#4806

# REGRESSION ANALYSIS OF RESIDENTIAL AIR-CONDITIONING ENERGY CONSUMPTION AT DHAHRAN, SAUDI ARABIA

**D.Y. Abdel-Nabi** **S.M. Zubair** **M.A. Abdelrahman** **V. Bahel**

## ABSTRACT

*The energy consumption of a house air conditioner located at Dhahran, Saudi Arabia, is modeled as a function of weather parameters and total (global) solar radiation on a horizontal surface. The selection of effective parameters that significantly influence energy consumption is carried out using general stepping regression methods. The problem of collinearity between the regressors is also investigated. The final model involves parameters of total solar radiation on a horizontal surface, wind speed, and temperature difference between the indoor and outdoor condition. However, the model coefficients are functions of relative humidity and/or temperature difference between the indoor and outdoor condition. Model adequacy is examined by the residual analysis technique. Model validation is carried out by the data-splitting technique. The sensitivity of the model indicates that relative humidity and temperature difference strongly influence the cooling energy consumption. It was found that an increase in relative humidity from 20% to 100% can cause a 100% increase in cooling energy consumption during the high cooling season.*

## INTRODUCTION

Present building technology in Saudi Arabia has evolved through the rapid economic development of the past two decades. During the course of this development, building practices of other countries were brought to the Arabian Gulf countries with little consideration for local design requirements and energy conservation. As a result, a typical new residence here consists of a well-built structure with heavy masonry construction materials and little or no insulation. Most residential buildings are generally equipped with oversized air-conditioning system(s). Some studies (Debs 1983; KFUPM 1984) have indicated that about 70% of the total residential electrical energy consumption in the Arabian Gulf region is used for space cooling of buildings. Any attempt to reduce this energy consumption should be preceded by a rigorous analysis of major factors affecting the thermal load of a building. These factors are: (i) the weather parameters in the location, (ii) the thermal characteristics of the building envelope, (iii) the tightness of the building envelope, (iv) the required indoor temperature and relative humidity, (v) the internal thermal loads, (vi) the schedules of air-conditioning and ventilation systems, and (vii) the electrical energy cost.

With the exception of the weather parameters and electrical energy cost, all these factors vary from one building to another. However, all of these factors are considered in detailed, hour-by-hour computer codes such as DOE-2 (LBL 1981). The use of this program is limited to those who have the financial support and access to computers, whereas the degree-day method is simple to use but does not provide the required accuracy. The degree-day method of energy analysis is an attractive approach to residential energy estimation when time and resources do not permit the use of computerized energy estimation procedures. This simplified method, which was first introduced in 1920, has been used for heating energy analysis but it has not been an accepted procedure for cooling energy estimation. This is because latent loads due to infiltration and occupants, and internal loads due to occupancy, lighting, equipment, and solar heat gain, are not only dependent on outdoor temperature. These loads, therefore, may not be accounted for in the degree-day method by computing the degree-days from the difference between some base temperature (i.e., zero load temperature) and the mean daily outdoor temperature. Recently, the degree-day method with a variable base temperature was used by Fels (1986) in a statistical model for calculating changes in energy consumption of a house.

The objective of this work is to investigate the effect of weather parameters and global (i.e., total horizontal) solar radiation on electric energy consumption of an air conditioner in a residential building at Dhahran, Saudi Arabia. The paper outlines the approach used and describes the results obtained by the use of multiple linear regression analysis techniques. The developed model is validated by comparison with the daily metered energy consumption for a 1076 $ft^2$ (96.5 $m^2$) residence located at Dhahran, Saudi Arabia, which has nearly year-round cooling requirements. Secondly, this study attempts to help fill a void in the literature on the energy consumption of residential building air conditioning as a function of the weather parameters and global solar radiation in a hot, humid climate.

## LOCATION AND CLIMATE

The region where Saudi Arabia borders the Arabian Gulf lies only a few degrees outside the tropics and extends from 24.8°N, 48.3°E to 28.7°N, 50.9°E (i.e., between 1.3° and 5° north of the Tropic of Cancer). Dhahran (26.32°N, 50.13°E) is located a few kilometers inland from the Gulf on the eastern coastal plain of Saudi Arabia. Although Dhahran is near the coast, it is located in a desert environment. Dhahran's climate, although extremely arid (approximately 80 mm total annual precipitation), is significantly influenced by the Gulf waters. However, the overall shallowness of the Gulf (average depth, 30 m), combined with the occurrence of the deepest waters near the Iranian coast against the Zagros mountain chain, results in the thermal control of regional weather by

**D.Y. Abdel-Nabi, S.M. Zubair, M.A. Abdelrahman**, and **V. Bahel** are Researchers in the Energy Systems Group, Division of Energy Resources, Research Institute, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia.

**TABLE 1**
**Climatic Temperature Data for Dhahran, Saudi Arabia, 1950 to 1976 Norms**

| Parameter | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Absolute Maximum** | 30 | 36 | 38 | 45 | 48 | 49 | 49 | 51 | 48 | 43 | 37 | 32 |
| **Average Daily Maximum** | 21 | 22 | 26 | 31 | 37 | 42 | 42 | 42 | 40 | 35 | 29 | 23 |
| **Mean** | 16 | 17 | 21 | 26 | 31 | 34 | 36 | 36 | 33 | 29 | 23 | 18 |
| **Average Daily Minimum** | 12 | 13 | 15 | 21 | 25 | 28 | 29 | 29 | 27 | 23 | 18 | 13 |
| **Absolute Minimum** | −1 | 3 | 5 | 10 | 18 | 22 | 23 | 24 | 21 | 16 | 8 | 3 |

the Gulf being much reduced relative to deeper water bodies of similar size, such as the Red Sea on the west coast of Saudi Arabia.

The climate of the Dhahran area is discussed in more detail in Williams (1979). The parameter of most general interest in Saudi Arabia is temperature. The temperature climate of Dhahran is summarized in Table 1, which lists means and extremes for the period 1950 to 1976 (Williams 1979). The absolute maximum temperature recorded at Dhahran was 51°C and the absolute minimum was -1°C. Average daily maxima range between 21° and 42°C and exceed 40°C for four months of the year. Average daily minima vary between 12°C in December and January and 29°C in July and August and reach below 20°C only during the months of November to March.

Despite the desert location of Dhahran, the nearness of the shallow Arabian Gulf induces high relative humidity. At Dhahran, relative humidity exhibits a large diurnal cycle (on the order of 60% throughout the year), with daily maxima often rising over the 80% level during most months. Wind speeds show a clear diurnal variability within the typical range, from near zero to 10 m/s. The synoptic wind direction exhibits long periods of more or less constant direction between north and northwest. An additional feature with some longevity is the tendency for the wind to swing to the east, in particular the quadrant between east and south.

