End Use Load Data Transferability: A Report from the Frontier

Marisha E. Chilcott and Nancy E. Ryan, Quantum Consulting Inc.
Richard Gillman, Center for Electric End-Use Data

End-use load data are increasingly valuable for a wide range of applications from DSM program evaluation to forecasting and program screening. Despite recent technological advances, end-use metering remains relatively expensive and time consuming. This paper explores one option for obtaining much of the value of a metering study through end-use load data transfer from "donor" to "recipient" utilities.

The paper extends previous research in this area by transferring data to recipient utilities for which metered end-use data are available. It presents results of applying previously developed transfer methods to metered end-use load data from a donor utility and measuring the validity of the results with the recipient utility's metered data. Two regression-based data transfer methods were applied to refrigeration, water heating and air-conditioning load data from different regions of the country. The findings from this study have implications about how existing data transfer methods should be implemented and offer promising directions for future research into novel approaches to data transfer.

Introduction

End-use load data are increasingly valuable for a wide range of applications from demand side management (DSM) program evaluation, to forecasting and program screening. However, despite recent technological advances, end-use metering remains relatively expensive and time consuming. At present there is a continuum of methods through which load data needs can be met, including conditional demand analysis, engineering simulations, disaggregation, data transfer and end-use data metering. This paper explores the emerging option of meeting such needs through transferring end-use load data from "donor" to "recipient" utilities.

Data transfer involves using metered end-use load data for a purpose other than that for which it was originally collected. One application of data transfer is to use load data collected in one region to meet a data need in another region; for example, developing end-use load shapes for DSM evaluation or planning purposes. Another application is to use data collected in one part of a utility's service territory to evaluate a targeted DSM program in a different sub-region of the same utility's service territory. The basic assumption underlying data transfer is that cross sectional differences in end-use load shapes can be controlled for through a modeling framework that incorporates data on weather, appliance characteristics, customer behavior, and other explanatory variables.

The current state of the art of load data transfer, the frontier, was evaluated by applying existing data transfer methods under conditions where the "answer" was known. This is a novel feature of our approach which sets it apart from earlier research on end-use load data transfer.[1] Two different data transfer methods were applied to each of three residential end uses, refrigeration, water-heating, and air-conditioning. Results for water heating and air conditioning are presented here.[2] For each end use, metered data were available for both the donor and the recipient utilities. Thus, it was possible to assess the results of applying these data transfer approaches by comparing the transferred load data to actual metered load data for the recipient.

A further advantage of this experimental design was that it afforded the opportunity to investigate explanations for unsuccessful data transfer efforts. Indeed, as the project progressed it became evident that the most useful results would not be whether or not the selected methods "worked" but rather what advances could be made toward a better understanding of what drives the particular end-use load shapes that were transferred. Performing these data transfer exercises also permitted the researchers to see the potential scope of what a data transfer method must be able to address and what issues arise when data transfer is attempted under different conditions. This scope varies by the end use, the distance, and the purpose of data transfer, where distance can be either a geographic or a temporal measure.

The remainder of this paper is organized into three sections. The following section describes the data transfer methodologies, the next presents the key results for each end-use, and the final section discusses conclusions and their implications for ongoing research into data transfer.

Methodology

Selection of Data Transfer Methods

A data transfer methodology must provide a framework for adjusting the donor's load data to account for the effects of differences between the donor's and the recipient's service territories. Among the factors which may cause cross-sectional variation in patterns of end-use energy consumption are differences in weather, customer demographics, dwelling characteristics and appliance portfolios. Different types of variables drive loads for different end uses. For example, among the factors which affect air conditioning loads are weather and dwelling characteristics (e. g. square footage, insulation, building orientation and number and size of windows). These factors are relatively unimportant (within a season) for water heating. Loads for this end use depend in part on ownership of other appliances (e.g. dishwashers and clothes washers), a factor which is relatively unimportant for explaining air conditioning loads.

