

6/8  
7/10/80  
SERI/RR-721-377

April 1980

DR. 1473

# Statistical Problems in Design Technique Validation

Joel S. Cohen

**MASTER**



# SERI

**Solar Energy Research Institute**

A Division of Midwest Research Institute

1617 Cole Boulevard  
Golden, Colorado 80401

Operated for the  
**U.S. Department of Energy**  
under Contract No. EG-77-C-01-4042

## DISCLAIMER

**This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.**

## **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

Printed in the United States of America  
Available from:  
National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Road  
Springfield, VA 22161  
Price:

Microfiche ~~\$3.00~~<sup>50</sup>  
Printed Copy ~~\$4.50~~<sup>6.40</sup>

#### NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

SERI/RR-721-377  
UC CATEGORIES: UC-58d,59c

STATISTICAL PROBLEMS  
IN DESIGN TECHNIQUE  
VALIDATION

JOEL S. COHEN

APRIL 1980

**DISCLAIMER**

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PREPARED UNDER TASK NO. 3525.40

**Solar Energy Research Institute**

1536 Cole Boulevard  
Golden, Colorado 80401

A Division of Midwest Research Institute

Prepared for the  
U.S. Department of Energy  
Contract No. EG-77-C-01-4042

THIS PAGE  
WAS INTENTIONALLY  
LEFT BLANK

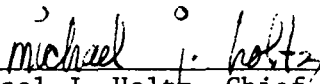
## FOREWORD

Designers, researchers, builders, and consumers who are examining the potential of solar energy to reduce auxiliary energy use of heating, cooling, and hot water constantly face the problem of quantifying this savings. Because of the limited data base of performance results from operating systems, we are dependent on design techniques to predict energy savings. Consequently, the accuracy of these design methods is called into question. For as long as there has been work to develop methods for predicting the performance of solar heating and cooling systems, the discussion of "validation" has ensued; and, one should add, without much agreement.

This work seeks to add to the literature information on the statistical aspects of validation. It is concerned with the statistical variability of the design technique and performance measurement processes, and the influence this variability can have on experimental design, validation data selection, accuracy of validation results, and appropriateness of the validation task.

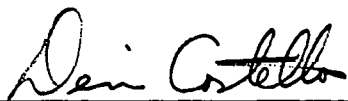
Joel Cohen, Assistant Professor of Mathematics at the University of Denver, on sabbatical at the Solar Energy Research Institute, is responsible for undertaking the research and preparation of this report. He is indebted to many individuals at the Solar Energy Research Institute who helped formulate the ideas in this report. In particular, he would like to thank F. Kreith, N. Kelley, L. Morrison, L. M. Murphy, R. McConnell, M. Connolly, J. Henderson, A. Eden, M. Edesess, C. Bishop, and G. Fegan.

This research has been supported by the Systems Development Division, Office of Solar Applications, DOE.

  
\_\_\_\_\_  
Michael J. Holtz, Chief  
Building Systems Development Branch

Approved for:

SOLAR ENERGY RESEARCH INSTITUTE

  
\_\_\_\_\_  
Dennis Costello, Manager  
Buildings Division

THIS PAGE  
WAS INTENTIONALLY  
LEFT BLANK



## TABLE OF CONTENTS

	<u>Page</u>
1.0 Introduction.....	1
2.0 What is Validation?.....	3
3.0 The Statistical Nature of the Validation Problem.....	7
4.0 Experimental Design for Statistical Validation.....	9
4.1 Introduction.....	9
4.2 Sources of Statistical Error in Validation Studies.....	9
4.3 External and Internal Validation Studies.....	13
5.0 Practical Design of External Validation Methods.....	17
5.1 The Accuracy of the Results.....	17
5.2 Population Definition and Sampling Strategies.....	19
5.3 Instrumentation for Statistical Validation.....	21
5.4 Sociological Survey Accompanying the Validation Study.....	21
5.5 Pilot Study of Validation.....	22
6.0 A Simplified Validation Method for Solar Domestic Hot Water Heaters.....	23
7.0 Conclusion.....	27
8.0 References.....	29

## LIST OF FIGURES

	<u>Page</u>
2-1 Distribution of $F_a - f_d$ .....	6

## LIST OF TABLES

4-1 95% Precision Confidence Interval Estimates for $B_d$ Using the t Distribution with $\sigma = 8$ .....	10
---	----

## NOMENCLATURE

$B_d$	design error bias
$E_d$	design error spread
$Q_{aux}$	annual auxiliary energy requirement
$Q_s$	annual energy provided by the solar system
$f$	annual solar fraction of the load
$UA$	overall heat transfer coefficient
$\mu$	mean or average
$\sigma$	standard deviation

## Subscripts

$a$	actual value occurring in the structure
$d$	design value
$m$	measured value

## SECTION 1.0

## INTRODUCTION

During the past ten years, many procedures have been developed for predicting the thermal performance of solar energy heating and cooling systems. These include simple hand computational procedures requiring a minimal amount of data and computation, computer-based design procedures such as F-CHART and SOLCOST to predict long-term annual performance, and elaborate computer codes such as TRNSYS, which simulate the performance of systems on a time-dependent basis. The following question is of primary interest to users and designers of these procedures:

How accurate are design techniques for predicting thermal performance of solar energy space heating, cooling, and hot water systems?

A design procedure is used in a situation that has a large degree of statistical uncertainty. Included in this uncertainty are basic model inaccuracy; inaccuracy or absence of necessary meteorological data; and imprecision in estimating thermal load and other important system parameters. The uncertainties involved may be so great that the question arises as to whether the system performance can be reliably predicted at all (D. S. Ward 1978).

Although much has been written about the accuracy of design techniques in isolated situations (National Solar Data Program; SERI 1978; Knasel et al. 1979b; Hill 1979; Fanney 1978; Winn 1978), and a validation methodology has been proposed (Kennish 1978; Kennish and Stolarz 1979; Knasel et al. 1979a; Anand et al. 1979), a detailed discussion of the statistical nature of validation has not been presented. This report is concerned with this statistical aspect; it includes a discussion of the statistical variability inherent in the design technique and performance measurement processes, and the ways in which this variability can affect the choice of experimental design, the selection of validation data, the accuracy of the validation results, and even the types of questions that can be reliably answered.

It seems impossible that any practical validation study could give special attention to all the aspects of experimental design outlined here. This work should be viewed primarily as a set of guidelines for the evaluation of proposed statistical approaches to validation. In reality, validation studies can range from an overall consensus of a few experts on the subject to a full scale statistical analysis of a random sample of the systems involved. A statistical approach has the potential of obtaining a more accurate picture of the situation; however, to have some assurance that the final results of the study can infer anything about the larger population of unmeasured systems, it is important to take into account the potential for statistical mistakes and to avoid obvious pitfalls in the experimental design. If not, there is little chance that the approach will give error estimates that are any more accurate than an educated guess and, in this case, there is little point in carrying out the statistical study in the first place.

In this work, validation\* refers to a statistical evaluation of the relationship between the predictions of thermal performance and the actual performance of a given class of systems over a given period of time--usually one year. Validation of the economic predictions, which involves another set of criteria, will not be considered. The study primarily relates to computer-based design techniques such as F-CHART or SOLCOST, although the discussion can also apply to any design technique--and to some extent to the validation of time-dependent detailed codes such as TRNSYS.

---

\*The term evaluation is more appropriate for the type of studies portrayed here. Validation often implies that part of the study involves a modification of the design code to obtain a greater correlation with reality. Here, we are only interested in techniques for evaluating the accuracy of the procedures. However, since the term validation is commonly used in the literature, we will use this terminology.