## DATA PREPARATION AND COLLECTION

In an energy conservation program sponsored by a Saudi university, personnel instrumented a few houses on the university campus at Dhahran with watt-hour meters for monitoring the air conditioner electrical power consumption on a weekly basis. The data considered in this study are daily air conditioner electrical power consumption for a single-story residence (composed of two bedrooms, a living room, and a kitchen). During the monitored period, it was occupied by a family of two adults and a child. The layout of the house is shown in Figure 1. As can be seen from this illustration, the house is attached to a similar house on the southeast side. Also, the house is surrounded by a green fence with a height of about 1.5 m and at a distance of about 5 m from the external walls. A large umbrella-type tree is located in front of the house, providing a shadow over part of it during the day. Air conditioner energy consumption and indoor temperature data were collected on a daily basis. The other meteorology and solar radiation data were monitored at the institute's meteorology and solar radiation station on a minute-by-minute basis, although only daily averages are considered in this study. The variation of energy consumption per square meter with time (Julian day) is shown in Figure 2. As can be seen from this illustration, the gaps are due to vacation periods when the house is not occupied. As expected, the air conditioner energy consumption is highest during the summer period (March through October), whereas the energy consumption during the winter season (November through February) is minimal. This low energy use during the winter season is due to the fact that in Dhahran the average daily temperature rarely reaches below 10°C.

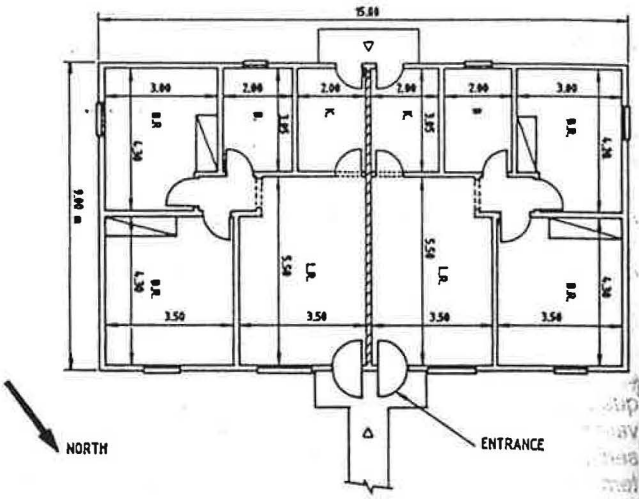


***Figure 1*** *House layout. Right-hand side of the house is the one considered in this study.*

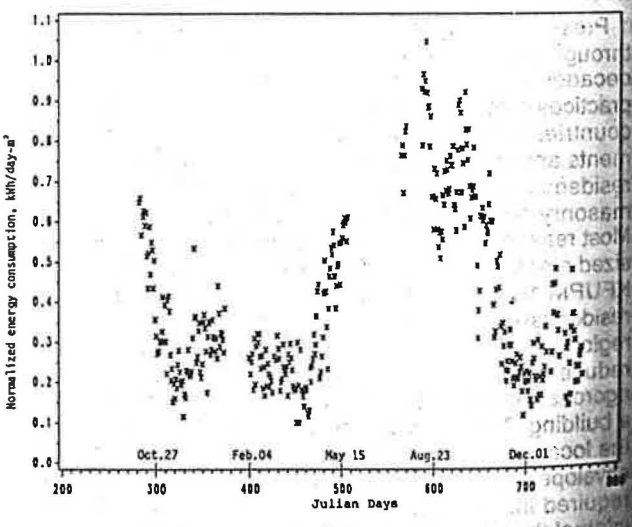


***Figure 2*** *Plot of normalized air conditioner energy consumption vs. time (Julian day)*

## MODEL SELECTION

To correlate residential air conditioner electrical power consumption with the weather parameters and global solar radiation, the dependent variable in the model is defined as the daily air conditioner energy consumption (EC), and the independent variables are the daily means of certain meteorological and radiation parameters. Hence, the preselection predictive model can be represented by the functional relationship:

$$EC_P = F(TA,TIN,WS,RH,THR) \quad (1)$$

where $EC$ is the measured energy consumption (kWh/day), $EC_P$ is the predicted energy consumption (kWh/day), $TA$ is the mean air temperature (°C), $WS$ is the mean wind speed (m/s), $RH$ is the mean relative humidity (%), $TIN$ is the mean indoor temperature (°C), and $THR$ is the mean global radiation (W·h/m²·day).

However, it is known that the temperature difference between indoor and outdoor conditions is a driving force in heat transfer across the wall; hence, for the present application, it has been assumed that the independent variables $TA$ and $TIN$ can be replaced by their difference $P1$ ($P1 = TA - TIN$). The other independent regressor variables may all be influential. In certain applications, theoretical considerations and/or prior experience can be useful in limiting the regressors for consideration in the model. However, for the problem under consideration, it is more appropriate, in the first instance, to use statistical techniques to perform this function. A number of approaches were employed to make the final variable selection. A statistical analysis package (SAS 1985) has been utilized for these tasks and for all other statistical analysis discussed in this paper.

It is important to note that the least-squares fitting techniques are heavily dependent on the coefficient of multiple determination, $R^2$, which is not a very good indicator of the model suitability. A systematic and powerful technique is described here as a practical guide for any type of energy use data. This technique is used to develop the air conditioner energy consumption correlation, which describes both heating and cooling periods, with the weather parameters and total solar radiation on a horizontal surface at Dhahran, Saudi Arabia. The procedure is detailed in Appendix A.

## ENERGY CONSUMPTION MODEL

As indicated by the procedure described in Appendix A, the best correlation of air conditioner electrical energy consumption data, for both heating and cooling periods, with the weather parameters and global solar radiation is given by:

$$EC_P = 18.243 - 3.803\text{E}-3 \cdot THR + 0.4268 \cdot WS + 8.715\text{E}-5 \cdot THR \cdot RH + 3.856\text{E}-2 \cdot P1 \cdot RH + 0.1208 \cdot P1 \cdot WS + 0.3452 \cdot P1^2 \quad (2)$$

The above air conditioner energy consumption model can be written as

$$EC_P = A + B \cdot THR + C \cdot WS + D \cdot P1 \quad (3)$$

*where*

$A = 18.243$
$B = -3.803\text{E}-3 + 8.715\text{E}-5 \cdot RH$
$C = 0.4268 + 0.1208 \cdot P1$
$D = 3.856\text{E}-2 \cdot RH + 0.3452 \cdot P1$

In the above model, $A$ is defined as the base load of the house, which is related to the indoor temperature settings as well as heat generated by appliances and occupants; $B \cdot THR$ is the contribution due to solar gain; $C \cdot WS$ is the infiltration load; and $D \cdot P1$ is the sensible and latent load of the house. It should be emphasized that the model coefficients $B$, $C$, and $D$ in Equation 3 are functions of relative humidity and/or temperature difference.