The effects of the relevant driving variables for an end use may be captured through engineering models, statistical approaches or a combination of the two. The first phase of this project consisted of an extensive survey of existing data transfer methods and applications of data transfer.[3] Following the literature survey, two statistically based approaches were selected for use in this study. These methods were selected because of their flexibility and because they represented the state-of-the-art, having recently been developed and used in data transfer projects. Both methods employ a statistical approach in which a combination of stratification and regression analysis are used to capture the effects of the variables which drive end-use load shapes. Stratification is used to control for the effects of customer behavior, dwelling characteristics and appliance ownership while the regression component of the methodology accounts for differences in weather and residual temporal variation in load shapes.

Because the methods employ regression analysis, they tend to perform better in predicting average load shapes. Method 1 consists of estimating 24 hour-by-hour regressions of hourly load on cooling or heating degrees, while Method 2 involves fitting a single regression equation in which hourly loads are modeled in terms of harmonic functions of time and a function of the temperature-humidity index.[4] In principle, characteristic variables may also be included in both regression specifications to account for the influence of driving variables other than weather.[5] It was determined, however, in preliminary analysis of the data that these models' implicit assumptions regarding daily patterns of end-use loads and weather responsiveness were valid only for averages across customers (air conditioning) or days within a daytype and season (refrigerators and water heating). Thus, customer-specific characteristic data could not be incorporated in the models explicitly, but had to be accounted for through stratification.

Method 1 may be summarized as follows.

$$L_{hdt} = a_h + b_h HD_{dt} + c_h CD_{dt} + e_{hdt} \tag{1}$$

where:

L_{hdt}	is the average kW load for stratum h at time t on day d;
HD_{hdt}	is heating degrees at time t on day d (calculated as the maximum of zero and the thermostat setpoint minus the dry bulb temperature);
CD_{hdt}	is cooling degrees at time t on day d (calculated as the maximum of zero and the dry bulb temperature minus the thermostat setpoint);
a_h, b_h and c_h	are the regression coefficients for strata h;
and e_{hdt}	is an independently and identically distributed error term.

For appliances which are not weather sensitive, Method 1 reduces to computing average hourly loads for each stratum. This approach offers several advantages. Most important, it imposes no temporal structure on the average load shape. It also allows for variation in patterns of weather response throughout the day. Although it allows for the realistic possibility that the variance of the error term is not constant over the course of the day, this modeling framework also embodies the less realistic assumption that the error term is independent across hours of the day.

Method 2 offers a more compact specification at the expense of making more restrictive assumptions about the underlying structure of end-use load shapes and their dependence on the weather. In Method 2 a single regression equation is estimated for each stratum. This approach may be summarized as follows:

$$
\begin{aligned}
L_{hdt} = {} & b_{h0} + b_{h1}\sin(t\pi/12) + b_{h2}\cos(t\pi/12) \\
& + b_{h3}\sin(t\pi/6) + b_{h4}\cos(t\pi/6) \\
& + b_{h5}\sin(t\pi/3) + b_{h6}\cos(t\pi/3) \\
& + b_{h7}FTHI_{dt} + e_{hdt};
\end{aligned} \tag{2}
$$

where:

L_{hdt}	is the average kW load for stratum h at time t on day d;
$FTHI_{hdt}$	is a transformation of the temperature humidity index (THI) analogous to cooling degrees, the maximum of zero and THI minus a switch-point;
b_{h0}, b_{hl}	are the regression coefficients for each stratum;
and e_{hdt}	is an independently and identically distributed error term.

The periodicity of the average load shapes is captured through the inclusion of harmonic functions of time. For weather sensitive end-uses the term based on the Temperature Humidity Index (THI)[6] is included in the model to control for the dependence of end-use consumption on weather. The cost of this model's greater simplicity is that it implicitly assumes that the both the weather response and the variance of the error term are constant throughout the day.

Application of Data Transfer Procedures

A three-step procedure was used to apply these methods to each end use.

First stratification variables were selected from the set of characteristic variables for which data were available for both the donor and recipient. Stratification variables were chosen based on their ability to explain cross-sectional variation in end-use load shapes within the donor utility's sample. Priority was placed on choosing stratification variables expected to have a structural (e.g. causal) relationship to end-use loads. If no structurally related variables were available for both utilities, instrumental variables were explored as an alternative.

