*To be Presented at the 1979 Annual Meeting of the Urban & Regional Information Systems Association, San Diego, CA, August 19-24, 1979.*

LBL-9233 c.2

RECEIVED
LAWRENCE
BERKELEY LABORATORY

JUL 23 1979

LIBRARY AND
DOCUMENTS SECTION

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 6782

**Energy and Environment Division**

# Data Needs for Regional Residential Electricity Sales Forecasting

*Carl Blumstein*

May 1979

**Lawrence Berkeley Laboratory University of California/Berkeley**

**Prepared for the U.S. Department of Energy under Contract No. W-7405-ENG-48**

LBL-9233 c.2

## DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal 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 its 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, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-9233

# DATA NEEDS FOR REGIONAL RESIDENTIAL ELECTRICITY SALES FORECASTING*

Submitted to URISA '79 Proceedings
(Proceedings of the 1979 Annual Meeting
of the Urban and Regional Information
Systems Association,
August 19-24, Hotel Del Coronado
San Diego, California)

Carl Blumstein
Energy and Environment Division
Lawrence Berkeley Laboratory
Berkeley, California

*Research supported by the California
Energy Commission
(Contract 400-088 (7/8))

Carl Blumstein
Energy and Environment Division
Lawrence Berkeley Laboratory
Berkeley, California 94720

# DATA NEEDS FOR REGIONAL RESIDENTIAL ELECTRICITY SALES FORECASTING

ABSTRACT. In recent years increasing attention has been devoted to the problem of forecasting electricity sales. One consequence of this increased interest has been the advent of "end-use" models for forecasting residential electricity sales. These models disaggregate electricity sales into end uses such as heating, cooling, refrigeration, cooking and lighting. They make it possible to examine future electricity requirements in some detail. However, the use of these models requires very large amounts of data. Regional residential electricity sales models developed by the California Energy Commission provide an illustration of these requirements and of the difficulties that may be encountered in satisfying them.

## INTRODUCTION

Not too long ago a utility company planner interested in forecasting the sales of electricity in order to determine the need for new generating stations could do a pretty good job with a straight edge and a few sheets of semi-log graph paper. From the late 1940's to the early 1970's, sales grew steadily at a rate of about 8 percent. Moreover, as long as the error was on the high side, the accuracy of the forecasts was not crucial, since the lead time for power plant siting decisions was only a few years and any excess capacity would soon be eroded by the steady growth in sales.

Times have changed. Since the Arab oil embargo, growth in electricity sales has both decreased and become erratic. The compound growth rate for the five years since 1973 has been less than 4 percent; in 1974 sales actually declined slightly. Moreover, both the lengths of lead times and the costs for power plant construction have escalated dramatically.

Changes in the rate of growth of sales and in the costs and risks of new supply options have been accompanied by growing involvement by government in utility planning. Part of the reason for this involvement is that the inevitable consequences of exponential growth are catching up with us. As the absolute magnitude of each new increment of supply increases, social and environmental costs become more and more significant. Government has become involved in efforts to minimize these costs and in attempts to insure that necessary costs are shared equitably.

A second reason for governmental involvement is that growth in electricity sales is no longer perceived as simply an exogenous factor to be coped with. Various policies are being undertaken or planned to reduce the rate of growth. There is a considerable interest both in estimating the likely impacts of these policies before they are initiated and in monitoring their effectiveness once they have been implemented.

One consequence of the changing times has been a great increase of interest in methods for modelling and forecasting electricity sales. Attendant on this increased interest has been the development of new types of models and with them demands for new kinds of data describing the hows and whys of electricity use.

## END-USE MODELS

Of course, the modelling of electricity sales did not begin with the Arab embargo (for a review of pre-embargo efforts, see (4)), but work in this field was greatly intensified after 1974. The area that has received the most attention thus far is the residential sector. Probably the most significant development in this sector has been the advent of end-use models. Three models of this type were developed at about the same time in Oak Ridge (3), in Princeton (6), and in Berkeley and Sacramento for the California Energy Commission (1;2).