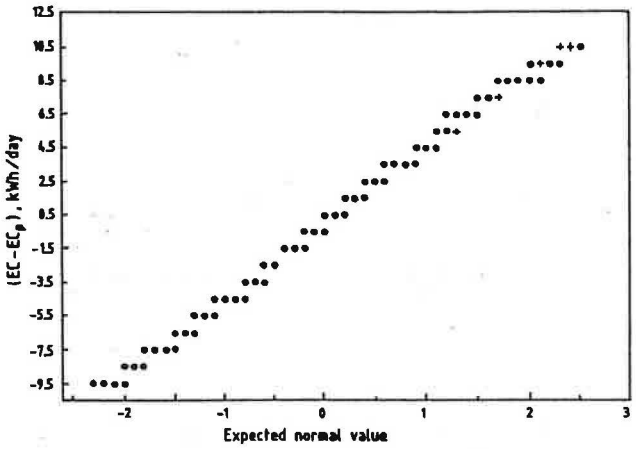


***Figure 3*** *Normal probability plot for residuals. Key: "+" represents the normal cumulative distribution, while "·" represents the residual cumulative distribution.*

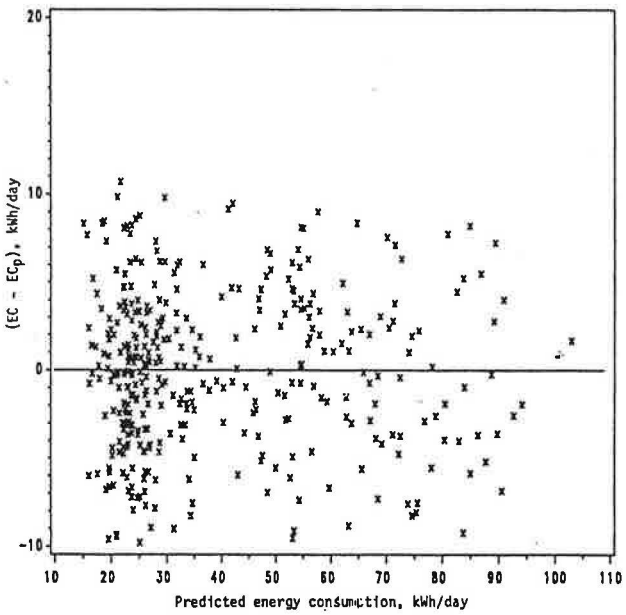


***Figure 4*** *Plot of residuals ($EC - EC_p$) vs. predicted air-conditioning energy consumption for the original data set*

## MODEL ADEQUACY

Evaluating model adequacy, which includes internal analysis to investigate the fit of the regression model to the available data, is an important part of any multiple-regression problem. Several methods can be used for this purpose (Montgomery and Peck 1982). Residual analysis was adopted for this study. The functional form of the multiple-regression model presented in Table A3 and discussed above is used to predict the power consumption over the domain of input data with the residuals (i.e., the differences between observed and predicted values of air conditioner energy consumption) being compared.

A normal probability analysis of the residuals was used to check the normality assumption (Figure 3); small departures from normality are statistically acceptable and do not cast the model into doubt. As can be seen from the figure, the points lie along a straight line (idealized type) and no obvious model inadequacies or defects are seen.

It is, however, important to consider the plots of the residuals vs. the corresponding predicted values of air conditioner

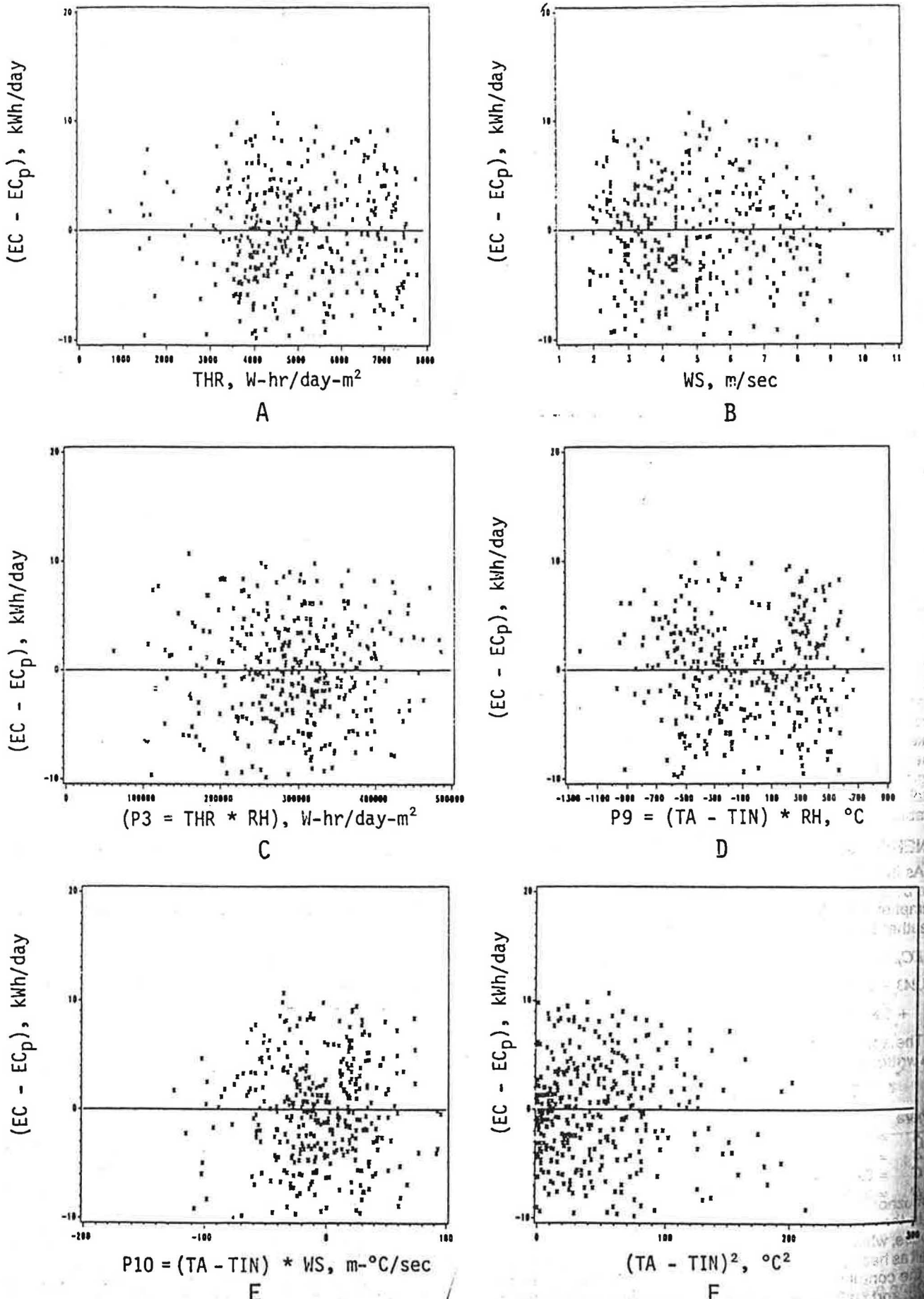


***Figure 5*** *Plot of residuals ($EC - EC_p$) vs. regressor variables. (a) global radiation, (b) wind speed, (c) crossproduct of global radiation and relative humidity, (d) crossproduct of temperature difference and relative humidity, (e) crossproduct of temperature difference and wind speed, and (f) temperature difference square.*