The second step of the process involved application of the regression-based components of Methods 1 and 2 to each stratum. For weather sensitive end uses, applying Method 1 consisted of fitting Equation 1 to each hour of the day for each stratum. For the weather-insensitive end uses, this approach reduced to calculating an average hourly load for each stratum. Applying Method 2 involved fitting Equation 2 to each stratum. For both methods, fitting separate equations for each stratum allowed for differences in the weather response across strata.

In the last stage of the data transfer process the fitted models were applied to the recipient utility's data. For weather sensitive appliances, transferred average hourly loads for each stratum were derived by using the recipient's weather data as input into the fitted model for each stratum. For weather-insensitive appliances, the donor's average hourly loads for each stratum were transferred directly to the corresponding strata in the recipient utility. Average load shapes for the recipient were then developed by applying the recipient's strata weights to the transferred stratum loads.

Data and Results

Overview

The end-use metered load data used in this study were collected from utilities in each of three regions of the U. S., the West, the Northwest and the Midwest. Refrigeration data were transferred between the West and the Northwest, water heating data were transferred between the Northwest and the Midwest, and air conditioning data were transferred between the West and the Midwest. All of these load data and their accompanying characteristic data were provided by the Electric Power Research Institute (EPRI) Center for Electric End-Use Data (CEED) Data Request Service (DRS.) Some weather data for these regions were also available through the DRS. Additional weather data were obtained from the National Oceanic and Atmospheric Administration (NOAA).

Water Heating

Water heater load data from the recipient (Northwest) and the donor (Midwest) were compared to one another before data transfer was attempted. Figure 1 compares average metered winter weekday load profiles for water heaters from these two regions.

The average load profiles from these two regions are very similar to one another. This is particularly interesting since the morning peak loads of greater than 1.0 kW are high in comparison to water heater loads which these researchers have observed in data from other parts of the country. The first step of the data transfer process revealed that of the available characteristic variables, the following were correlated with differences in load for these data sets:

- number of occupants and
- ownership of a dishwasher.

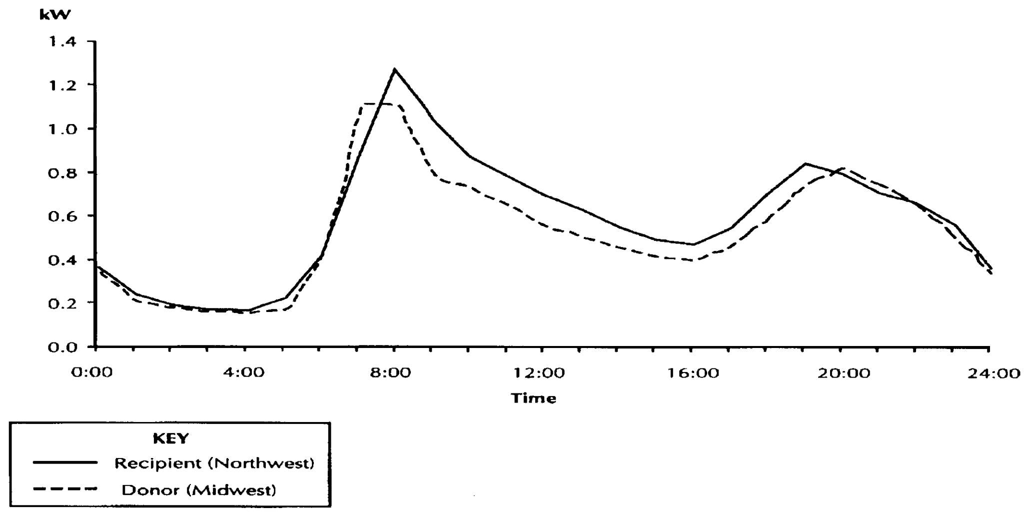

Figure 1. Average Winter Weekday Water Heater Load Shapes, Donor (Midwest) and Recipient (Northwest)

In addition to these variables, the size of the home, the presence of a clothes washer and the tank size were also examined for correlations with the average load data. As was the case with refrigerators, the data for these variables were either all missing, all the same or incompatible between the two regions' data sets.

