P-H	0.69	16.2	9	3	1430	87.25	12.65	5.74	11.54

H = heating only, C = cooling only, HC = heating and cooling
* Units for electricity are in kWh/day per 1000 ft^2 + Units for gas are in ft^3/day per 1000 ft^2

of the two individual segments of the change point models may be widely different by as much as a factor of 3 in some cases (Table 4). For example, both the CVs for gas use at AB2 are close indicating that errors are probably additive over the entire range of variation. A wide difference between both values, such as for gas use at AB3, indicates that the model for region (2) is much poorer than region (1) (see Figure 2) and are due to errors being probably multiplicative. Thus, the nature of how the errors in the data affect the model differs from one Army base to another. The last column of Table 4 presents the WCV values as calculated from Equation (20) with m, m_1, and m_2 replaced by n, n_1, and n_2, respectively. Note that n = 12 in this case because year long utility bill data are considered. The WCV values are generally slightly lower than the CV values (column 4), but this difference is usually small. This suggests that the user could simply use the CV value of the entire model if a preliminary estimate of the savings uncertainty fraction is required. If the CVs of both segments are close, then the analyst considers the errors to be additive and uses Equation (15) along with the overall CV to determine fractional uncertainty in savings. If the CVs are widely different, the WCV values have to be computed from Equation (20) and this value is used in Equation (15) for the CV.

IMPLICATIONS TOWARDS REQUIRED LEVEL OF MONITORING AND VERIFICATION

A situation may exist where the energy manager of a certain facility wishes to reduce the energy bill by having certain ECMs performed by an energy service company. It is assumed that utility bills of the facility are available for at least one year but that no sub-metered data is available. The statistical expression given by Equation (15) or Equation (19) depending on whether the model residuals are additive or multiplicative, can be used to directly provide an indication as to whether sub-metering is necessary or not.

As discussed earlier, the selection of the particular value of $\Delta E_{save,m}/E_{save,m}$ depends on the energy manager and the ESCO while negotiating the energy conservation contract. It is assumed that both parties have reached a mutually agreeable value and that the energy savings fraction F has also been determined by an audit. Equation (15b) is then used to determine the maximum CV (called CV_{max}) of the baseline model. A model is then fit to the utility bills of the past year. If the model CV < CV_{max}, then no sub-metering is necessary and the required savings verification could be done from utility bill analysis alone. If this is not the case, then some sort of sub-monitoring is required. Let us illustrate this with the data of the Army bases shown in Table 4.

The values $\Delta E_{save,m}/E_{save,m} = 0.15$, and $F = 0.2$ are assumed for illustration. Then for one year of pre-retrofit and one year of post-retrofit, $n = m = 12$. These are substituted in Equation (15b), $CV_{max} = 7.6\%$. From Table 4, for only two cases (electricity use at AB1 and AB4) are the WCV values sufficiently less than 7.6% so as not to require sub-monitoring. There are five more models with CV values within plus/minus two percentage points, and whether sub-monitoring is required or not is negotiable. In the remaining five cases, there is a definite need to perform sub-monitoring if the desired confidence in the resulting savings is to be satisfied.

CONCLUSIONS

A previous study suggested absolute criteria for the model goodness-of-fit to ascertain whether a baseline model is adequate to subsequently determine savings due to energy conservation measures (Reynolds and Fels 1988). The contention in this paper is that baseline model development is not an end in itself, and that explicit consideration should be given to the relative impact of the retrofit on the baseline energy use. Under this paradigm, we argue that there will be no absolute statistical cutoff criteria for model goodness-of-fit indices. Subject to certain simplifications that are intuitively appealing, expressions for the fractional uncertainty of energy savings in terms of model CV and the relative impact of the retrofit on the baseline energy use for both the case of a model with uncorrelated residuals (as encountered in utility bill analysis) as well as for correlated residuals (as often encountered with models based on hourly or daily data) are proposed. The modification to the standard CV statistic so as to be applicable to change point models with multiplicative errors is presented. The effect of the two different types of model residual errors, additive and multiplicative, on fractional uncertainty of energy savings has also been explicitly treated. How these concepts would logically lead to insights into the selection of the most appropriate monitoring and verification protocol to verify savings has also been discussed and illustrated with energy use utility bill data from several Army bases nation-wide. The approach suggested in this paper for determining the uncertainty in the savings due to an ECM should be useful to energy service companies and professionals engaged in the ever-expanding field of energy conservation in buildings.