**TABLE 2**
**Summary Over a 150-Day Validation Period for the Prediction Model**

Also included are comparable numbers for the full domain (369 oberservations) used in formulating the model. Additionally listed results are from Perrone and Miller (1985) for forecast using the GEM and MOS techniques. Key: $AE$ absolute error (kWh/day), $GE$ algebraic error (kWh/day), $LE$ large errors (> kWh/day), $H_d$ modified hit rate (± 5.0 kWh/day).

| | $AE$ | $GE$ | $LE$ (%) | $H_d$ (%) |
|---|---|---|---|---|
| Model (369 days) | 3.7 | 0.3 | 0.3 | 71 |
| Model (150 days) | 3.7 | −3.8 | 0.7 | 72 |
| GEM | 2.8 | −0.9 | 1.2 | — |
| MOS | 3.0 | −0.3 | 2.4 | — |

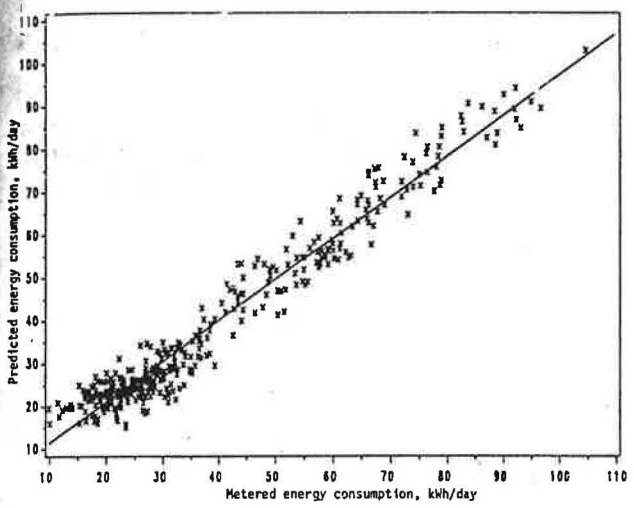


***Figure 6*** *Metered air conditioner electric power consumption (EC) vs. predicted power consumption ($EC_P$) over 369 days*

energy consumption for detecting several common types of model inadequacies. Such plots typically exhibit certain patterns (i.e., horizontal band, outward-opening, double bow, and curved band [Montgomery and Peck 1982]). The desirable pattern for a good model is one where the residuals are contained within a horizontal band. A plot of residuals vs. predicted values of air conditioner energy consumption is shown in Figure 4. This figure shows all the residuals to be contained within a horizontal band (idealized plot). There is no indication of any model defects.

Plotting residuals against the corresponding values of each regressor variable is also helpful in investigating model adequacy. These plots often show the patterns discussed above and once again an impression of a horizontal band containing the residuals is desirable. Plots of the residuals vs. all the effective regressor variables are also shown in Figure 5. No obvious model defects are revealed. Finally, measured and predicted values of the air conditioner power consumption rates are compared graphically in Figure 6 for the period from which the model was developed. This figure clearly indicates model suitability for its intended aim.

## MODEL VALIDATION

Model validation, which is different from model adequacy, aims to determine if the model will function successfully in its intended operating field. Three differing techniques can, in general, be used to validate a regression model (Montgomery and Peck 1982). These techniques are: collection of fresh data to investigate the model's predictive performance; anal-

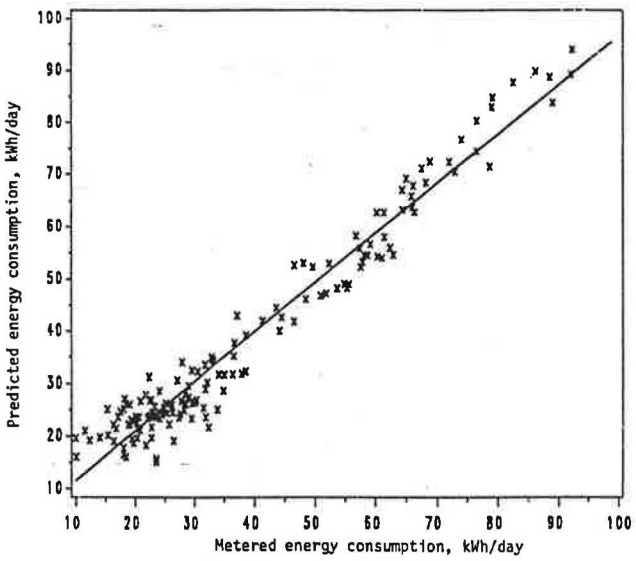


***Figure 7*** *Metered air conditioner electric power consumption (EC) vs. predicted values ($EC_p$) over the independent test period (150 days)*

ysis of the model coefficients and predictive values, including comparison with prior experience; and use of the data-splitting technique. Since no fresh data are available, the splitting technique is adopted here. The splitting step was completed by randomly choosing a sub-data set from the original data set used to develop the model. One hundred fifty observations were randomly chosen from the original data set of 369 observations. Figure 7 compares measured and predicted air conditioner energy consumption using the energy consumption model for each day of the validation interval.

On another aspect, four statistical parameters, which were introduced by Perrone and Miller (1985), are modified here to meet our requirements; these four statistical parameters can be defined in terms of the metered power consumption ($EC$) and the predicted power consumption for the same day ($EC_p$) as:

$H_d$ = percent number of hits, where a hit occurs when $|EC - EC_p| \leq 5.0$ kWh/day;

$AE$ = absolute error, defined as the mean of $|EC - EC_p|$;

$GE$ = algebraic error, defined as the mean of $(EC - EC_p)$;

and

$LE$ = percent large errors, i.e., percent occurrences of $|EC - EC_p| \geq 10.0$ kWh/day.

These parameters were determined for the statistical model defined in Table A3 as well as for the 150 split observations. The results are shown in Table 2, which also includes generalized equivalent exponential Marcov (GEM) and model output statistics (MOS)— quantitative statistical values for comparison purposes.