## SECTION 2.0

## WHAT IS VALIDATION?

Validation of design techniques means different things to different people. In validation, the predictions of mathematical or computer models are compared with physical reality. But agreement stops there. For example, validation can be a measure of how close a design technique compares to a highly controlled experimental situation in which the thermal load, insolation, and other important parameters are known to a high degree of accuracy. At the other extreme, validation may be done on a design procedure used in a realistic residential situation in which the full statistical variability of the design process is encountered. In fact, a validation can often be classified by the amount of statistical variability that the situation includes. In the first of the above cases, the discrepancy between predicted and measured quantities is a result of the limitations of the mathematical model used in the design technique. In the second case, the discrepancy is due to the inadequacies of the input data and the limitations of the mathematical model. Obviously, the results of these two situations do not have much in common, and it would be inappropriate to infer that one situation says much about the other. A similar distinction must be made in the choice of data. If the accuracy of a design technique in a realistic residential situation is of interest, it may be inappropriate to use performance data from an accurately instrumented experimental structure even if typical design data is used for design predictions. Such structures are often offices or laboratories, with thermal loads different from homes. In addition, if there is constant observation by a team of research workers, there is likely to be a higher level of system performance than in a typical residence. If the validation study is concerned with the residential application of design techniques, then the data must be chosen from a random sample of houses that accurately reflects this target population.

Fundamental to any validation study is an unambiguous statement of the question being investigated; only then is a criterion available for deciding which data are appropriate to the study. The validation question may go beyond comparing predicted and measured results. For example, a design technique may be used in choosing components to optimize the performance of a system. Even if the design technique is inaccurate in an absolute sense, it may still be able to predict the degree to which one collector system outperforms another. Validation questions of this type appear to be more difficult to answer than the question of absolute accuracy.

This study is primarily concerned with the development of experimental designs for the statistical analysis of the absolute accuracy of design techniques used in field systems. The basic question is as follows:

How well do design techniques predict the annual thermal performance of an "operating," "well-designed," and "well-constructed" solar space heating system, solar cooling system, or solar domestic hot water heater when using input data "generally available" to the typical user of the technique?

Criteria that are independent of design predictions must be formulated to determine if a system is "operating," "well designed," and "well constructed." The results of the study only apply to systems satisfying these criteria. Without this distinction, confusion arises because two related questions are being tested simultaneously: (1) are operating, design, and construction errors inhibiting the performance of the system; and (2) can the design procedure accurately predict the performance of the system? In this context, it is important to define what constitutes an operating solar energy system. For many problems (a malfunctioning pump, for instance), the distinction is straightforward. For minor problems, such as an incorrectly adjusted thermostat, the distinction is more difficult. A decision must be made as to which problems characterize a malfunctioning system and which are part of the statistical variability the validation study is designed to measure. The question is not how well design techniques predict the performance of optimally tuned systems, but how do they predict the performance of the typical operating solar energy system.

A definition of "generally available" data must also be formulated. For design data, this might mean ASHRAE or manufacturers design data. The selection of meteorological data is more difficult. Most design techniques use weather data averaged over many years to predict long-term average performance of the system. By comparing predictions using yearly averaged weather data to a single year's actual performance, additional design error is introduced since the weather during that particular year may deviate markedly from the average.\* On the other hand, it may be inappropriate to use weather data measured during the year of the study, because this is data that is not generally available to the user. In addition, the measured weather data taken during the year of the study will undoubtedly be more accurate than the long-term averaged data that is currently available. Therefore, by using this data, the study will not reflect the inherent statistical variability of insolation measurement error. By using the averaged data in the study, long-term design error is overestimated; by using weather data taken during the year of the study, long-term design error is underestimated. It seems that the only reasonable solution is to perform the analysis for both sets of data.\*\*

As the validation question now stands, the statistical variability introduced by the general (possibly inexperienced) user is included. This variability

---

\*In some cases, variations in solar fraction due to yearly weather change is relatively small. For example, using the temperature and insolation data given in Klein et al. (1976), the variation in solar fraction for system C (solar fraction 69%), using eight years of Madison, Wisc., weather data, is  $\pm 4\%$ . This variation is considerably smaller than the variation due to other inaccuracies in weather data, parameter error, and uncertainty in thermal load. However, other variables, such as auxiliary energy use, are considerably more sensitive to yearly variation.

\*\*If meteorological data is measured during the year of the study, it should be measured at the collection site of the long-term averaged data, not at the system site. In this way, the statistical variability of extrapolating temperature and insolation data from nearby or similar locations is included in the experiment.

can occur because of an inappropriate choice of input data for the technique; during the use of the technique; and in interpreting the results of the technique. To eliminate this variability, the design technique calculations should be done by an "expert," making a uniform selection of the data, doing and interpreting the procedure correctly. (The statistical variability introduced by the general user can be examined in a separate experiment.) On the other hand, the validation question should include the variability occurring because of the use of generally available data. For instance, solar insolation and temperature input data are frequently estimated using data from nearby or similar locations. The statistical variability that this introduces should be part of the validation.

Restating the validation question:

How well do techniques predict the annual thermal performance of an "operating," "well-designed," and "well-constructed" solar space heating system, solar cooling system, or solar domestic hot water heating system when using input data that is "generally available" and when the technique is performed by an "expert?"

The design and implementation of a method to obtain statistical information about this question is referred to here as statistical validation.

A statistical validation should produce confidence interval estimates for variables such as solar fraction, displaced energy, auxiliary energy, or appropriately normalized versions of these quantities. By way of an example, suppose that a design technique predicts the solar fraction\* of a system. Let  $f_d$  be the predicted solar fraction and let  $f_a$  be the actual solar fraction. A confidence interval estimate for the error  $f_a - f_d$  is of the form  $B_d \pm E_d$ , where  $B_d$  is the average error in the predictions of this variable, and  $E_d$  is a measure of the variability of this prediction (see Fig. 2-1). The value  $E_d$  is defined in such a way that at the time of the validation study, 95% of the systems in the target population have a prediction error  $f_a - f_d$  that falls within the interval  $B_d \pm E_d$ .\*\*

The interpretation of the results of a validation study would be as follows. Assume that the results for solar fraction are  $B_d = -10$  and  $E_d = 8$ , and that

---

\*Because of difficulties in accurately measuring load, solar fraction is one of the hardest variables to validate statistically. However, since solar fraction is frequently used as a measure of system performance, it will be used as an example in this paper.

\*\*The 95% confidence interval is somewhat arbitrary. If 99% confidence intervals are used, the expected variability  $E_d$  is larger and the codes may seem less accurate. If  $f_a - f_d$  is normally distributed,  $E_d = 1.96\sigma$ , where  $\sigma$  is the standard deviation of the values  $f_a - f_d$ . If the prediction error is not normally distributed,  $E_d$  can be determined numerically.



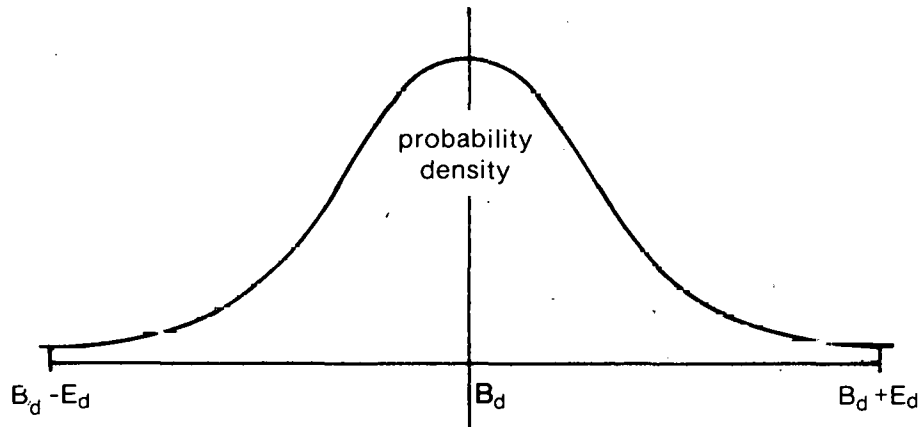


Figure 2-1. Distribution of  $f_a - f_d$

for a given system chosen at random from the population, the design procedure predicts a solar fraction  $f_d = 57\%$ . Then

$$B_d + E_d = -2$$

$$B_d - E_d = -18$$

There is a 0.95 probability that  $-18 \leq f_a - f_d \leq -2$ . Therefore, there is a 0.95 probability of  $f_a$  being between 39% and 55%. In an actual study, it is more appropriate to consider normalized variables such as  $(f_a - f_d)/f_d$ .