Both Methods 1 and 2 were applied to the donor's (Midwest) data using household size (1-2 people versus 3 or more people) and ownership of a dishwasher as stratification variables. For all seasons and day-types, Method 1 performed considerably better than Method 2. Figure 2 shows results of using Method 1 to transfer a load shape for winter weekdays.

As shown in the figure, the transferred load shape matches the recipient's (Northwest) metered data more closely than the donor's (Midwest) unadjusted average load shape. The improvement is limited to the morning peak hour.

The relatively successful application of data transfer to water heating provides a positive demonstration of the value of having complete and compatible data for driving variables. In addition to the stratification variables which were actually used here, the investigation of potential stratification variables also yielded insights into other important driving variables for water heating (several of which were mentioned above). Another variable which

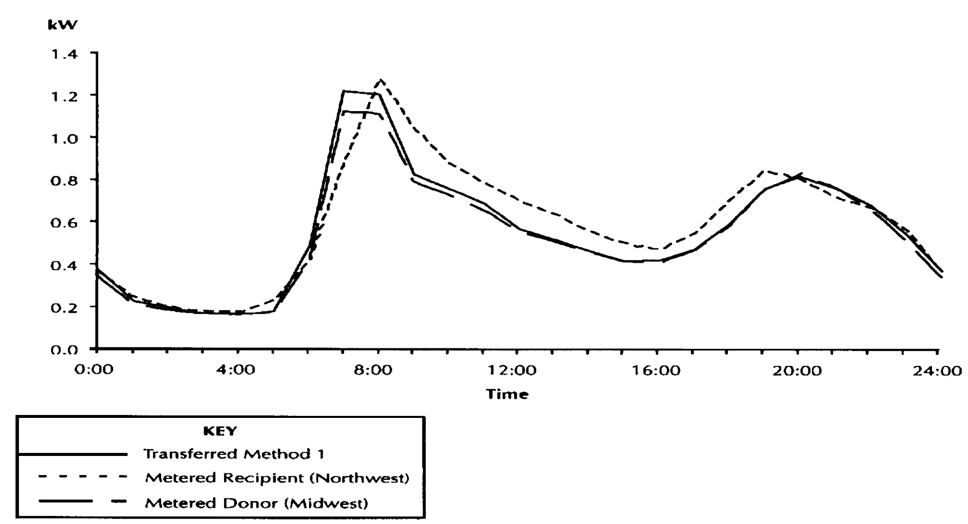

Figure 2. Results of Water Heater Data Transfer, Transferred vs. Metered Recipient (Northwest) and Donor (Midwest) Load Shapes, Winter Weekdays

may be important, is the temperature of the intake water. This factor may vary across seasons and regions (or even within regions). The implications of the energy requirements due to differences in intake water temperature are 2.44 Wh, per gallon of water, per °F. This translates into a difference of approximately 100 Wh per 40 gallon water heater full, per °F. This variable should be considered when transferring water heater data between regions where the water intake temperatures may be different.

Air Conditioning

The data transfer for air conditioning was implemented using central air conditioner and heat pump load data gathered during the summer of 1990. The donor utility was located in the Western U.S. and the recipient utility was located in the Midwest. The donor (West) and recipient (Midwest) samples contained 54 and 39 customers respectively.

Transferring air-conditioning loads between two distant and very different regions presented the greatest challenge of this research project. An indication of the difficulty of this undertaking is provided by Figure 3, which depicts average weekday air-conditioning load profiles for the donor (West) and recipient (Midwest) utilities. During afternoon and evening hours, the recipient's (Midwest) air-conditioning loads were significantly higher than the donor's (West). Also, the peak hour was later for the recipient (Midwest), and following the peak, the recipient's (Midwest) air-conditioning loads tapered off more gradually than the donor's (West).