The name "end use" derives from the fact that these models attempt to forecast electricity sales for separate end uses. That is, they produce forecasts for refrigerators, freezers, space heating, air conditioning, etc. Conceptually, end-use models are quite simple. As illustrated in Figure 1, there are three basic components in such models. The households component provides forecasts of the number of households by housing type. The saturation component provides forecasts of the number of energy consuming units (i.e., appliances and space conditioning equipment) per household. The unit energy consumption (UEC) component provides forecasts of average electricity use per energy consuming unit. The electricity sales for each end use are given by the product of households, the number of energy consuming units per household, and the electricity use per unit.

End-use models have two important advantages. First, they tie electricity sales forecasts to the physical processes involved in energy consumption. Models that are based solely on the extrapolation of trends in factors such as household income, electricity price, and total residential sales may produce results that are not physically practical (e.g., do not take account of the fact that many major end uses have saturated). Second, end-use models make it possible for policy makers to examine the implications of specific conservation measures. For example, end-use models have been used to forecast the impacts of changes in building codes and of energy performance standards for appliances.

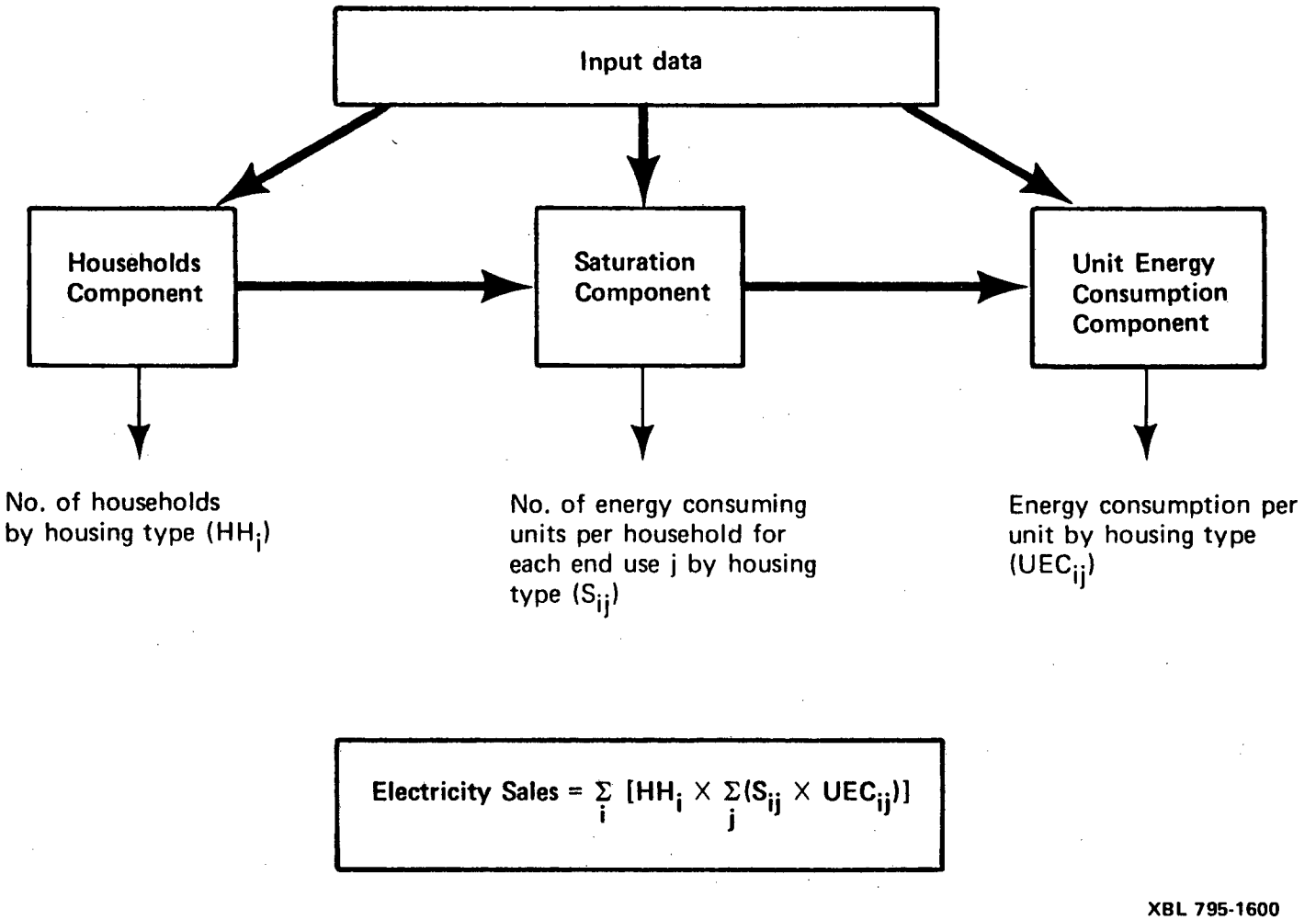



FIGURE 1. The components of an end-use model. Each component has data requirements. The saturation component also requires input from the households component and the unit energy consumption component requires input from the saturation component.