## SECTION 3.0

## THE STATISTICAL NATURE OF THE VALIDATION PROBLEM

This section is a summary of the statistical variables that can contribute to the inaccuracy of design techniques.

- Basic Model Error. The equations of the design technique are approximations, and certain aspects of the system have been deliberately eliminated for simplicity.
- Errors in Weather Data. This includes errors in temperature and solar insolation data due to inaccurate measurements, yearly variation, absence of data at a given location, and microclimate variation.
- Errors Due to the Habits of the Occupants. This includes the effect of the thermostat settings, appliance use, amount and pattern of hot water use, installation of window covers and storm windows, and the occupant's contribution to infiltration. The Twin Rivers conservation study has shown that variations in energy consumption for identical structures can vary up to 100% due to occupant habits alone (Socolow 1977/1978).
- Parameter Data Error. This includes incorrect estimation of heat-loss parameters, collector parameters, etc.
- Errors Due to Minor System Design, Construction, and Performance Problems. This includes the mismatch of components, the incorrect placement of sensors, minor control problems, etc. These minor problems are viewed as a legitimate component of the statistical variability that the validation study is designed to measure.
- Errors in the Use of the Design Procedure. This includes the use of inappropriate data, incorrect use of the procedure, and misinterpretation of the results. (Because of numerical truncation, the choice of units can have an effect on the predictions of some computer-based design procedures).
- Miscellaneous Errors. This includes the effects of snow cover and shading; house color; system degradation; and surrounding topography, vegetation, and other structures.

It is impossible to get reliable statistics on the individual effects of most of these variables. D. S. Ward (1978) has shown that it is reasonable to expect errors in the prediction of solar fraction of as much as 10% or 15% from errors in load calculations and insolation data alone.

In the application of a design technique to a solar energy system, the above variables can make either a positive or negative contribution to the overall design error  $f_a - f_d$ . However, from a statistical point of view, the situation is somewhat different. To see this, let  $\mu_i$  and  $\sigma_i$  ( $i = 1, 2, \dots$ ) be the mean and standard deviation of the error contribution to  $f_a - f_d$  produced

by each of the sources of error listed above. The error bias  $B_d$  of the design process is given by

$$B_d = \mu(f_a - f_d) = \sum_i \mu_i \quad (3-1)$$

Assuming the error contributions are independent, the variance of  $f_a - f_d$  is given by

$$\sigma^2(f_a - f_d) = \sum_i \sigma_i^2 \quad (3-2)$$

If  $f_a - f_d$  is normally distributed, then

$$\begin{aligned} E_d &= 1.96 \sigma(f_a - f_d) \\ &= 1.96 \left( \sum_i \sigma_i^2 \right)^{1/2} \end{aligned} \quad (3-3)$$

From Eq. 3-1, it can be seen that the various error contributions can again make either positive or negative contributions to the average bias  $B_d$ . However, if the error contributions are independent, Eq. 3-3 shows that every source of error will make a positive contribution to the variability  $E_d$ , although the contributions with large  $\sigma_i$  will have a more notable effect. There has been a tendency in some validation work to dismiss the error contributions of some of the above variables because of the relative insensitivity of design predictions to their variation. Sometimes this insensitivity can be demonstrated for specific examples, but usually the limits of this insensitivity are not explicit. It is important to note that the collective contribution of a large number of relatively insensitive parameters can cause an increase in the variability  $E_d$ ; by ignoring any of these contributions, a false impression of the accuracy of a design technique can be created.

## SECTION 4.0

## EXPERIMENTAL DESIGN FOR STATISTICAL VALIDATION STUDIES

## 4.1 INTRODUCTION

Because validation is statistical, the techniques of statistical sampling theory are appropriate. Quality control in the experimental design is important to ensure that the results reliably infer information about the larger, target population. A poor experimental design will have little chance of being statistically accurate and may not give results that are any more accurate than an educated guess. The problem is that solar energy systems are difficult and expensive to instrument and monitor; if usual computer-based observational techniques are part of the experimental design, financial and engineering constraints will make it unreasonable to obtain the statistical precision of traditional sampling studies. Fortunately, time-dependent variables, such as incident insolation, collector temperature, or storage tank temperature, which are important for engineering studies, are not really needed for this type of study. In the case of statistical validation, one is primarily interested in measurements of annual solar fraction  $f_a$ , the annual displaced energy  $Q_s$ , and the annual auxiliary energy requirements  $Q_{aux}$ . By restricting the study to system variables that can be accurately monitored and by paying careful attention to the statistical experimental design, it is possible to increase the chances for success within prescribed financial limitations.

## 4.2 SOURCES OF STATISTICAL ERROR IN VALIDATION STUDIES

It has been stated that the object of statistical validation is to obtain confidence-interval-type estimates of the form  $B_d \pm E_d$  for the prediction errors of the variables of interest. The final appraisal of a validation is based on the degree of statistical precision that can be assigned to estimates for  $B_d$  and  $E_d$ . The factors that can contribute to the loss of precision are given below:

The Random Sample Does Not Accurately Represent the Target Population

It is important to emphasize the sample/population question, because it is an issue in much of present validation work. The data must be accurate in the statistical sense as well as the engineering sense. The data must accurately reflect all the statistical variability present in the situation. Though it is difficult to achieve a random sample in practice, this criterion for study data selection can help evaluate the quality of the study and can indicate validation data that are not appropriate and that might subtly bias the results. Referring again to the question of design accuracy in the realistic application environment: if the data set (the sample) consists only of locations where solar insolation data is routinely available, the results of the study would not reflect the statistical variability caused by estimating insolation data from nearby or similar locations (i.e., would be appropriate for only a very narrow target population).

In evaluating a random sample, the time-dependent character of a validation study must also be considered. Because the solar energy industry is rapidly

changing, a validation study could become obsolete shortly after its completion. Changes in system and control components, implementation of load-matching strategies, refinements in insolation data, improvements in installation techniques, changes in the precision of parameter estimation, and sociological changes in the characteristics of the user population can alter the precision of the results. The design techniques will also evolve with time. Finally, if measurements for the validation study are carried out while the weather is significantly different from average, the precision of the results will be altered. (Statistical methods can be used to test the significance of this difference.) The conclusion of a validation study should include a realistic estimate of the limitations of the precision of its results. (Practical approaches to sample selection are discussed in Section 5.2.)

#### The Sample Size is Too Small to Give Accurate Results

Since the random sample is a small fraction of the target population, there is uncertainty because of random sampling error. It is necessary to determine the size of the experiment required to ensure a desired degree of sampling precision in the estimates for  $B_d$  and  $E_d$ . For example, consider the problem of estimating  $B_d$  and  $E_d$  for the variable  $f_a - f_d$ . Suppose that in the target population  $f_a - f_d$  is normally distributed, with  $\mu(f_a - f_d) = -10$  and  $\sigma(f_a - f_d) = 8$ . In this case,  $B_d = -10$  and  $E_d = 1.96\sigma = 15.6$ . In Table 4-1, the 95% precision confidence interval estimates for the bias  $B_d$  is given. The precision confidence interval is a measure of the expected precision of the study with respect to errors due to random sampling mistakes. For example, if the sample size is  $n = 20$ , one would expect that, with probability 0.95, a validation study would estimate that  $B_d$  would fall within the interval from -13.7 ( $-10 - 3.7$ ) to -6.3 ( $-10 + 3.7$ ).<sup>\*</sup> Larger values of the sample size  $n$  will give smaller precision confidence intervals. Precision confidence interval estimates for  $E_d$  are obtained using the chi-square distribution and in this case are roughly the same size.