Ideally the data transfer approach would make use of both weather and characteristic data to explain the differences between patterns of air-conditioning consumption in the donor's (West) and recipient's (Midwest) samples. Unfortunately very limited characteristic data were available for the recipient (Midwest) utility. Premise size (measured in square feet) was the only potential stratification variable common to both data sets. *Ex ante* air-conditioning consumption could be expected to be positively correlated with premise size (since larger units and/or more intensive operation would be required to cool a larger structure). Premise size was rejected as a stratification variable, however, because this expected relationship was not observed for the donor (West) utility (on average larger homes actually had lower air-conditioning consumption than medium sized homes)

Two other potential stratification variables were also considered and rejected. These were total household energy consumption and the size of the air-conditioning units as measured by connected load. Total household energy consumption could not be used for stratification purposes because there was little overlap between the two samples in the range of this variable: most houses in the donor (West) sample used more energy than all but the largest premises in the recipient (Midwest) sample. Connected load could not be used as a stratification variable because connected load could not be computed for a high proportion of the houses in the donor sample. The connected load is usually calculated as the maximum load observed for an appliance over a season.[8] However, because many of the customers in the donor (West) sample appear never to have operated their air-conditioners, their connected loads were not observable. Clearly the differences between the two samples, revealed through the process of screening potential stratification variables, will play an important role in understanding the results of the attempt to transfer air-conditioning load data.

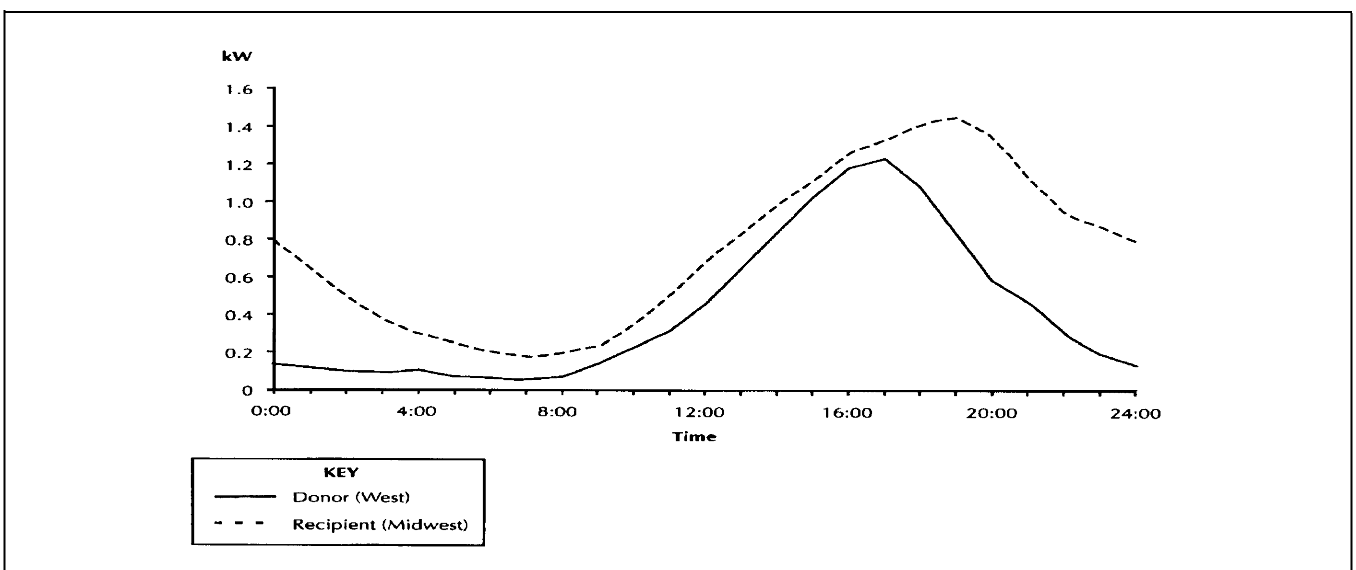

Figure 3. Average Summer Weekday Air Conditioner Load Shapes, Donor (West) and Recipient (Midwest)

Due to the lack of suitable stratification variables, Methods 1 and 2 were applied directly to average hourly air conditioning loads for the entire donor sample (averages were computed across customers for each day). Thus only a partial data transfer for air-conditioners was undertaken. This transfer only took into account the role of differences in weather between the two regions. Humidity is also an important determinant of air-conditioning loads. This variable was accounted for in Method 2 through the use of the THI as a regressor. It could not be included in Method 1 because hourly humidity and dry-bulb temperature readings were highly correlated within each sample. [9]