## SENSITIVITY ANALYSIS

The energy consumption of a house air conditioner as a function of the outdoor and indoor temperature difference, global (i.e., total horizontal) solar radiation, average wind speed, and average relative humidity is plotted in Figures 8 through 10 for the design conditions indicated in Table 3. These plots are generated by using the air conditioner energy consumption model described in Table A3. Although some of these variables do not exist linearly in the regression model, it is important to note that these variables are the basic regres-

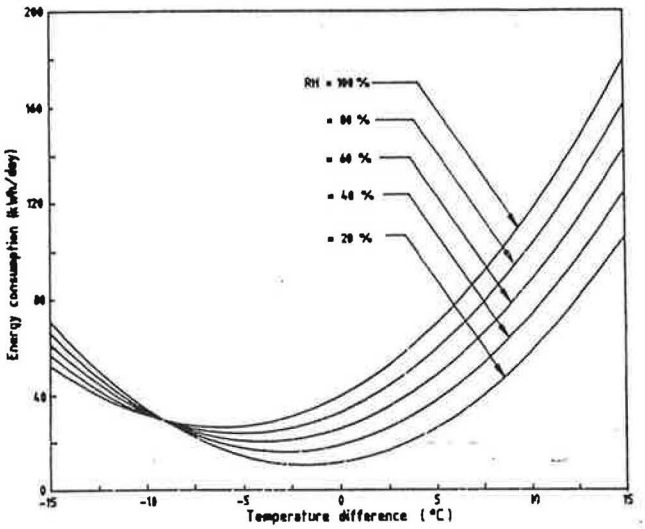


*Figure 8* *Plot of air conditioner energy consumption (EC) vs. the temperature difference between indoor and outdoor conditions; effect of relative humidity*

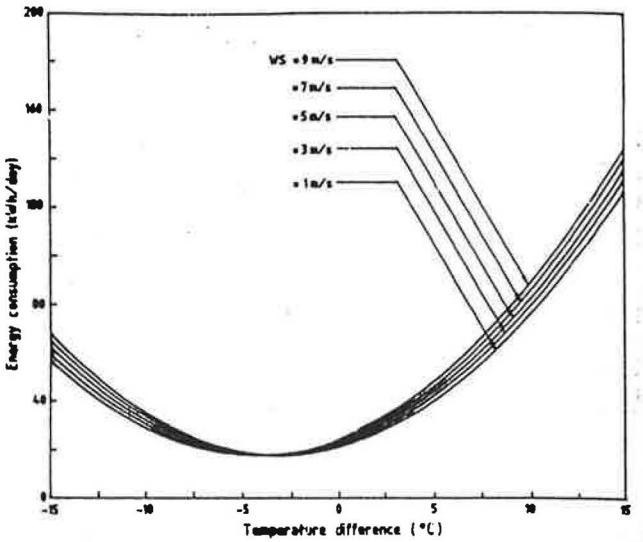


*Figure 9* *Plot of air conditioner energy consumption (EC) vs. the temperature difference between indoor and outdoor conditions; effect of wind speed*

**TABLE 3**
**Design Data for the Sensitivity Analysis**

| Variable | Value |
|---|---|
| Wind speed | 4.0 m/s |
| Relative humidity | 50.0% |
| Global radiation | 4000 W·h/m²·day |

sors that are used to develop the model, as shown in Equation 3.

As expected, these curves indicate that the temperature difference between the outdoor and indoor conditions, which causes the heat to flow from the hot body to the cold body, strongly influences the energy consumption of the air conditioner. The negative temperature difference implies the heating period, when the electric-strip heater is maintaining the indoor condition. At a temperature difference of about −3°C, the energy consumption of a house air conditioner approaches the zero value. The non-zero value of energy consumption in the vicinity of the 0°C temperature difference is attributed to the base load of the house, which is related to intrinsic gains such as heat generated by appliances and occupants.

Figure 8 indicates that the cooling energy consumption is a strong function of the average relative humidity. It may be noticed from this figure that at a temperature difference of 10°C, an increase in relative humidity from 20% to 100% causes about a 100% increase in the energy consumption of the air conditioner. This may be explained by the fact that at high ambient temperature and high relative humidity the amount of moisture in the air is very high, which results in a large increase in the latent load of the structure. On the other hand, as expected, the effect of relative humidity during the heating period is minimal.

The effect of wind speed and total horizontal radiation is demonstrated in Figures 9 and 10, respectively. The wind speed, which is expected to cause an increase in the infiltration load of the house, apparently does not change significantly for the range investigated. Similarly, the total horizontal radiation alone does not contribute much to the energy consumption of the residence. It should, however, be noted that it is not only these variables contributing linearly to the model, as shown in Table A3, but that these variables, with the temperature difference, contribute significantly to the model coefficients, as discussed in Equation 3.

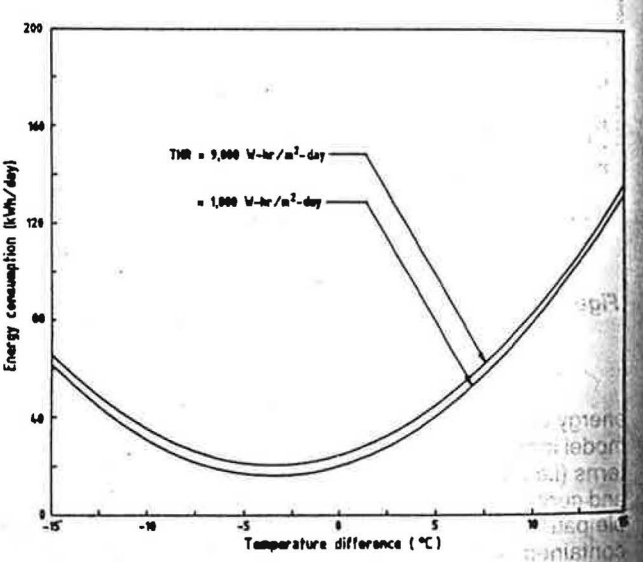


*Figure 10* *Plot of air conditioner energy consumption (EC) vs. the temperature difference between indoor and outdoor conditions; effect of global solar radiation on a horizontal surface*

## DISCUSSION AND CONCLUDING REMARKS

In this paper, the techniques being used to develop a multiple linear regression model have been laid out and the case of daily air conditioner energy consumption of a residence at Dhahran, Saudi Arabia, has been discussed in detail. The air conditioner energy consumption model adequacy was extensively verified within the data domain from which the model was developed. The final model for air conditioner energy consumption is seen to satisfy all criteria of statistical adequacy. The measured ($EC$) and predicted ($EC_p$) air conditioner energy consumption for each day of the complete data interval are compared in Figure 6.

Model validation is also verified within the validation interval. Figure 7 compares $EC$ and $EC_p$ for each day of this validation interval. The model gave very acceptable results, with a mean absolute error ($AE$) of 3.7 kWh/day. The large error ($LE$) is 0.3%, while rates between 1.2% and 2.4% were

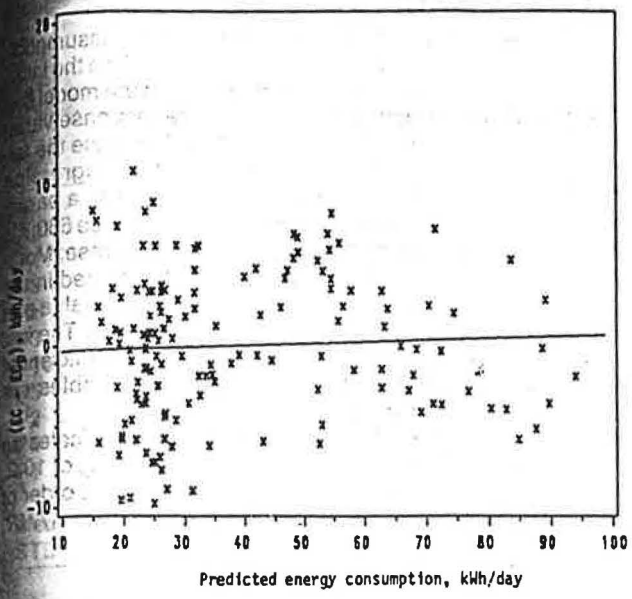


*Figure 11* *Plot of residuals ($EC - EC_p$) vs. predicted values ($EC_p$) for the validation interval (150 days)*

acceptable for forecast models GEM/MOS (Perrone and Miller 1985). Furthermore, a comparison of Figures 6 and 7 shows the degradation of model performance between application in both domains (split and complete) to be small. Quantitatively, the comparison is: 71% number of hits within the domain of the complete data set used in generating the model, 72% within the split data set, $AE$ 3.7 kWh/day for both domains, large error 0.3% for the complete data set, and 0.7% for the split data set (Table 2).