## DATA REQUIREMENTS FOR END-USE MODELS

End-use models have very substantial data requirements. Table 1 shows the types of data required by the California Energy Commission's model. The latest version of this model (5) makes forecasts for six utility service areas (the five major utilities and "other"). Since the saturation and energy requirements for space conditioning equipment are very dependent on climate, the service areas are further divided into climate zones. Eleven zones have been used to describe the California climate (see Figure 2). The overlay of these zones on the service areas produces fifteen forecast areas.

This geographical disaggregation is only the beginning. Since the saturation and energy requirements of end uses are dependent on housing types, households are disaggregated into single-family, multi-family, and mobile units. Moreover, there are twelve end uses in the model: refrigeration, freezing, cooking, hot water heating, dishwashing, clothes washing, drying, television, space heating, air conditioning, and pumping for swimming pool filters and sweeps.

Obviously, this can generate a lot of numbers. Twelve end uses times three housing types times fifteen forecast areas would give 540 separate end use forecasts. Things are not quite this bad in the California model since most end uses are only disaggregated to the service area level and housing types are not disaggregated in some forecast areas. Still, there are indeed quite a lot of numbers in the model.

Automation is, of course, essential to keeping manageable the problems created by these data requirements. It is fortunate that some of the most important data for the model are available on magnetic tape (especially the Sixth Count of the 1970 Census of Housing and the utility company surveys). Nevertheless, several person years of effort are required to get all of the data together. Consequently, running an end use model on this scale is an expensive proposition.