The results demonstrated that adjusting for differences in weather via these methods can account for only a small part of the regional differences in air-conditioning load profiles. Results for days with comparable weather provide the best basis for assessing the performance of the data transfer approaches. In general Method 1 outperformed Method 2. Figures 4 and 5 present comparisons between transferred load shapes (Method 1) and average load shapes derived from the donor's (West) and recipient's (Midwest) metered load data. Results for days with a maximum temperature of 940 F (the hottest day for which load data were available for both samples) are shown in Figure 4, while Figure 5 presents results for days with a maximum temperature of 88 "F. As suggested by the figures, Method 1 performs better on less extreme days. On these days the transferred load shapes generally fit the recipient's (Midwest) data better than the donor's unadjusted data. Neither method yields a transferred load profile which closely tracks the recipient's (Midwest) metered load shape; nor do these methods accurately predict either the magnitude or the timing of the recipient's (Midwest) peak air-conditioning load. Method 1 performs better than Method 2, especially in capturing the relatively gradual decline of air-conditioning loads during the early evening hours.

In addition to providing a basis for assessing the results of the data transfer, the availability of metered air-conditioning data for the recipient (Midwest) afforded the means to explore explanations of why the transfer methods did not perform better. Several possible explanations were considered in this investigation. The most important finding was that customers in the two samples demonstrate different patterns of response to variations in the weather. Figure 6 illustrates this finding using data from both utilities. In the figure average air-conditioning loads at noon on each summer weekday are plotted against coincident dry bulb temperatures. The fitted regression equations from applying Method 1 to each data set are also included in each figure. The figure shows that in the recipient (Midwest) sample customers begin to cool their homes at lower temperatures (the "set-point" is lower for the recipient) and that on average these customers display greater sensitivity to rising temperatures (the regression line is steeper for them).

This finding contradicts a central assumption of both approaches to data transfer employed in this study—that the same underlying model of the relationship between end-use loads and weather variables applies to both utilities.

There are several possible explanations for this finding. First, there may be important differences between the customers represented in the two samples which could not

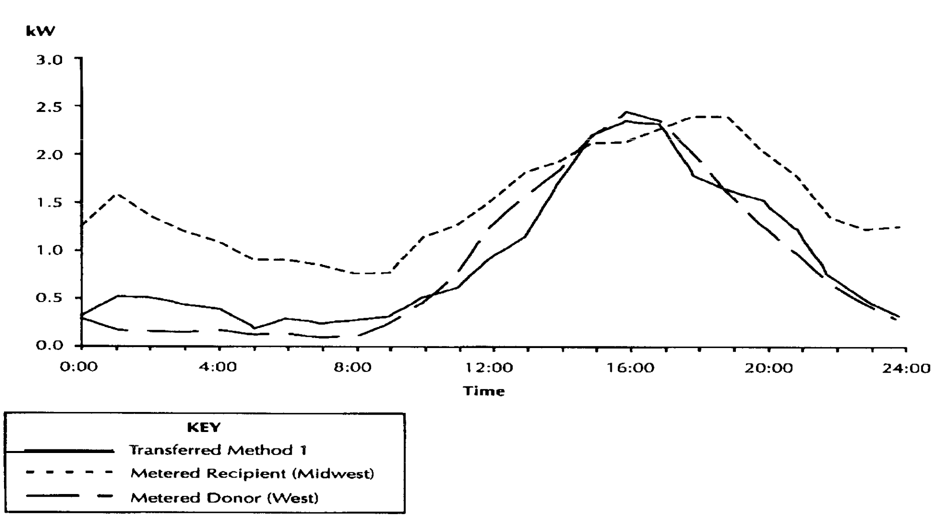

Figure 4. Results of Air Conditioner Data Transfer Transferred vs. Metered Recipient (Midwest) and Donor (West) Load Shapes Weekdays with a Maximum Temperature of 94°F