Table 4-1. 95% PRECISION CONFIDENCE INTERVAL ESTIMATES FOR  $B_d$  USING THE  $t$  DISTRIBUTION WITH  $\sigma = 8$ .

$n =$ size of random sample	Precision Confidence Interval Estimate
4	$\pm 12.7$
10	$\pm 5.7$
20	$\pm 3.7$
50	$\pm 2.2$
100	$\pm 1.6$

<sup>\*</sup>This confidence interval presumes a random sampling procedure; negligible measurement error; and a measuring process that does not affect system performance.

It should be noted that confidence interval estimates depend on the value for  $\sigma$ . Smaller (or larger) values of  $\sigma$  will give greater (or less) precision in these estimates. Of course, the estimate for  $E_d$  (or  $\sigma$ ) is a desired product of the validation study. Rough estimates of  $\sigma$  for use in estimating the appropriate sample size in an actual study may be obtained from a pilot study or from existing data.

### Effect of Measurement Error

It is difficult to obtain accurate data for some performance variables of solar energy systems. Because measurement error can be of the same magnitude as prediction error, it is important to examine its effect. Measurement error affects estimates of both  $B_d$  and  $E_d$ . For example, let  $f_m$  represent the measured solar fraction of a given system and let

$$y = f_a - f_d ,$$

$$x = f_m - f_d ,$$

$$\epsilon = f_m - f_a .$$

In a statistical validation study, one is interested in a statistical analysis of the variable  $y$ . Because of measurement error  $\epsilon$ , the variable  $x$  is used as an approximation for  $y$ . In this case, the error bias and spread for a given variable is approximated by the equations

$$B_{d,\text{observed}} = \mu(x) \quad (4-1)$$

$$E_{d,\text{observed}} = 1.96 \sigma(x) . \quad (4-2)$$

Because  $x = y + \epsilon$ ,

$$\mu(x) = \mu(y) + \mu(\epsilon) . \quad (4-3)$$

Furthermore, because  $y$  and  $\epsilon$  are, in general, independent,

$$\sigma^2(x) = \sigma^2(y) + \sigma^2(\epsilon) . \quad (4-4)$$

The quantities  $\mu(\epsilon)$  and  $\sigma^2(\epsilon)$  represent the overall average bias and variance of the measurement technique. From Equation 4-4, it follows that

$$E_{d,\text{observed}} \geq E_d .$$

If  $\mu(\epsilon)$  and  $\sigma^2(\epsilon)$  are comparable in magnitude to  $\mu(y)$  and  $\sigma^2(y)$ , the measurement error could significantly alter the results of the study. In order for data to be useful for statistical validation, either the data must be extremely accurate or estimates for the statistics of the measurement error must be

available. Without some knowledge of the measurement error statistics, it is impossible to tell whether the observed design technique error is primarily due to measurement error or due to the limitations of the design code and the inaccuracies of the design input data. When measurement error statistics are available, Eqs. 4-3 and 4-4 can be used to subtract the effect of measurement error. In this case

$$B_d = \mu(x) - \mu(\epsilon) \quad (4-5)$$

$$E_d = 1.96 [\sigma^2(x) - \sigma^2(\epsilon)]^{1/2} . \quad (4-6)$$

Because of measurement error, some variables cannot be statistically validated reliably. Variables that can not be accurately measured, or that depend on energy balances (which in turn depend on approximated quantities), can introduce a large observed statistical variability into validation. There is no justification for doing an expensive statistical validation study for variables that cannot be readily defined and measured with available engineering techniques. To be useful and reliable, statistical validation should only be done for variables that can be precisely defined, have an unequivocal relationship to the reality to be measured, and can be accurately measured. From the economic point of view, the consumer is interested in saving electricity and fossil fuel by using a solar energy system. Design techniques give the consumer an idea of what to expect by estimating such things as thermal load, solar fraction of the load, energy displaced by the solar system, and expected backup fuel consumption. Although concepts like solar fraction, thermal load, and displaced energy are intuitively easy to grasp, they are difficult to define precisely and measure accurately even with sophisticated monitoring systems. For the consumer, the only variable that is the least bit concrete is the expected backup fuel usage, and in most cases this is the only measure that is available to determine if the system is operating as advertised. The auxiliary energy consumption is also the easiest and least expensive variable to statistically validate. A study designed to measure prediction error for this variable alone could give some measure of design accuracy. Statistical validation studies for other performance variables depend on the availability of accurate, inexpensive techniques to measure them.

#### Errors Due to the Effects of Measuring the System

Alterations in a system's performance that occur as a result of monitoring must be minimized. In some cases, alterations can occur because of the actual physical measuring process. Further, the people using the solar energy system are an important and highly unpredictable aspect of the system, and monitoring the performance of a system may subtly alter their energy behavior. (There is some evidence that the fact that the system is solar does alter the energy use patterns of the participants.) Dealing with these types of performance modifications is difficult. One possibility is to obtain baseline data (such as kilowatt-hour usage) before or after the experiment; but because of weather variability, this might be of limited usefulness. The alternative (and perhaps better) tactic is to make the measuring devices as unobtrusive as

possible. For ethical reasons, the participants must be made aware of the kinds of measurements that are being made. On the other hand, they do not have to know all the details of the questions that are being investigated until the end of the study. (It might even be useful to consult social scientists, who routinely face questions of this sort).

#### 4.3 EXTERNAL AND INTERNAL VALIDATION

It is possible to divide validation studies into two classes (external and internal) with contrasting approaches to the statistical analysis of design accuracy. The goal of an external-type validation is simply a "bottom line" evaluation of the accuracy of the techniques. This involves sampling a large number of systems and a statistical analysis of the difference between predicted and actual values of annual performance. The entire variability in the situation is characterized by the statistical analysis of the prediction errors of a few annual energy variables. In this case, no attempt is made to investigate the statistical distributions of the various sources of design error discussed in Section 3.0. Alternately, an internal approach involves a statistical analysis of some (or all) of the different sources that can contribute to design inaccuracy. For example, one might attempt a statistical analysis of the difference between predicted and actual UA values or predicted and actual solar insolation at a randomly chosen collection of sites and then examine the effect of this variability on the output of a given design technique. In this case, the "bottom line" absolute accuracy of a design technique is obtained by a suitable statistical combination of all the various contributions to design error.

An internal validation method has recently been proposed (Kennish and Stolarz 1979; Knasel et al. 1979a; Anand et al. 1979). In this method, the statistical analysis is divided into three levels:

Detailed Code vs. Experiment (Level 1). This level involves an analysis of the predictive capability of detailed, time-dependent computer codes (such as TRNSYS) when compared to a reality in which all the parameters of the system and the meteorological data are assumed to be known to a high degree of accuracy. It is proposed that the comparisons be done for components as well as for entire systems.

Detailed Codes vs. Design Codes (Level 2). This level is concerned with a statistical analysis of the difference in output of a detailed level code (such as TRNSYS) and a design code (such as F-CHART) when identical data is used in each code.

Input Errors with Typical Data Bases (Level 3). This level concerns the effects of errors in parameter and weather data on the predictive capability of the design techniques. It is proposed that the errors in weather data, and the differences between the parameter values generally available to the public (such as ASHRAE UA values) and actual measured values, should be analyzed statistically, and the effects of these errors in design technique prediction should be analyzed.



This method claims that an evaluation of the accuracy of a design technique can be found via a suitable statistical combination of the variability observed at the three levels. It also suggests that an external-type validation (called Level IV) should be performed for a few systems as a check of the combined statistics of the first three levels.