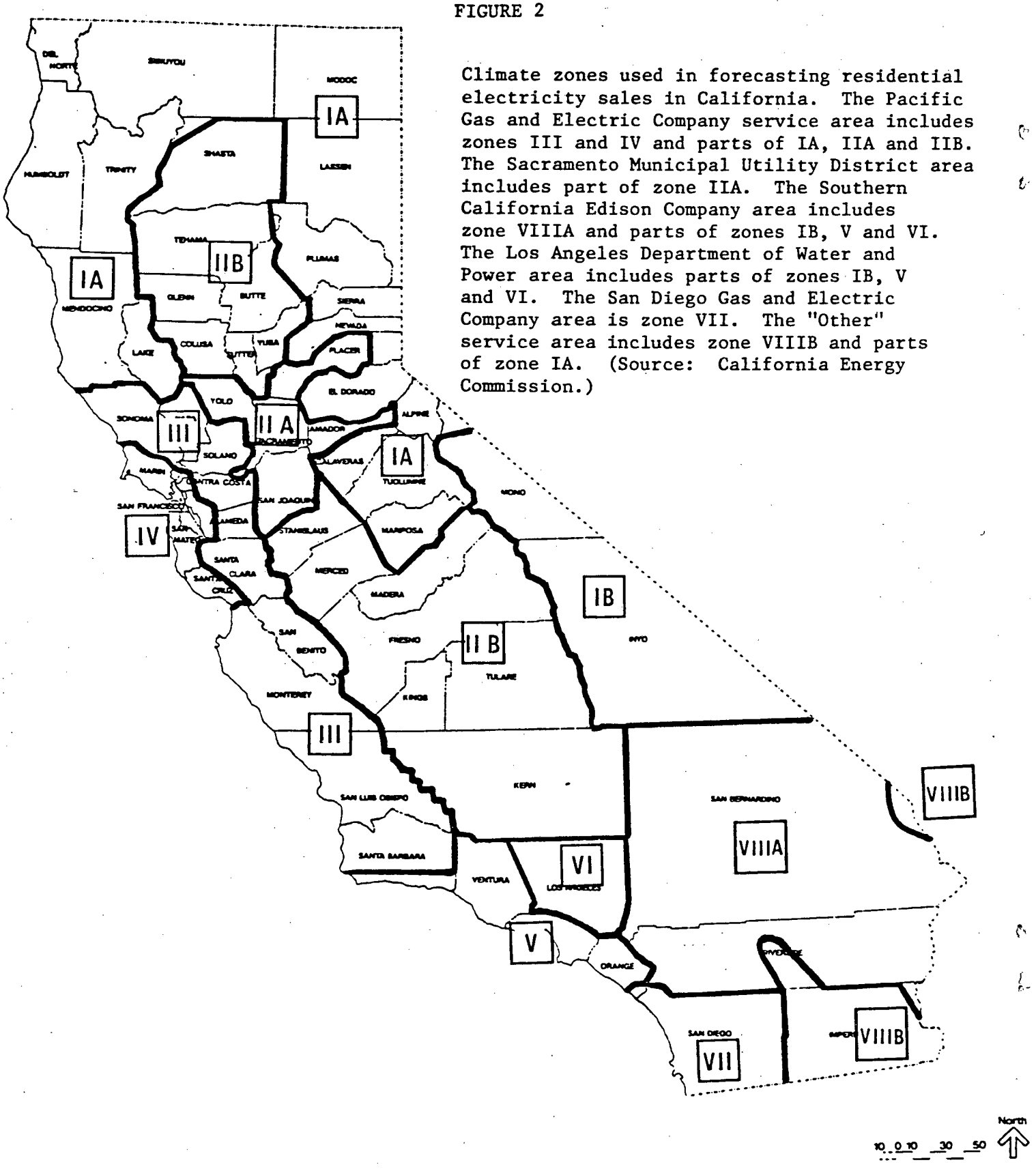
Existing data sources have proved reasonably adequate to the task of end-use modelling for the residential sector in California. Perhaps the best evidence for this is that it has been possible to reconstruct total electricity sales in the residential sector from end uses for each of the years 1975 to 1978 with errors of only a few percent (5). However, data in other regions of the country appear to be much less adequate. Even in California there are still quite a number of problems that will require the attention of modellers and information specialists for some years to come.

TABLE 1

DATA REQUIREMENTS AND SOURCES FOR THE CALIFORNIA ENERGY COMMISSION'S END-USE MODEL

| DATA | SOURCES |
|---|---|
| Households | Census of Housing<br>Calif. Dept. of Finance estimates |
| New Construction | Security Pacific Bank's "Calif. Construction Trends" |
| Removal Rates | Census of Housing (Components of Inventory Change) |
| Appliance Saturation | Census of Housing<br>Surveys by the major electric utilities in California |
| Appliance Lifetimes | Survey by the U.S. Dept. of Agriculture |
| Appliance Efficiency | Manufacturer's test reports<br>Estimates derived from analysis of utility surveys<br>Engineering estimates |
| Quality of Housing Stock (insulation, etc.) | Utility surveys |
| Climate | National Weather Service |
| Persons per household and household income | Bureau of Economic Analysis<br>Bureau of the Census<br>Calif. Dept. of Finance |
| Electricity Prices | Utility company reports |