Figure 11 represents the variation of residuals with predicted power consumption within the validation interval. This figure shows the residuals falling within the idealized horizontal band, thus indicating no model deficiency (compare with Figure 4 of the complete data set). However, a small asymmetry toward negative residuals can be seen in Figure 11 and this is confirmed by the −3.8 kWh/day mean algebraic error $GE$ listed in Table 2. This validation discussion has demonstrated, through application to a nearly independent data set, that the developed model is an effective model for prediction and estimation of residential air conditioner energy consumption.

It is important to note that the present energy estimation procedure is different from the degree-day method in three major aspects: (i) the base temperature is variable and is equal to the daily average of the indoor temperature instead of the constant value (18.3°C) used in the degree-day method; (ii) the outdoor temperature is the mean of the hourly averages, while in the degree-day method it is the average of the daily maximum and minimum temperatures; and (iii) most of the meteorological parameters are included in the present analysis instead of considering the ambient and indoor temperatures only, as is the case in the degree-day method.

This study shows, as expected, that air conditioner electrical power consumption is a strong function of the temperature difference between the ambient and indoor, but the study revealed that energy consumption is not zero at a 0°C temperature difference due to intrinsic gains such as heat generated by appliances and occupants. This investigation indicates that at a temperature difference of about −3°C the energy consumption of a house air conditioner approaches the zero value. It has also been found that air conditioner cooling power consumption is a strong function of the daily average relative humidity. At a temperature difference of 10°C, an increase in relative humidity from 20% to 100% causes about a 100% increase in the energy consumption of an air conditioner. This is due to a large increase in the latent load of the structure at high relative humidity and ambient temperature.

The total horizontal radiation, as a linear term, has a small effect on the power consumption of the air conditioner. This may be due to the partial shading provided by a large umbrella-type tree located in front of the house. The wind speed effect is also small due to the green fence, the attached house, and the tree in front of the house.

It should be emphasized that the air conditioner energy consumption model discussed in Equation 3 is general, and includes all the important meteorological parameters as well as global solar radiation contributing to the air conditioner load of a given house or building. Furthermore, to compare air conditioner energy consumption of different houses, it is recommended that the variables used in Equation 3 be fixed, permitting the model parameters to be compared for analysis and interpretation of various components of a building load.

## ACKNOWLEDGMENT

This work was supported by the Research Institute at King Fahd University of Petroleum and Minerals under Project No. 12031.

## NOMENCLATURE

$AE$ = absolute error defined as the mean of $|EC - EC_p|$
$C_P$ = $(SSE_p/S^2) - n + 2p$
$EC$ = measured electric power consumption (kWh/day)
$EC_P$ = predicted electric power consumption (kWh/day)
$GE$ = algebraic error defined as the mean of $(EC - EC_p)$
$H_d$ = percent number of hits, where a hit occurs when $|EC - EC_p| \leqslant 5.0$ kW·day
$LE$ = percent large errors, i.e., percent occurrences of $|EC - EC_p| \geqslant 10.0$ kW·day
$n$ = number of observations
$p$ = number of model parameters (i.e., $p - 1$ regressors plus intercept)
$P1$ = $TA - TIN$
$P2$ = $THR^2$
$P3$ = $THR \cdot RH$
$P4$ = $RH^2$
$P5$ = $RH \cdot WS$
$P6$ = $WS \cdot THR$
$P7$ = $WS^2$
$P8$ = $P1 \cdot THR$
$P9$ = $P1 \cdot RH$
$P10$ = $P1 \cdot WS$
$P11$ = $P1^2$
$RH$ = mean relative humidity (%)
$SSE_p$ = sum-of-squares error for a model with $p$ parameters
$S^2$ = full model mean square error
$TA$ = mean air temperature (°C)
$TIN$ = mean inside temperature (°C)
$THR$ = mean global radiation (W·h/m²·day)
$VIF_i$ = variance inflation factor, $1/(1 - R_i^2)$
$WS$ = mean wind speed (m/s)

## REFERENCES

Belsley, D.A.; E. Kuh; R.E. Welsch. 1980. *Regression diagnostics: identifying influential data and sources of collinearity*, p. 292. New York: John Wiley & Sons.

Debs, A.S. 1983. "Energy conservation in Kuwait buildings." *Proceedings Energy Conservation Measures*, ed. J.D. Parker. KFAS-Kuwait.

Fels, F.M. 1986. "PRISM: an introduction." *Energy and Buildings*, Vol. 9, Part 1, pp. 5-18.

KFUPM/RI. 1984. "Energy conservation on the campus of King Fahd University of Petroleum and Minerals." Final Report, P.N. 12031. Dhahran, Saudi Arabia: King Fahd University of Petroleum and Minerals, Research Institute.

LBL. 1981. *DOE-2.1 reference manual.* Lawrence Berkley Laboratory Report LBL-8706, Rev. 3, Los Alamos Scientific Laboratory Report LA-7689-M.

Montgomery, D.C., and E.A. Peck. 1982. *Introduction to linear regression analysis*, p. 504. New York: John Wiley & Sons.

Perrone, T.J., and R.G. Miller. 1985. *A comparative verification of GEM and MOS.* World Meteorological Organization, Short-Range and Medium-Range Weather Prediction, Research Publication Series No. 9, p. 67.

SAS. 1985. *SAS user's guide: statistics, version 5*, p. 956. Cary, NC: SAS Institute Inc.

Williams, R.O. 1979. *ARAMCO meteorologic and oceanographic data book for the eastern province region of Saudi Arabia. Dhahran, Saudi Arabia*: Environmental Unit, Arabian American Oil Co. (revised June 1979), p. 133.

## APPENDIX A

### MODEL TYPE

The first step was to determine the relative contributions of linear, quadratic, and crossproduct expressions derived from the variables list described in Equation 1. This step was completed by fitting the parameters, using the SAS RSREG procedure, of an optimized full quadratic response surface and then determining critical values to optimize the response with respect to the factors in the model. This analysis gave the values (listed in Table A1) for the coefficient of multiple determination ($R^2$), F-ratio, and model sum of squares ($SS_{model}$), respectively.