From the point of view considered here, analysis at each level is a statistical sampling problem; if statistical precision is of interest, it is necessary to consider the sources of error discussed in Section 4.2. Although it may be possible to obtain some useful information related to design technique development using the proposed method, it is extremely difficult to use it for a statistical validation, for the following reasons:

- The designers of this method are aware that it is difficult to obtain good, long-term data on a large number of solar energy systems. They rely on the premise that Level I validation, which involves this type of data, can be accomplished by combining the results of component validation with the validation results of a few well-instrumented systems. The proposers give few details as to how this would be done, although it has been suggested that some of the data currently being collected at various instrumented sites around the country, and some of the previous validation work of others, might be useful at this stage of the study.
- Although component validation is of interest, it is only very tenuously related to system validation. In addition, the available system data is of questionable quality (for statistical validation) because it is often inaccurate in both the engineering and the statistical sense. This is especially important when it comes to population-descriptive terms like "operating," "well designed," and "well constructed," and the energy-behavior characteristics of the occupants. As we have pointed out, for any evaluation of design accuracy, it is essential that an estimate of measurement accuracy be available. This is going to be difficult to obtain for the data sources currently available. Accurate statistical studies should not be attempted using highly processed data of questionable accuracy from a variety of sources simply because that data happens to be available.
- Because Level I is itself a statistical problem, the techniques of sampling theory are applicable. However, because the expected variability at Level I is smaller than that expected in an external validation study, it may actually be possible to use a smaller sample size. The random sample at this level is designed primarily to represent the expected variability due to basic model error; for this reason, the sampling could be easier to implement than that required for external validation, where a larger degree of statistical variability is involved. However, depending on how this level is viewed, some or all of the variability that results from the occupants' lifestyle, from minor design and operational problems, and from the effects of shading and the surrounding environment might also be included. If this variability is not incorporated, it would have to be added at Level III: For example, if the systems at Level I are all tuned to eliminate most operational problems, then the variability due to minor operational problems will have to be introduced at Level III. If all this additional

variability is included at Level I, the required random sample is similar to that of an external validation and the data could be used for an external approach in the first place. If the variability is incorporated at Level III, the complications are even more extensive. All the statistical variability must be accounted for at one of the three levels.

- There is some question as to whether Levels I and III are amenable to statistical analysis at all. Both levels require accurate determination of the parameters of the systems. The most important system parameter, the UA value of the house, is not only difficult to predict, it is almost impossible to accurately determine in practice without resorting to energy balances based on quantities that in turn can not be measured and must be approximated. It is doubtful whether a meaningful statistical analysis can be based on variables that cannot be accurately measured. It will also be difficult to get the accurate estimate of the errors in weather data that is needed for Level III. Error estimates are useless if the variables cannot be accurately defined or measured.
- There are additional statistical problems associated with combining the various sources of variability to obtain the final estimate of design accuracy, including: the determination of proper variable normalization and the dependence between variables at Level III; the dependence between the variability at the various different levels; and the assessment of the effect of measurement error for all the different statistical analyses.

The Level II validation, comparing the output of a detailed code to that of a design technique, is solely a computer study, which does not appear that difficult. Similar work (but without the statistical orientation) was done in the development of the FCHART design technique (Klein 1976).

This has been an attempt to interpret an internal validation methodology proposed by Kennish and Stolarz (1979), Knasel et al. (1979a) and Anand et al. (1979). Other interpretations are also possible. There are so many potential sources of statistical error in this approach that it is difficult to see how anything but extremely rough estimates of design accuracy could be obtained. However, some aspects of the internal approach could be used to help quantify weak points in design technique development and use. For evaluation of design accuracy, a careful external approach is much more amenable to statistical analysis and has a much greater chance of success.



## SECTION 5.0

## PRACTICAL DESIGN OF EXTERNAL VALIDATION METHODS

When choosing an approach for design technique validation, it is important to be aware of the questions that need to be answered, the necessary precision of the results, and the administrative and financial resources that are available. In this section, some guidelines for the practical design of an external statistical validation are presented.

## 5.1 THE ACCURACY OF THE RESULTS

Validation studies have virtually ignored the question of the inferential accuracy of the studies. Given the high degree of statistical variability, the difficulty of obtaining good data, and the rapidly changing state of the solar energy field, one can question whether any validation study, restricted to a small sample of the population at a given point in time, can reliably predict a measure of design accuracy. Therefore, should validation be restricted to inexpensive studies that give only a rough idea of design accuracy? To answer this, consider the consequences of a faulty estimate of design accuracy. For example, suppose the results of a validation study indicate that a given design method overestimates the solar fraction by between 0% and 20% ( $B_d = -10$  and  $E_d = 10$ ), and suppose in reality the technique is actually accurate to  $\pm 10\%$  ( $B_d = 0$  and  $E_d = 10$ ). Could this difference cause the industry to overdesign systems, which would increase the cost and would have a negative impact on commercial penetration?

How accurate need a validation be? If rough estimates will suffice, then there are inexpensive approaches that should be considered before a statistical method (involving added instrumentation) is undertaken. These approaches include:

- Expert opinion: Elicit the opinions of several neutral, expert observers, thoroughly familiar with the practical use of design techniques, and (in some way) average the results.
- The opinions of design technique users: Take a random sample of the actual users of the techniques, elicit their opinions, and average the results.
- Use of available data. As has been stated, much of the presently available system data is of limited use for statistical validation studies because it is of unknown statistical and engineering accuracy. Nevertheless, by carefully selecting data appropriate to the question at hand, it might be possible to obtain a rough estimate of design accuracy. It is important to note that, from the point of view of statistical validation, additional data similar to that which is currently available will do little to improve the situation.
- Sensitivity and worst-case studies. It is easy to find the effects that an error in one of the many variables of a design technique will have on predictions. Combining the results of a number of these

"sensitivity" studies, one may get some feel for design accuracy. A study of this type, examining the effects of errors in insolation and UA values on solar fraction and economic predictions, was done by D. S. Ward (1978). Of course, studies of this type can give no information about basic model error, and they depend on assumed error distributions of the variables. Since errors in different variables can often have opposing effects, worst-case studies, assuming errors in all the variables, will make a design technique appear less accurate than it actually is.

- Comparison studies. It is possible to get some measure of solar heating system performance and design accuracy by comparing the electricity, natural gas, and heating oil usage of matched pairs of conventional and solar houses. The results of a study of this type for a single housing development have been given by J. C. Ward (1978). Because of the difficulty of controlling for differences in construction and resident energy behavior, this approach can also make a design technique appear less accurate than it actually is.

The advantages of these procedures are their moderate expense and the fact that, in some cases, the experiment can easily be repeated at a later date to estimate the effects of industry and design method evolution. In fact, even if a more accurate statistical method is used, it may be useful, for comparison purposes, to also sample the opinions of the users of the techniques. The disadvantage of all of these procedures is that there is really no possibility of knowing how rough the final answers actually are: the design accuracy could be either over- or underestimated. However, a complex statistical method, using a random sample of systems and involving additional instrumentation, should be done only if the inaccuracy of the above techniques is unacceptable, and more accuracy is needed to ensure a high level of user confidence in the techniques. Note, however, that a large-scale statistical experiment, carried out without thorough and continuous attention to proper experimental design, will not give answers that are any more accurate than the above simpler, less costly methods.

In Section 4.2, the effect of sample size on the reduction of random sampling error was discussed. To make similar computations to decide on the sample size for a given level of precision in an actual statistical validation, there must be an estimate of the standard deviation  $\sigma$  for the variable of interest (such as  $f_a - f_d$ ) before the study begins. A rough estimate for  $\sigma$  must be obtained from pilot studies or from previous data. It should be noted that the precision confidence intervals in Table 2-1 have been computed assuming accurate measurements of the variable in question. If large measurement error is suspected, and the statistics of this measurement error are not known, the precision confidence intervals cannot provide useful information. However, when measurement error statistics are known, it is possible to analyze the effect of measurement error on the actual error predictions (Eqs. 4-5 and 4-6) and on the size of the precision confidence intervals.\* In this case, it is

---

\*All of this applies to nonnormalized versions of the variables. The problem is more involved for normalized variables. In some cases, it may actually be possible to obtain an estimate for the statistics of the measurement error for nonnormalized variables: one such possibility is presented in Section 6.0.

safe to assume that precision confidence intervals for both  $B_d$  and  $E_d$  will be no more than twice the size of the intervals given in Table 2-1.