FIGURE 2




Climate zones used in forecasting residential electricity sales in California. The Pacific Gas and Electric Company service area includes zones III and IV and parts of IA, IIA and IIB. The Sacramento Municipal Utility District area includes part of zone IIA. The Southern California Edison Company area includes zone VIIIA and parts of zones IB, V and VI. The Los Angeles Department of Water and Power area includes parts of zones IB, V and VI. The San Diego Gas and Electric Company area is zone VII. The "Other" service area includes zone VIIIB and parts of zone IA. (Source: California Energy Commission.)

## SOME PROBLEMS WITH EXISTING DATA SOURCES

Problems with the data sources used for the California model fall into three categories: completeness of the sources, reliability of the sources, and comparability between sources.

### Completeness

In some cases, data that would be quite useful are not available. For example, the California model requires data on the saturation of appliances in 1970 and 1977. Most of the necessary 1970 data are available from the Census of Housing. However, the Census does not contain information on the number of refrigerators per household. One might suppose that this is no great problem and that the number was probably very nearly 1.0, but in 1977 the number of refrigerators per household in California was closer to 1.2 and there is evidence to suggest that the number was significantly greater than 1.0 in 1970.

Another example of a "hole" in the data is fuel availability. The availability of natural gas is of course a critical factor in a homebuilder's decision on what fuel to use for space heating and water heating. Surprisingly, even the gas utilities do not have a good handle on the fraction of households within their franchises that do not have gas available (i.e., are too far from a gas main to obtain service at a reasonable cost).

It is unlikely that the data sources will ever be complete enought to satisfy the modellers. Many of the data related to past energy usage patterns are beyond recall. Moreover, new data requirements keep coming up. For example, recent developments in fuel choice have made it desirable to know whether the fuel used for a household's water and space heating systems was chosen at the time of construction or at some later time. Still, there is reason to hope that there will be significant improvements in the completeness of the sources. The key is closer collaboration between the modellers and the survey researchers. Some efforts, both in California and Washington, D.C., have already been made to bring this about.

### Reliability

Reliability is always a concern when working with survey data. As the data analyst quickly learns, there are a lot of ways that data reliability can be compromised. The list runs from biases in the sample through incorrect answers from respondents to improper management of the raw data. Trouble was experienced in all of these areas during the development of the California model.

The greatest concern has been with incorrect responses. The most serious problem area has been space heating fuel choice. The responses in the utility surveys that are used to determine fuel choice are matched with the respondents utility bills. Thus, it is possible to check whether those who report using electricity as

their main heating fuel have winter bills that are consistent with electric heating. A sample of respondents reporting electric heat was not reassuring on this point. A significant percentage (more than forty percent) had bills that appeared to be inconsistent with electric heating.

Ceiling insulation is another area where survey responses are suspect. The surveys indicate the presence of much more ceiling insulation than is found by actual inspection in the field. However, field work thus far has been rather limited in scope so the seriousness of this problem is still uncertain.

There is little that can be done by modellers or survey researchers to get at the root cause of incorrect responses (i.e., lack of knowledge on the part of respondents), but there may be ways to reduce the frequency of incorrect responses or at least to detect and compensate for them. First, careful attention should be given to the design of survey questions. A well constructed question may at least elicit a "don't know" rather than a wrong answer. [For example, it is probably better to ask whether a household is served by "utility gas" rather than "natural gas" since, as indicated by responses from areas with no natural gas, there appears to be some confusion between natural gas and liquified petroleum gas (LPG).] Second, more attention needs to be given to the development of methods for checking responses. These checks can be internal to the surveys (consistency checks) or external (field work and other sources).

Comparability

The utility surveys on which the California model relies for much of its information on appliance saturation had their origins in the marketing departments of two of the State's major utilities. The surveys were originally designed for market research and little, if any, thought was given to using the results in combination with other data for end-use models. As a consequence, it is sometimes difficult to compare data from the utility surveys with data from other sources such as the Census of Housing.