**TABLE A1**
**Contribution of Linear, Quadratic, and Crossproduct Terms to the Regression Model**

| Model | $R^2$ | F-ratio | $SS_{model}$ |
|---|---|---|---|
| Linear Regression | 0.594 | 1203.4 | 95,937 |
| Quadratic Regression | 0.325 | 659.1 | 52,550 |
| Crossproduct Regression | 0.037 | 50.6 | 5,994 |

The results clearly show considerable model enhancement from the quadratic and crossproduct terms and hence a multiple linear regression model is deemed the most appropriate for the present application. Hence, the regressed energy consumption equation is redefined as

$$EC_P = F(P1, RH, WS, THR, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11) \quad \text{(A1)}$$

*where*

$P1 = TA - TIN$
$P2 = (THR)^2$
$P3 = THR \cdot RH$
$P4 = (RH)^2$
$P5 = RH \cdot WS$
$P6 = WS \cdot THR$
$P7 = (WS)^2$
$P8 = P1 \cdot THR$
$P9 = P1 \cdot RH$
$P10 = P1 \cdot WS$
$P11 = (P1)^2.$

### STEPWISE REGRESSION

Four variable selection procedures are appropriate to the present application. All should be applied with care (Montgomery and Peck 1982).

### Forward Selection

The forward selection process begins with the assumption that there are no regressors in the model other than the intercept. The first regressor selected for entry into the model has the largest simple correlation ($R^2$) with the response variable, $EC_P$. This is also the regressor that will produce the largest value of the F-statistic used in testing regression significance. The first regressor entered, $P8$, has a partial correlation of 0.650 and an F-to-enter of 680.7. Since 680.7 is greater than a preselected F-value (2.17 for this case; Montgomery and Peck 1982), the variable $P8$ is included in the model. This process continues until the partial F-value at a particular step falls below the preselected F-value. The last regressor entered, $P5$, has a partial correlation coefficient of 0.0004 and an F-to-enter of 3.0. Note that variables are entered into the model up to 0.1 significance level.

Application of this forward selection process indicates an 11-variable model with an $R^2$ of 0.956, Mallow's $C_p$ of 10.2, and F-ratio of 708.1; the effective regressors, in the order of insertion in the model, are $P8$, $P11$, $P9$, $RH$, $P3$, $P10$, $THR$, $WS$, $P4$, $P1$, and $P5$.

### Backward Elimination

The backward elimination process involves a search to find the best parameter combination by working in the opposite direction, starting with a model that includes all candidate regressor variables. This process is carried out to examine the effect of including all the candidate regressors as well as the order of the regressors in the model. The partial F-statistic is computed for each regressor as if it were the last regressor to enter the model. The process of removal continues until the partial F-value of all parameters exceeds the F-to-remove value. The last regressor removed is $WS$ with a partial F-value of 0.8. This process suggests a 10-variable model with $R^2$ of 0.956, Mallow's $C_P$ of 8.6, and F-ratio of 780.2; the effective regressors are $RH$, $P1$, $P2$, $P3$, $P4$, $P5$, $P8$, $P9$, $P10$, and $P11$.

### Stepwise Regression

The stepwise regression process is a modification of forward selection, with the difference that at each step all regressors previously entered into the model are reassessed based on their current partial F-statistics. Hence, a regressor added at an earlier step may now be removed. In this case, the model obtained by applying stepwise regression is a six-variable model with $R^2$ of 0.954, Mallow's $C_P$ of 17.5, and F-ratio of 1252.4. The effective regressors are $THR$, $WS$, $P3$, $P9$, $P10$, and $P11$.

### Mallow's $C_P$ Statistics

Mallow's $C_P$ is a criterion related to the mean square error of the model. It is defined as (Montgomery and Peck 1982):

$$C_P = (SSE_p/S^2) - n + 2p \quad \text{(A2)}$$

where $SSE_p$ is the sum-of-squares error for a model with $p$ parameters, $S^2$ is the full model mean square error, $n$ is the number of observations, and $p$ is the number of model parameters (i.e., $p - 1$ regressors plus intercept).

This step was completed, using the RSQUARE procedure, by finding subsets of independent variables that best predict the dependent variable and by employing correlation coefficient statistics as the selection criteria. Mallow's $C_P$ is calculated for each subset and the subset with minimum $C_P$ is recommended for incorporation into the model. In this case, this criterion indicated a 10-variable model. The variables $P8$, $P11$, $P9$, $RH$, $P3$, $P10$, $P2$, $P1$, $P5$, and $P4$ ($C_P = 8.6$) are the effective regressors, equal to the backward elimination result described earlier.

The above four processes suggest different models. Only the backward elimination and Mallow's $C_P$ procedures indicate the same model, while the forward selection and step-

**TABLE A2**
**Correlation Matrix**

| | P8 | P11 | P9 | RH | P3 | P10 | P5 | P2 | P4 | P1 |
|---|---|---|---|---|---|---|---|---|---|---|
| P8 | 1.00 | 0.09 | 0.90 | −0.46 | 0.16 | 0.90 | −0.48 | 0.68 | −0.43 | 0.97 |
| P11 | | 1.00 | −0.17 | −0.29 | −0.30 | −0.17 | 0.00 | 0.08 | −0.26 | −0.08 |
| P9 | | | 1.00 | −0.37 | 0.37 | 0.88 | −0.51 | 0.63 | −0.37 | 0.96 |
| RH | | | | 1.00 | 0.35 | −0.42 | 0.44 | −0.68 | 0.99 | −0.42 |
| P3 | | | | | 1.00 | 0.23 | −0.06 | 0.35 | 0.30 | 0.29 |
| P10 | | | | | | 1.00 | −0.54 | 0.65 | −0.38 | 0.94 |
| P5 | | | | | | | 1.00 | −0.45 | 0.43 | −0.50 |
| P2 | | | | | | | | 1.00 | −0.66 | 0.69 |
| P4 | | | | | | | | | 1.00 | −0.40 |
| P1 | | Symmetric | | | | | | | | 1.00 |
| Mean | 4.1E3 | 45.9 | −76.2 | 62.2 | 3.0E5 | −5.9 | 306.3 | 2.7E7 | 4.1E3 | −0.5 |
| STD | 3.6E4 | 42.1 | 404.2 | 15.7 | 7.5E4 | 38.3 | 137.4 | 1.4E7 | 1.9E3 | 6.8 |

wise regression procedures suggest two other models. Hence, all the regressors are held for further consideration.

## COLLINEARITY DIAGNOSTICS

If any selected regressor can be closely approximated by a linear relation with one or more of the other regressors in the model, then the affected estimates (i.e., the model coefficients of the collinear terms) are unstable and have high standard errors. This collinearity (or multicollinearity) problem is not statistical in nature (i.e., it is not related to the model) but it is a problem inherent in the data (Belsley et al. 1980). It is essential to investigate this problem with regard to the previously selected variables listed above.