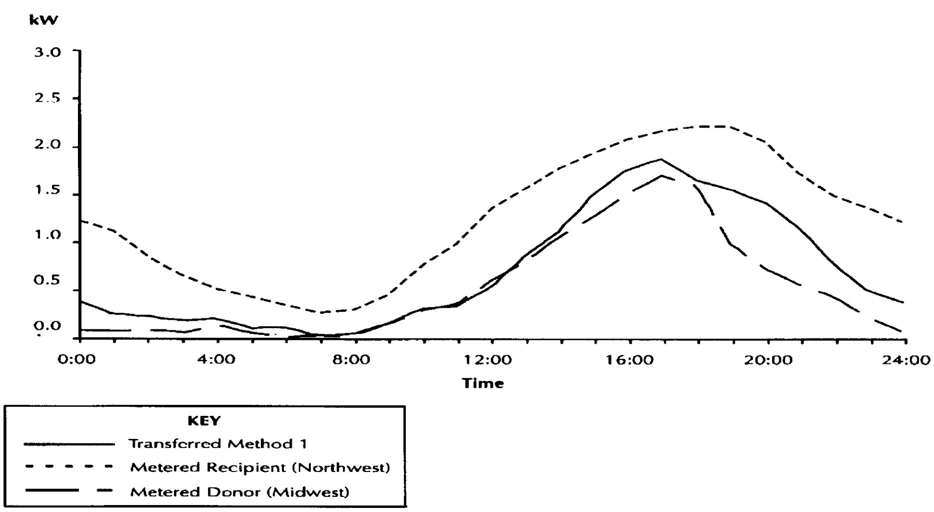

Figure 5. Results of Air Conditioner Data Transfer, Transferred vs. Metered Recipient (Midwest) and Donor (West) Load Shapes, Weekdays with a Maximum Temperature of 88°F

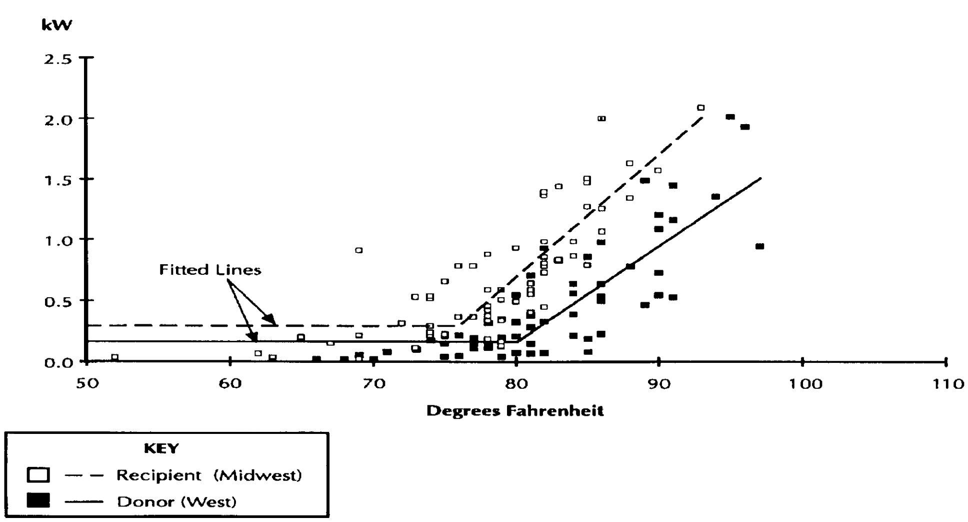

Figure 6. Air Conditioning Load vs. Dry-Bulb Temperature at Noon, Donor (West) and Recipient (Midwest)

be taken into account due to the lack of overlapping characteristic data. That such differences exist was strongly suggested by the observed differences in the distributions of total premise energy consumption and air-conditioner connected loads. Another closely related possibility is that customers in different regions experience the same weather differently and therefore respond to it in different ways. This may be due to different personal experiences with weather, regional variation in dwelling characteristics, or the inability of these models to adequately account for the role of humidity.[10] For the samples used in this study, differences in dwelling characteristics were probably of great importance. The donor (West) utility is located in a state which has very high efficiency standards for new buildings. Also, over the last decade economic growth has been much more rapid in the donor's (West) service territory than in the region where the recipient (Northwest) is located. Thus, the customers in the donor (West) sample are more likely to live in newer homes and to have more efficient appliances. Both of these explanations should be investigated further in future research on data transfer.