Comparability is important for two reasons: First, many of the model's results are based on the analysis of trends in parameters that are determined in 1970 by the Census and in 1977 by the utility surveys. If the questions on the Census and the utility surveys are different, the apparent trend may be biased. There are unfortunately quite a few cases where the questions are different (e.g., the Census asks about utility gas and the surveys ask about natural gas, the Census asks about automatic clothes washers and "wringers or spinners" and the surveys ask only about clothes washers).

Second, if the utility survey data are comparable with data from other sources, it will be easier to detect and adjust for biases in the surveys. For example, there has been some suspicion

that the responses in utility surveys are biased toward higher income households. In principle, it should be possible to stratify the responses by income and then to correct for the bias by weighting the strata according to the best estimate of the true income distribution. However, this would probably be quite difficult in practice because the income questions on the surveys are not well matched to income data from other sources.

Most problems with comparability are probably avoidable. The first rule is, don't phrase questions differently from the Census unless there is a very good reason for doing so. When it becomes necessary to break with the conventions used by other sources (or when other sources use conflicting conventions), it is important to consider the comparability problem carefully. A lot of work and uncertainty can be avoided if ways to preserve comparability are not overlooked when a survey is designed.

## CONCLUSIONS

The field of energy modelling is still in its infancy. This brief report has noted only a few of the areas that will require attention as the field advances. But even the briefest discussion should mention one of the most important lessons learned thus far: success depends on interdisciplinary teamwork. The California model is the result of a collaboration between engineers, physical scientists, economists and information specialists. Progress on the problems discussed above will require a still broader collaboration. In particular, the modellers and survey researchers must work together. The former must find better ways to communicate the requirements of their models, the latter must devise new ways to meet these requirements as well as to communicate both the value and the limitations of their results.

## ACKNOWLEDGEMENTS

Much of the information in this paper was obtained while the author was on leave at the California Energy Commission. Thanks are due to Don Carner, Scott Cauchois, Steve George, Tom Gorin, Barbara Greene, Mike Jaske, Lucinda Larson, Dale Paige, and Sarita Sarvate who worked with the author on the Commission's residential electricity sales model and to Thom Kelly and Bob Weatherwax who supervise the Commission's modelling efforts. All of these people provided both good advice and friendship. Ed Kahn, Will Siri, and Carl York read a draft of this paper and made helpful comments. Jan Bacon prepared the manuscript. The opinions expressed and any errors that may be contained in this paper are the responsibility of the author.

This work was supported by the U. S. Department of Energy.

## REFERENCES


(1) Berman, S.M., C. Blumstein, et al., (1976) Electricity Consumption in California: Data Collection and Analysis, Berkeley, California, Lawrence Berkeley Laboratory, UCID-3847, 1976.

(2) Goett, A., et al., (1976) Residential Electricity Forecasting Model: Technical Documentation, Sacramento, California, California Energy Commission, November 1976.

(3) Hirst, E., and J. Carney (1978), The ORNL Engineering-Economic Model of Residential Energy Use, Oak Ridge, Tennessee, Oak Ridge National Laboratory, ORNL/CON-24, July 1978.

(4) Taylor, Lester (1975), "The Demand for Electricity: A Survey", Bell Journal of Economics, Vol. 6, No. 1, Spring 1975.

(5) Weatherwax, R.K., et al., (1979), Second Biennial Forecast of Electricity Sales for California, Sacramento, California, California Energy Commission, in preparation.

(6) Weatherwax, R.K., and R.H. Williams (1977), Energy Conservation/Load Management Analysis for the Residential Sector of Consolidated Edison Electric Franchise District (Draft), Princeton, New Jersey, Center for Environmental Studies, Princeton University, March 1977.

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

*TECHNICAL INFORMATION DEPARTMENT*
*LAWRENCE BERKELEY LABORATORY*
*UNIVERSITY OF CALIFORNIA*
*BERKELEY, CALIFORNIA 94720*