One method for the identification of simple collinearity is the inspection of the off-diagonal elements of the correlation matrix shown in Table A2; collinearity exists if the absolute value of an element is near unity. Table A2 reveals a high correlation between relative humidity (*RH*) and the square of *RH* (*P*4) (0.99), between *P*1 and *P*8 (0.97), and between *P*1 and *P*9 (0.96). However, examining pairwise correlations is not a complete diagnostic measure since one is unable to distinguish among several coexisting near dependencies.

Another diagnostic criterion is based on variance inflation factor (*VIF*) analysis, where $VIF_i = 1/(1 - R^2_i)$ and $R^2_i$ is the multiple correlation coefficient of the ith explanatory variable regressed on the remaining explanatory variables. The *VIF* of each term in the model measures the combined effect of the dependencies among the regressors on the variance of that term. A high *VIF* value must point to collinearity. *VIF*s below 10 are statistically acceptable (Montgomery and Peck 1982). The *VIF* values for the 10 effective regressors suggested by the backward elimination procedure and Mallow's $C_p$ criteria are: *P*8—86.1, *P*11—3.5, *P*9—35.2, *RH*—155.1, *P*3—16.0, *P*10—12.3, *THR*—20.7, *P*5—1.7, *P*4—93.5, and *P*1—158. The *VIF*s of *RH*, *P*4, and *P*1 are exceptionally high. The correlation matrix and *VIF* results reveal that there is simple collinearity between *RH* and *P*4, between *P*1 and *P*8, and between *P*1 and *P*9. This collinearity is sufficient to affect the accuracy with which the regression coefficients can be calculated.

However, there are several techniques that can be applied to overcome collinearity problems (Montgomery and Peck 1982). Elimination of one of the variables (i.e., respecification of the model) is the most appropriate solution here. The model was respecified by eliminating the regressors *RH* and *P*1.

Repeating the above steps (i.e., the forward selection, backward elimination, and stepwise regression procedures) and investigating the Mallow's $C_p$ criteria for the reduced parameter list gave the following results: (i) forward selection indicated that the effective regressors are *WS*, *P*2, *P*3, *P*4, *P*8, *P*9, *P*10, and *P*11, model $R^2$ is 0.954, model $C_p$ is 8.2, and model F-ratio is 932.6; (ii) backward elimination indicated that the effective regressors are *THR*, *WS*, *P*3, *P*9, *P*10, and *P*11, model $R_2$ is 0.954, model $C_p$ is 3.6, and model F-ratio is 1252.4; (iii) the stepwise regression procedure indicated that the effective regressors are *THR*, *WS*, *P*3, *P*9, *P*10, and *P*11, model $R_2$ is 0.954, model $C_p$ is 3.6, and model F-ratio is 1252.4; and (iv) Mallow's $C_p$ criteria indicated that the effective regressors are *P*9, *P*11, *P*3, *THR*, *P*10, and *WS*, model $R_2$ is 0.954, model $C_p$ is 3.6, and model F-ratio is 1252.4.

**TABLE A3**
**Mallow's $C_p$ Criteria Analysis**

| Number of Parameters | $R^2$ | $C_p$ | C.V. | Variables in Model |
|---|---|---|---|---|
| 1 | 0.650 | 2368 | 31,6 | P8 |
| 2 | 0.892 | 477 | 17.5 | P9 P11 |
| 3 | 0.936 | 135 | 13.5 | P9 P11 P4 |
| 4 | 0.946 | 66 | 12.9 | P9 P11 P4 P3 |
| 5 | 0.953 | 12 | 11.7 | P9 P11 P3 THR P10 |
| 6 | 0.954 | 3.6 | 11.5 | THR WS P3 P9 P10 P11 |

Backward elimination and Mallow's $C_p$ criteria resulted in the elimination of another two variables. The best model thus obtained (as indicated by three procedures: backward elimination, stepwise regression, and Mallow's $C_p$) includes the effective regressors *THR*, *WS*, *P*3, *P*9, *P*10, and *P*11. The corresponding *VIF*s are, respectively, 2.5, 1.2, 1.7, 5.3, 5.5, and 1.3; these values are all statistically acceptable.

Mallow's $C_p$ criteria also indicated the best 1, 2, 3, 4, 5, and 6 parameter models. The results are summarized in Table A3.

## GENERAL LINEAR MODELING

The effective regressors recommended by the stepwise regression, backward elimination, and Mallow's $C_p$ criteria are; *THR*, *WS*, *P*3, *P*9, *P*10, and *P*11, were introduced to the general linear modeling procedure to determine the unknown coefficients of the model as well as other statistical parameters; the results are summarized in Table A4. In this particular case, the significance level terms, $PR>|T|$, for all the regressors included are small and much less than 0.05 (the acceptable limit value), indicating that all submitted regressors contribute significantly to the model. Multiple linear regression models were also developed for the first six models listed in Table A3 and the results are shown in Table A5.

**TABLE A4**
**General Linear Models Procedure Results for the Selected Model**

| Source | DF | SS | MS | F-value | $R^2$ | C.V. | RMSE |
|---|---|---|---|---|---|---|---|
| Model | 006 | 154112 | 25685 | 1252.4 | 0.954 | 11.5 | 4.5 |
| Error | 362 | 7427 | 20.5 | | | | |
| Corrected total | 368 | 1611536 | | Adj $R^2$ = 0.953 | | | |

| Parameter | Estimate | T for H0: Parameter = 0 | PR>\|T\| | STD Error of Estimate |
|---|---|---|---|---|
| Intercept | 18.243 | 11.8 | 0.0001 | 1.54 |
| THR | −3.803E−3 | −14.6 | 0.0001 | 2.6E−4 |
| WS | 0.4268 | 3.23 | 0.0013 | 0.13 |
| P3 | 8,715E−5 | 21.0 | 0.0001 | 4.2E−6 |
| P9 | 3.856E−2 | 28.8 | 0.0001 | 1.3E−3 |
| P10 | 0.1208 | 8.4 | 0.0001 | 1.4E−2 |
| P11 | 0.3452 | 54.6 | 0.0001 | 6.3E− |

**TABLE A5**
**Derived Coefficients of the 1 to 6 Variable Models**

| Number in Model | Intercept | P8 | P9 | P11 | P4 | P3 | THR | P10 | WS |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 37.41 | 4.74E−4 | | | | | | | |
| 2 | 29.57 | | 0.0442 | 0.2867 | | | | | |
| 3 | 16.99 | | 0.04955 | 0.3266 | 0.00271 | | | | |
| 4 | 9.19 | | 0.04624 | 0.3321 | 0.00207 | 3.37E−5 | | | |
| 5 | 20.32 | | 0.03828 | 0.3449 | | 8.34E−5 | −3.5E−3 | 0.1136 | |
| 6 | 18.24 | | 0.03856 | 0.3452 | | 8.71E−5 | −3.80E−3 | 0.1208 | 0.4268 |