ACKNOWLEDGMENTS

This research was performed as part of the Texas LoanSTAR Monitoring and Analysis Program funded by the State Energy Conservation Office of the State of Texas. Stimulating discussions with Dr. J. Hebert of Sam Houston State University are acknowledged. An insightful review by G. Reeves is also highly appreciated. The Army Base data has been taken from a former project funded by the U.S. Army Construction Engineering Research Laboratories (CERL). Finally, we would like to thank S. Deng for his help in generating some of the figures.

NOMENCLATURE

- E energy use, baseline energy use
- F ratio of the energy savings to baseline energy use
- m number of post-retrofit observation points (months, days, hours,...)
- n number of pre-retrofit observation points (months, days, hours,...)
- p number of model parameters in the regression model
- T outdoor dry-bulb temperature

t t-statistic
$\bar{X}$ mean value of $\boldsymbol{X}$
$\hat{X}$ model predicted value of $\boldsymbol{X}$
ρ auto-correlation coefficient
CV coefficient of variation (either the CVRMSE or the CVSTD)
WCV weighted CV defined by Equation (20)
R^2 coefficient of determination
ΔE uncertainty in E
MSE mean square error
STD standard deviation

Subscripts

post post-retrofit
pre pre-retrofit or baseline
save saving

REFERENCES

Beale, E. 1960. Confidence Regions in Non-Linear Estimation. *J.R. Stat. Assoc.* B-22: 1-88.

Draper, N., and H. Smith. 1981. *Applied Regression Analysis*, 2nd Ed. John Wiley & Sons, New York.

Fels, M.F. (Ed.). 1986. Special Issue devoted to Measuring Energy Savings, The Princeton Scorekeeping Method (PRISM). *Energy and Buildings* 9(1) and 9(2).

Fels, M.F., K. Kissock, M. Marean, and C. Reynolds. 1995. PRISM (Advanced Version 1.0) Users' Guide. Center for Energy and Environmental Studies, Princeton University, Princeton, NJ.

Goldberg, M. 1982. A Geometric Approach to Nondifferentiable Regression Models as Related to Methods for Assessing Residential Energy Conservation. Ph.D. Thesis, Department of Statistics, Princeton University, Princeton, NJ.

Hinkley, D.V. 1969. Inferences about the Intersection on Two-Phase Regression. *Biometrica* 56: 495-504.

IPMVP. 1997. International Performance Measurement and Verification Protocol. Department of Energy, DOE/EE-0157, Washington DC.

Katipamula, S., T.A. Reddy, and D.E. Claridge. 1998. Multivariate Regression Modeling. *ASME Journal of Solar Energy Engineering* 120(9): 177.

Kissock, J.K., T.A. Reddy, and D.E. Claridge. 1998. Ambient-temperature Regression Analysis for Estimating Retrofit Savings in Commercial Buildings. *ASME Journal of Solar Energy Engineering* 120(8): 168.

Neter, J., W. Wasserman, and M.H. Kutner. 1989. *Applied Linear Regression Models*, 2nd Ed. Irwin, Homewood.

Reddy, T.A., N.F. Saman, D.E. Claridge, J.S. Haberl, W.D. Turner, and A. Chalifoux. 1997a. Baselining Methodology for Facility-Level Monthly Energy Use—Part 1: Theoretical Aspects. *ASHRAE Transactions* 103(2).

Reddy, T.A., N.F. Saman, D.E. Claridge, J.S. Haberl, W.D. Turner, and A. Chalifoux. 1997b. Baselining Methodology for Facility-Level Monthly Energy Use—Part 2: Application to Eight Army Installations. *ASHRAE Transactions* 103(2).

Reddy, T.A., J.K. Kissock, and D.K. Ruch. 1998. Uncertainty in Baseline Regression Modeling and in Determination of Retrofit Savings. *ASME Journal of Solar Energy Engineering* 120(8): 185.

Reynolds, C., and M. Fels. 1988. Reliability Criteria for Weather Adjustment of Energy Billing Data. *Proceedings of the 1988 ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 10.236-51.

Ruch, D.K., and D.E. Claridge. 1993. A Development and Comparison of NAC Estimates for Linear and Change-Point Energy Models for Commercial Buildings. *Energy and Buildings* 20: 87-95.

Ruch, D.K., J.K. Kissock, and T.A. Reddy. 1999. Prediction Uncertainty of Linear Building Energy Use Models with Autocorrelated Residuals. *ASME Journal of Solar Energy Engineering* 121(2): 121.

Thiel, H. 1971. *Principles of Econometrics*. John Wiley & Sons, New York.