## 5.2 POPULATION DEFINITION AND SAMPLING STRATEGIES

Unambiguous criteria defining the population of solar energy systems of interest and a sampling procedure for obtaining an accurate cross section of this population are essential for a successful validation. The problem is complicated by the fact that the initial consumers of solar energy systems are not particularly representative, sociologically or geographically, of the general population; to obtain meaningful results, the sampling procedures may have to be modified to take this into account. This stage of the study is crucial, and it may be useful to consult with a statistical polling organization that has the experience and personnel for handling these problems. In this section, we discuss some approaches to problems that arise at this stage of the study.

The problem of population definition can be addressed as follows:

- Primary classification. The study is restricted to a class of systems, such as solar domestic hot water heaters, and this class includes only systems for which the design technique was intended. Because design techniques are often used for situations in which they are not quite applicable, criteria will have to be formulated to establish which systems are appropriate for the study. For example, newer systems often include advanced control techniques and opportunities for economical modes of operation that will affect the accuracy of the design technique.
- Secondary classification. This involves the formulation of criteria to quantify descriptive terms such as "operating," "well designed," and "well constructed." Ideally, these criteria should depend only on short-term observations of the system and should be independent of design predictions.
- Additional classification factors. This might involve further administrative limitations, such as restricting the population to systems with electrical backup or only considering recently constructed systems to better represent the current state of the industry. In addition, there should be criteria to establish that the system is being used in an environment that reflects the purpose of the study. Finally, criteria should be developed to make sure that the codes are uniformly applied to the different systems with the appropriate input data to obtain the intended results of the study.

To choose a random sample from the above population, one could, in principle, list all the solar energy systems that satisfy the criteria and then, using a random sampling strategy (such as random numbers from a computer), select the systems to be monitored. This is impractical and may not even reflect the statistical variability we would like to measure. To begin with, listing all the systems that satisfy the population criteria is a major task. In addition, a random sample would scatter the systems, making accurate monitoring time-consuming and expensive. As we have pointed out, the geographic

distribution of solar energy systems is not uniform throughout the country and is even quite different from that of the human population. To avoid these difficulties and to more accurately reflect the intent of the study, an alternative, multistage sampling procedure should be used. For example, if one were interested in validation on a nationwide basis (a difficult problem), a possibility would be to block-out a large number of regions by placing a rectangular grid over a map of the country. Then a number of geographic regions could be chosen at random. The second stage would be to sample one or more systems from each of the selected regions. Strictly speaking, if a large number of regions are chosen, a proper interpretation of the results of this validation should mention that it assumed a uniform geographic use of the solar energy system throughout the country. Other possibilities for the first stage of a multistage procedure are geographic sampling with a tendency to sample in highly populated regions or in regions of expected rapid expansion of the industry. It should be noted that if (for administrative purposes) the study is restricted to a few randomly chosen regions of small area, the precision of the validation study will be reduced with respect to the effect of variability in weather data.

When it comes to actually choosing systems, it appears that the only practical approach on a nationwide basis is to sample the installers or dealers of solar energy systems and randomly choose one or more of their recently installed systems. Lists of dealers and installers are currently being assembled at SERI. Of course, care must be taken to ensure that the dealers do not simply list their most successful systems. Other approaches to this problem might involve the utilities, solar energy societies, lists of municipal building permits, or simply asking a statistical polling organization what to do. Of course, the final sample should only include those systems satisfying the population criteria. This will require an on-site inspection by an engineering team thoroughly familiar with the population criteria and the intent of the study.

Some statistical studies divide the population into classes (or strata) and choose random samples from each strata. For example, there are something like fifteen types of solar water heaters on the market, and sampling systems from each of these classes might be considered. There are two main advantages to stratified sampling. First, if the sample size in each strata is large enough, a measure of design accuracy in that particular strata can be obtained. In addition, if the statistical variability within each strata is small, the precision of the results of the entire study can be increased. However, this approach does require information on the relative frequency of each strata in the overall population. In the case of design validation, it does not appear that the precision of the study will be significantly increased by taking a stratified approach. A properly chosen random sample of the overall population of interest should contain members of each strata at roughly the same relative frequency as that occurring in the population. If a system is omitted from the overall random sample, then its relative frequency in the population is probably too small to have any affect on the estimation of design error.

Much of what has been said here about population definition and sampling strategy represents an ideal that is difficult to achieve in practice. A

discussion of the advantages and disadvantages of various sampling strategies for general sampling studies is given in Cochran (1977) and Hansen et al. (1953).

### 5.3 INSTRUMENTATION FOR STATISTICAL VALIDATION

The data requirements and performance evaluation procedures for an engineering analysis of solar energy systems are not discussed in detail here because they can be found in Streed et al. (1976) for active systems and in Palmiter et al. (1979) for passive systems. However, there are differences between data requirements for statistical validation and for an engineering analysis. With statistical validation, unlike a technical study, interest does not lie in energy balances, subsystem performance, component and system efficiencies, and component interactions. Interest resides in the integrated values of a few well-defined and easily measured variables that describe system performance. Time-dependent data is not required. What is needed is to monitor a large number of systems. Although there has been some interest in simple, low-cost monitoring (for example, U.S. DOE 1978; Cummings et al. 1978), it will probably be more effective to develop instrumentation that is specifically designed for the study at SERI.

In Section 4.2, the effect of measurement errors on the precision of validation studies was discussed. Without knowledge of the measurement error statistics, it is impossible to tell whether the observed design error is primarily due to measurement error or to the limitations of the design technique and the inaccuracies of its input data. In this context, measurement error is defined to be any phenomena that gives rise to an inaccurate picture of system performance. This might include instrument error, the effects of data sampling rates, misplacements of sensors, the effect of the measuring process on system performance, data acquisition errors, or perhaps the inability to define and measure the variables in the first place. In practice, statistical validation is done with respect to a reality given by the definitions of the measured variables and a uniform placement of sensors in the various systems that make up the random sample. An accurate determination of the statistics of the measurement error with respect to this reality is important and should include a statistical model of the measurement process and estimates of the mean and standard deviation of the measurement error of the integrated values of each variable of interest. If accurate estimates of measurement error statistics are available, a correction for this uncertainty, based on Eqs. 4-5 and 4-6, can be made for nonnormalized variables. For normalized variables, an attempt to subtract out the effects of measurement error can lead to statistical difficulties. From a practical point of view, it may be difficult to correct for measurement error even for nonnormalized variables. What is important is that the measurement errors be unbiased, with a standard deviation that is much smaller than the standard deviation of the design error that the validation study is attempting to find. An approach in which it may be possible to estimate the measurement error is discussed in Section 6.0.

### 5.4 SOCIOLOGICAL SURVEY ACCOMPANYING THE VALIDATION STUDY

It will be useful to conduct two sociological surveys in conjunction with the validation study to obtain information about both system users and system



installers. For system users, the survey should include information on thermostat settings, infiltration patterns, hot water usage patterns, family size and characteristics, impressions of system performance, use of load-matching techniques, and system repair problems. In addition, information should be obtained about the effect of monitoring on energy use. It might also be helpful to obtain statistical technical information on the systems being sampled, including a measure of the adequacy of system design, construction, and level of operation. Installers should be surveyed to obtain information on design techniques (sources of data, performance predictions, etc.), imprecisions of system performance, and impressions of design accuracy. Because many design techniques are available, this survey may not provide enough information for a good measure of the contribution to design error from the user of the technique. This contribution to validation should be examined in a separate statistical sampling study.