Conclusions

This study has yielded several conclusions regarding the current state of the art of end-use data transfer. These findings have implications about how existing data transfer methods should be implemented and offer promising directions for future research into novel approaches to data transfer.

This study has underscored the importance of incorporating appropriate data on customer demographics, dwelling characteristics and appliance holdings in a data transfer method. Ideally the characteristic variables employed in the data transfer process should be causally related to the load shape to be transferred. This was the case in the relatively successful water heater transfer described in this paper. Data collection to support data transfer should focus on ensuring the availability of compatible data for such variables. Surveys and/or audit instruments should be designed and administered with a view toward obtaining high response rates for these data elements.

Another key finding is that future research should focus on developing transfer methods for weather sensitive end uses which allow for differences in the response of energy consumption to weather variables. Improved methodologies may need to draw on engineering models which take into account dwelling and appliance characteristics, and can better account for the effects of humidity.

Finally, an important insight acquired through this study is that flexibility is required in developing and implementing data transfer methods. In future load data transfer applications the choice of methodology should be determined bearing in mind both the nature of the load shape to be transferred and the purpose for which the transferred data are to be used. Just as different modeling frameworks work best for different types of end uses, different methodologies may be appropriate for different applications of transferred data.

Acknowledgments

The authors gratefully acknowledge support for this project provided by the Electric Power Research Institute under Research Project number RP 2980-14.

Endnotes

1. Earlier studies which evaluated the results of residential and commercial whole-premise data transfer using metered data were Moe 1992, AEIC 1982, and Linder and Breese 1984.

2. More detailed results for these end uses, as well as results for refrigerators, are presented in Ryan et al. 1994.

3. Results of this survey are presented in Chilcott et al. 1993.

4. Method 1 was originally developed to transfer total premise load shapes. See Powers et al. 1992. Method 2 was originally developed to transfer ground source heatpump load shapes. It is presented in Moe 1992.

5. Indeed, Method 2 was originally conceived as a two step regression approach. In the first step, coefficients are estimated for the weather variables and harmonic functions of time. These coefficients are then regressed on characteristic variables in the second step.

6. THI = 0.55*dry bulb temperature + 0.2*dewpoint temperature + 17.5

7. An instrumental variable is one which is correlated with both the dependent variable and the unobserved driving variable, but is not itself structurally related to the dependent variable.

8. If the maximum appears to be an extreme outlier then the 99th or 95th percentile of the appliance's load may be used instead.

9. The implication of this multicollinearity was that the effect of humidity could not be distinguished from that of dry-bulb temperature. In principle the *joint effect* of these two variables could be estimated by including both in the Method 1 specification. However, this was not justified because the relationship between the two variables differed across regions. Consequently the joint impact measured for one area would not be transferable.

10. Developing techniques which explicitly account for the influence of humidity is one way of addressing this issue.

References

Association of Edison Illuminating Companies. 1982. Data Transferability, *Load Research Committee Report on Methods Used to Develop Current Or Future Load Patterns Based on Historic or Borrowed Data* (Papers A, E).

Chilcott, M. L.; Mullin, C.; and Wilson, A. 1993. *Inventory of Load Data Transfer Methods,* Final Report prepared by Quantum Consulting, Inc. for the Electric Power Research Institute under Research Project PR 2980-14.

Linder, K.P. and Breese, J.S. *1984. Load Data Transferability,* Electric Power Research Institute Report EA-3255, ICF Incorporated.

Moe, Ronald J. 1992. *Draft Final Report, Development of Load Shape Transferability Methodology,* Synergies Resources Corporation.

Powers, J., and Lockwood, M. 1992. *Workshop on Load Data Transferability: Mission Impossible or Not?* Report prepared by Quantum Consulting Inc. for the Center for Electric End-Use Data under EPRI Research Project 2980-14.

Powers, John, et al. 1992. *Drafi Final Report, The End-Use Technology Assessment Project*, EPRI Research Project 2342-8, Quantum Consulting Incorporated, 1992.

Ryan, N. et al. *1994. Residential End-Use Load Data Transfer: An Assessment of Current Methods.* Report prepared by Quantum Consulting for the Center for Electric End-Use Data under EPRI Research Project 2980-14.