APPENDIX A

We shall present an unsuccessful attempt to redefine some of the standard goodness-of-fit criteria so that different models could be compared on an absolute basis. In order to be able to compare models with different slopes and magnitude, we may consider a simple linear transformation such as:

$$x' = (x - x_{min})/(x_{max} - x_{min}) \text{ and } y' = (y - y_{min})/(y_{max} - y_{min}) \qquad \text{(A1)}$$

where the minimum and maximum points of a linear model are as shown in Figure A-1a.

In physical terms, this normalization implies that the data points have been collapsed to fit into a square with the origin as one end point and (1,1) as the other (see Figure A1b). If we were to replace the y terms in Equation (1) by the above definition of y', define $\Delta y = y_{max} - y_{min}$ and $y^* = y_{min}/(y_{max} - y_{min})$, and assume that regression is based on the least square criterion, we obtain the following:

$$\begin{aligned} \text{Modified } R^2 &= 1 - \sum (y'_i - \hat{y}'_i)^2 / \sum (y'_i - \bar{y}')^2 \\ &= \frac{1 - \sum\left(\frac{y_i}{\Delta y} - y^* - \frac{\hat{y}_i}{\Delta y} + y^*\right)^2}{\sum\left(\frac{y_i}{\Delta y} - y^* - \frac{\bar{y}}{\Delta y} + y^*\right)^2} \\ &= 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2 \end{aligned} \quad \text{(A2)}$$

This expression is the same as Equation (1) and so we can conclude that the above linear transformation of the data does not affect the value of R^2. Thus R^2 is independent of variable rescaling when Equation (A1) is used.

If we were to do the same with Equation (2), we find that the new or normalized CV (which we shall denote by NCV) depends on variable scaling. This can be shown as follows:

$$\begin{aligned} \text{NCV} &= \left[\sum\left(\frac{y_i}{\Delta y} - y^* - \frac{\hat{y}_i}{\Delta y} + y^*\right)^2 / (n-p)\right]^{0.5} \Big/ \left(\frac{\bar{y}}{\Delta y - y^*}\right) \\ &= [\sum (y_i - \hat{y}_i)^2]^{0.5} / (\bar{y} - y_{min}) / (n-p)^{0.5} \\ &= \text{CV} \frac{\bar{y}}{(\bar{y} - y_{min})} \end{aligned} \quad \text{(A3)}$$

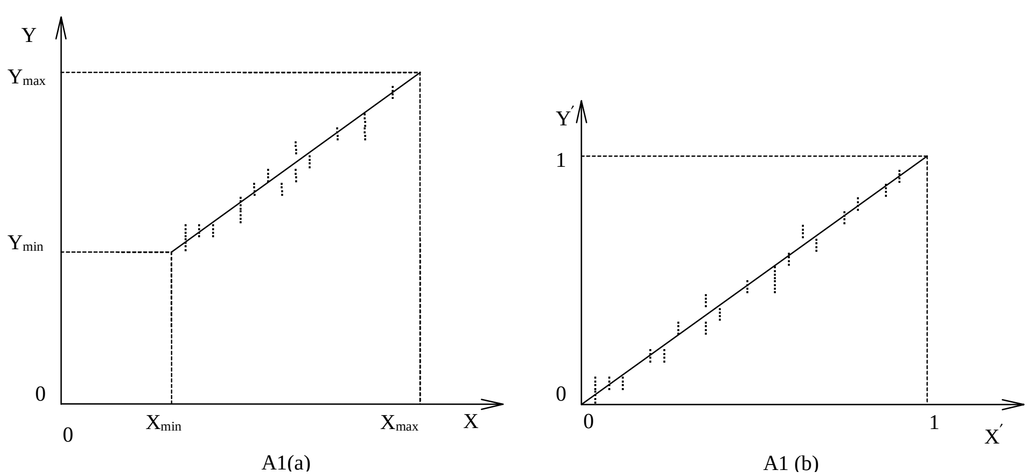

Figure 6. Transformation of linear model to fit into square with origin (0,0) and (1,1) as other end point

In order to remove the undue sensitivity to one data point only, viz. y_{min}, a more stable solution is to use the value of y_{min} as predicted by the identified model using $x = x_{min}$. Though the above transformation may have intuitive appeal, we found certain insurmountable limitations to this type of normalization while considering change point models.

Two major limitations have been identified with the type of transformation given by Equation (A1) when applied to change point models. For a 3-P model, the normalization does not make sense for the mean portion of the model. So the CVSTD rather than the CVRMSE should be used for the base portion. For 4-P models with one of the segments having a very low slope, the denominator of the normalization factor for y tends to be a very small number ($\bar{y} \approx y_{min}$), which artificially inflates the normalized CVRMSE measure. Consequently correspondence between different model types is lost and the same measure cannot be used to evaluate different model types in the same framework. Hence we do not recommend the above transformation.