#### 5.5 PILOT STUDY OF VALIDATION

A pilot study, on a limited scale, would be useful to reveal problems before a large-scale study is attempted. All aspects of the larger study should be included in the pilot study, including the formulation of the questions to be answered, the definition of the populations, the random sample selection, and system instrumentation. It may also be useful to place the validation instruments in structures that are already being monitored by other instruments to compare the results with unmonitored systems. For practical reasons, the pilot study would probably have to be done on a limited time scale in a small geographic region and would only give rough estimates of the statistics of interest.

## SECTION 6.0

A SIMPLIFIED VALIDATION METHOD FOR SOLAR DOMESTIC  
HOT WATER HEATERS

Among the commercially available active solar energy systems, domestic hot water heaters are the most economically accessible to the consumer and also the easiest (relatively) to validate. It is appropriate to concentrate the initial work of statistical validation on these systems. Although there has been much speculation about the correctness of various design techniques, including the accuracy of the input data, no attempt has been made to measure the overall statistical variability in the design process. A well-designed solar water heater validation study would give a measure of the overall statistical variability in this process due to uncertainty in the actual design techniques, inaccuracies in insolation data, hot water loads, and system parameters, and the effects of minor design, construction, and operational problems. In addition, it would give the solar community firsthand experience with the possibilities and limitations of the statistical validation process for a case in which the system is relatively easy to monitor and for which there is an ample population. Finally, the experiment could provide information on the differences among the various approaches to system design when applied in the realistic design environment.

A wide range of instrumentation strategies is possible for solar water heaters, from a simple monitoring of backup fuel usage to sophisticated computer-based monitoring capable of obtaining energy balances for time-dependent data. In the remainder of this section, a simple, low-cost validation method (suggested by F. Kreith) is discussed. This approach attempts to evaluate the accuracy of design predictions for the auxiliary energy requirements  $Q_{aux}$ , the solar energy contribution  $Q_s$ , and the solar fraction  $f_d$  for domestic hot water heater systems with electrical backup.\* The idea is to monitor electrical backup usage for lengthy time intervals each month for two different operational states of the system. In the first state, the system is operating normally, with the solar components providing a fraction of the system's energy requirements. In the second state, the solar components are disconnected, and the system operates as a standard hot water heater. Kilowatt-hour and time measurements are recorded several times per month, including the times when the system is switched from one state to another. This data can be analyzed to obtain estimates for the annual values of  $Q_{aux}$ ,  $Q_s$ , and  $f_d$ . All that is required is a clock, a kilowatt-hour meter,\*\* and a mechanism for switching the system from one state to the other. One possibility for data collection is to place the meter and state switch outside the residence. Actual data collection and state changes could be accomplished by neighbors, students, polling organizations, EPRI, or the utilities and sent back to a data

---

\*Only electrical backup systems are considered, because of the simplicity of the measurement and to reduce the cost of the study. A majority of solar domestic hot water systems use electrical backup.

\*\*In practice, there could be a single digital readout consisting of time digits, energy digits, and error-checking lights.

collection center on postcards.\* In order to reduce the error in this measurement sampling process, it is essential that the actual user not know which state the system is in.

The measurement error, in this case, is due to the inaccuracies of the kilowatt-hour meter and the fact that solar system performance is being sampled. Because insolation, outside temperature, and thermal load are not uniform in time, and because a relatively long time period (about two weeks) will be used, some error can be expected in the estimates of  $Q_{aux}$ ,  $Q_s$ , and  $f_d$ . However, what is important for statistical validation is not the error in individual cases, but the statistics of error for many systems, and there is a way to obtain an estimate for this quantity. A time-dependent, quasi-steady-state program (such as TRNSYS), although of unknown absolute accuracy, does "perform" qualitatively like a solar water heater. An estimate of the measurement sampling error can be obtained by running this detailed model in two different ways for a randomly selected sample of locations. First, the computer model is run in the usual manner.\*\* In the second case, the model is modified to reflect the two operational states in the actual validation experiment and is rerun with the same data, the system being switched between the two states in the same way as in the actual measured systems. Computed auxiliary energy usage calculations are recorded at the same times that the recordings are taken in the measured systems. The data from this computer run are then compared to the values for  $Q_{aux}$  and  $Q_s$  that have been obtained in the first simulation run. A statistical analysis of the differences in output of the two runs for a large number of systems can be used to estimate the error due to the measurement sampling process. This analysis, together with an estimate of the error for the kilowatt-hour meter, will give an estimate for the measurement error in the study.

There are technical problems to be considered before using this type of approach. The systems may have to be modified to ensure complete state separation and to avoid collector stagnation or other physical effects. Transient effects occurring at the time of state change can be minimized by allowing time for the systems to come into equilibrium and by increasing the number of

---

\*The Department of Energy has recently supported work on a national solar data collection system (Delima et al. 1978).

\*\*We can only run this model where solar and temperature data are available. Although it would be useful for this data and load data to be available for each of the measured systems, it would defeat the purpose of keeping the measurement process simple and it is not essential for error estimation. Any collection of sites where data happened to be available and that accurately reflect the random sample being tested can be used. In fact, since measurement sampling error is being estimated, it is not even necessary that the data be accurate. However, to make the error estimate more realistic, this computer experiment should model the actual experiment as much as possible. For example, if data are taken at irregular intervals in the real systems, the simulation should also use irregular intervals. To make the thermal load realistic, hot water load data from other measured sites would be input into the detailed model.

measurements per month. The computer simulation experiment can also be used to estimate these effects.

There are other variations to this approach. For example, the systems could be operated with the solar components supplying a fraction of the load from the winter solstice to the summer solstice and disconnected the rest of the year. Another possibility is to run the experiment for two years with one year for each state. In these cases, stagnation can be prevented by covering the collectors; however, this could influence the energy behavior of the occupants. In either case, simulations could again be used to approximate the measurement error.

This type of approach tries to utilize the statistical nature of the validation problem.\* It has the advantage that expensive instrumentation and data acquisition systems are not required, making it possible to sample the large number of systems necessary for statistical validation. Although a validation study based on highly instrumented systems has the potential of greater precision, the per system expense is much higher, and the results to date have not been promising. However, studies of the type discussed above do have the potential of achieving a level of precision greater than the simple approaches used in Section 5.2.

---

\*In another simulation context, the statistical nature of the solar energy system analysis problem has been used. In Lameiro (1979), a simple stochastic model has been used to approximate the annual performance of solar energy systems.

**SERIO** 

## SECTION 7.0

## CONCLUSION

This work has been concerned with an analysis of the statistical validation process for measuring the accuracy of design techniques for solar energy systems. In the view taken here, it is important to evaluate design technique performance in the context of the realistic process of system design. Statistical validation is an approach to measuring the accuracy of design methods in the environment defined by this process.

In this context, validation is not simply a matter of gathering some data and making comparisons to reality. In actuality, it involves:

- an assessment of how the design techniques are used in practice;
- an honest appraisal of the needs for a validation study;
- an appraisal of the precision required in the study;
- an initial formulation of questions to be answered and variation to be measured;
- a realistic assessment of measurement capability from both an engineering and fiscal point of view;
- a restriction of the study to variables that can be precisely defined, have an unequivocal relationship to the reality of interest, and can be easily measured;
- a precise definition of the target population of interest;
- a judicious choice of statistically meaningful data;
- an analysis of all sources of statistical error in the validation process;
- a candid interpretation of the possibilities and limitations of the results; and
- an assessment of the expected impact of the study, including the sensitivity of this impact to inaccuracies in the final results.

A statistical validation study is complicated by:

- the difficulty of measuring some of the variables of interest;
- the difficulty of separating measurement error and system performance problems from the problem of design accuracy;
- the difficulty of obtaining an accurate random sample of the system population of interest; and
- the rapidly changing state of the design methods and the solar energy industry.

There are two important questions that have not been considered in this report:

- Is it possible to perform an accurate statistical validation study?

The answer is "no" if the statistical validation process is expected to produce statistically accurate error estimates for all variables of interest for all the different solar energy systems on the market. However, it is the opinion of the author that by restricting the scope of the study to system variables that can be accurately monitored, by paying careful attention to statistical experimental design, and by making a candid interpretation of the meaning and accuracy of the results, it is possible to do statistical validation.

- Is it necessary to do an accurate statistical validation study?

The answer to this question is not so obvious. It really depends on the way the design techniques are used, the expectations of the users of the techniques, and the expected impact of such a study.

This work is primarily an attempt to present a framework to help evaluate the statistical precision of potential validations. Any validation can produce numbers. The question is: do these numbers accurately reflect the reality they propose to measure? Only with the proper attention to experimental design can statistical validation give good estimates of design accuracy.

## SECTION 8.0

## REFERENCES

- Anand, D. K.; Kennish, W. J.; Knasel, T. M.; Stolarz, A. C. 1979 (Aug.). "Validation Methodology for Solar Heating and Cooling Systems." Energy. Vol. 4 (No. 4): pp. 549-560.
- Cochran, W. 1977. Sampling Techniques. Third Edition. New York: John Wiley and Sons.
- Cummings, R. D.; Pech, D. J.; and Hall, W. J. 1978. Low Cost Monitoring System for Solar Water Heaters. Palo Alto, CA: Electric Power Research Institute; Er-795, Project 554-1.
- Delima, H.; Hirshberg, A.; and Dick, C. 1978. "A National Heating and Cooling Evaluation System." Conference Proceedings: Solar Heating and Cooling Systems Operational Results. Colorado Springs, CO: Nov. 28-Dec. 1. Golden, CO: Solar Energy Research Institute; Report No. SERI/TP-49-063.
- Fanney, A. H. 1978 (July). Experimental Validation of Computer Programs for Solar Domestic Hot Water Heating Systems. Center for Building Technology, National Bureau of Standards; Letter Report.
- Hansen, M.; Hurwitz, W.; and Madow, W. 1953. Sample Survey Methods and Theory, Methods and Applications. Vol. 1. New York: John Wiley and Sons.
- Hill, J. 1979. "Comparison of F-CHART Predictions to Performance of Solar Water Heaters." Presentation at May 1979 SSEA Working Group Meeting. Golden, CO: Solar Energy Research Institute.
- Kennish, W. J. 1978 (March). An Approach to the Validation of Computer Simulation Codes. Sandia Laboratories; SAND-78-0433.
- Kennish, W. J. and Stolarz, A. C. 1979 (Jan.). A Case Study of the Proposed Validation Methodology. Beltsville, MD: TPI, Inc.; SR#79-01.
- Klein, S. A. 1976. "Design Procedure for Solar Heating Systems." Ph.D. Thesis. University of Wisconsin.
- Klein, S. A.; Beckman, W. A.; and Duffie, J. A. 1976. "A Design Procedure for Solar Heating Systems." Solar Energy. Vol. 18: pp. 113-127.
- Knasel, T. M.; Kennish, W. J.; and Cassel, D. 1979a (March). Validation Methodology for Solar Heating and Cooling Simulation Models. Sciences Applications, Inc.; Contract EM-78-C-04-4261.
- Knasel, T. M.; Mansoor, Y.; and Kennish, W. J. 1979b (March). Validation Status of Solar Heating and Cooling Systems Models. Sciences Applications, Inc.; Contract EM-78-C-04-4261.



- Lameiro, G. F. 1979. "A Markov Model of Solar Energy Space and Hot Water Heating Systems." Solar Energy. Vol. 22: pp. 211-219.
- National Solar Data Program. ND. Solar Energy System Performance Evaluation Reports; National Solar Heating and Cooling Demonstration Program. U.S. Department of Energy. (Available from Huntsville, AL: IBM).
- Palmiter, L. S.; Hamilton, L. B.; Holtz, M. J. 1979 (Oct.). Low Cost Performance Evaluation of Passive Solar Buildings. Golden, CO: Solar Energy Research Institute; SERI/RR-63-223.
- Socolow, R. H. 1977/1978. "The Twin Rivers Program on Energy Conservation in Housing: Highlights and Conclusions." Energy and Buildings. Vol. 1: pp. 207-242.
- Solar Energy Research Institute. 1978. Conference Proceedings: Solar Heating and Cooling Operational Results. Colorado Springs, CO; Nov. 28-Dec. 1. Golden, CO: Solar Energy Research Institute; Report No. SERI/TP-49-063.
- Sonderegger, R. 1977 (Nov.). Diagnostic Tests Determining the Thermal Response of a House. Lawrence Berkeley Laboratory; LBL-6856.
- Streed; E. et al. 1976 (Aug.). Thermal Data Requirements and Performance Evaluation Procedures for the National Solar Heating and Cooling Demo Program. National Bureau of Standards; Pb-257-770, NBSIR 76-1137.
- United States Department of Energy. 1978. Conference on Performance Monitoring Techniques for Evaluation of Solar Heating and Cooling Systems. April 3-4; Solar Research and Development Branch and Solar Demonstration Branch.
- Ward, D. S. 1978. "Realistic Sizing of Solar Heating and Cooling Systems." Boer, K. W.; Franta, C. E., eds. Proceedings of the Annual Meeting of the International Solar Energy Society. August; Denver, CO. Vol 21: pp. 105-112.
- Ward, J. C. 1978. "Electricity and Gas Consumption of 24 Solar Homes Compared with 26 Conventional Homes Having Identical Heating Loads." Conference Proceedings: Solar Heating and Cooling Systems Operation Results. Nov. 28-Dec. 1; Colorado Springs, CO. Golden, CO: Solar Energy Research Institute; SERI/TP-49-063.
- Winn, C. B.; Parkinson, B. W.; and Duong, N. 1978. "Validation of Solar System Simulation Programs." Boer, K. W.; Franta, G. E., eds. Proceedings of the Annual Meeting of the International Solar Energy Society. August; Denver, CO. Vol. 2.1: pp. 120-124.

<b>Document Control</b> <b>Page</b>	1. SERI Report No. RR-721-377	2. NTIS Accession No.	3. Recipient's Accession No.
4. Title and Subtitle  Statistical Problems in Design Technique Validation		5. Publication Date April 1980	
7. Author(s) Joel Cohen		6.	
9. Performing Organization Name and Address  Solar Energy Research Institute 1617 Cole Boulevard Golden, Colorado 80401		8. Performing Organization Rept. No.	
12. Sponsoring Organization Name and Address		10. Project/Task/Work Unit No. 3525.40	
		11. Contract (C) or Grant (G) No. (C)  (G)	
15. Supplementary Notes		13. Type of Report & Period Covered Research Report	
16. Abstract (Limit: 200 words)		14.	
<p>This work is concerned with the statistical validation process for measuring the accuracy of design techniques for solar energy systems. This includes a discussion of the statistical variability inherent in the design and measurement processes and the way in which this variability can dictate the choice of experimental design, choice of data, accuracy of the results, and choice of questions that can be reliably answered in such a study. The approach here is primarily concerned with design procedure validation in the context of the realistic process of system design, where the discrepancy between measured and predicted results is due to limitations in the mathematical models employed by the procedures and the inaccuracies of input data. A set of guidelines for successful validation methodologies is discussed, and a simplified validation methodology for domestic hot water heaters is presented.</p>			
17. Document Analysis			
<p>a. Descriptors Solar Energy Systems; Measuring Methods; Validation; Design; Statistical Variability; Errors; Sampling; Accuracy; Solar Water Heaters; Solar Domestic Hot Water Systems</p> <p>b. Identifiers/Open-Ended Terms</p> <p>c. UC Categories</p> <p>58d, 59c</p>			
18. Availability Statement National Technical Information Service U.S. Department of Commerce 5285 Port Royal Road Springfield, VA 22161		19. No. of Pages 38	
		20. Price \$4.50	