Using Geographic Information Systems to Support Market Transformation

Robert M Wirtshafter, Wirtshafter Associates, Inc.
Susan L. Radke, Berkeley Geo-Research Group, Inc.
Shahana Samiullah, Southern California Edison

ABSTRACT

Utility-sponsored energy-efficiency market-transformation programs use many types of geographic information in the planning, implementation, and evaluation of the programs. Unfortunately in most cases, there is no coordination of the spatial data across applications. This not only makes for confusion or lost information, but also limits the analysis potential of the data. Fortunately, advances in the application of geographic information systems (GIS) make it now possible for a unifying system to be established to control all spatial data requirements. As this paper will demonstrate, using a GIS to coordinate spatial analysis enhances the planning and analysis of these market transformation efforts. Two case studies, using the data to determine how well the existing approved contractor base can supply services to the potential households in a particular area, and estimating the income level, race and, and ethnicity of program participants.

Introduction

The Residential Contractor Program and its Data Needs

The Residential Contractor Program (RCP) now being implemented statewide in California by the four investor own utilities, Pacific Gas & Electric (PG&E), Southern California Edison (SCE), Southern California Gas (SoCalGas), and San Diego Gas and Electric (SDGE), provides an excellent case study to demonstrate the value of the GIS application. The RCP program is a program centered on contractors working in existing residential homes who supply HVAC, windows, insulation, and weatherization services. The program focuses on market interactions that occur between a variety of residential contractors and single family homeowners. Customers may obtain and utilize vouchers for payment to qualifying "eligible" contractors who provide energy efficiency services in accordance with program guidelines. It is intended that a significant number of these approved contractors will exist in the marketplace such that consumers may select their preferred provider in the marketplace. To develop this market, the utility administrators have provided training services to interested contractors, developed screening criteria for eligible contractors, and made lists of eligible contractors available to customers.

The RCP already uses geographic information in a variety of ways, though there is little coordination of these uses. Table 1 shows the variety of data used by RCP.

Table 1: Geographic Data Used by RCP Program

	Source	Issues
Utility boundary data	California's four investor owned utilities	There is overlap in gas and electric service boundaries between utilities. We used electric boundaries to demarcate boundaries.
Climatic data	California Energy Commission	CEC has divided state into 16 zones. Some of the RCP program measures are offered in only some of the climate zones.
Contractor service areas	Contractor reported	Each contractor operates in a self-determined local area. The program needs to know this territory so that it can accurately refer customers to local contractors willing to service their homes. Utilities are using zipcodes or county boundaries to mark boundaries.
Baseline Market Characteristics	Survey of licensed contractors	The market for energy efficient products is not uniformly distributed across the state. In some areas with extreme weather conditions and/or with homeowners who are more informed and have higher-incomes, the use of energy-efficient measures is more accepted. There are other areas of the state where code enforcement is lax and measures are being used with energy efficiency levels below standard levels. The program managers know that these differences in application exist, but they do not have actual or even mental maps of the boundaries.
Location of Actual Program Participants	Utility tracking databases	The utility has a need to analyze this information for a number of reasons. They need to track the distribution of the program to determine whether program funds are being fairly distributed across their service territory. In addition, consistent with the market transformation objectives of this program, they want to know how program influences existing practices, that is, are vouchers going into areas where the energy efficiency measure is already well established, or are they pushing into new market areas?

As part of the RCP evaluation, a small budget of around $40,000 was dedicated to demonstrating the development and application of GIS for both planning and evaluation purposes. The project established a GIS system that included data acquisition and/or input of utility boundary, climatic, census economic and demographic factors, and contractor and household participation data. This paper discusses the means by which all of these data were acquired and assembled in the GIS. It also discusses two specific applications of the GIS to address specific planning issues identified by the RCP program managers. Finally, the paper discusses other uses for this database, and broader applications of GIS for energy efficiency programs and program evaluation.

Definition of GIS and Description of ArcInfo/ArcView

A geographic information system (GIS) is a computer-based tool for mapping and analyzing things that exist and events that happen on earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. (ESRI 1999b).

The prominent software tools used to generate and manage GIS databases are ArcInfo and ArcView (ESRI 1999a). ArcInfo is the parent GIS software that provides the processing foundation for managing, manipulating, analyzing and displaying geographic data in map form. ArcView is the lighter desktop application that allows the user to integrate, analyze, query and view geographic information. Information can be displayed as either a raster image (e.g. satellite image or aerial photograph) or vector data (e.g. point, lines and polygons of geographic features), or a combination of both for a fully integrated GIS solution. The real strength of a GIS system is its ability to combine or overlay numerous maps with different boundary definitions and to maintain the underlying data attached to each location.

For example, in the RCP case, we are able to take one map showing number of occupied housing units by census track and overlay that with a map of the utility service territories. In many places, utility service territories intersect the census tracks. The composite map of the overlay now contains a series of whole and partial census tracks that are within each utility service territory. The computer processing recalculates the areas of the subdivided tracks and apportions the occupied housing units proportionally based on the partial tracks' relative areas. Instantly, the GIS provides an accurate measure of the number of occupied housing units within each service territory. Also provided are housing densities, and all of the other underlying data associated with each census track.

What also makes the GIS really powerful is its geocoding/address identification ability to locate from street addresses the coordinate location of any data point. By geocoding the location of participants, we can associate the appropriate census track data to that location. Similar overlays of climatic data make it possible to attach localized climatic data to each participant's location.

The GIS Process

The following discussion outlines the sources and types of data used in producing the RCP GIS. It also provides a brief overview of the methodology used to set up the GIS application. Later in the paper, we will use two examples to illustrate how the GIS is then used to address several specific applications.

Types of Data, Sources and Preparation

Using GIS has become feasible for many new applications because spatial data are now readily available, and the software is more user-friendly. Detail political boundary data, street files, and census information is available on the Internet. Many utilities have established GIS operations for transmission/distribution planning so service territory information is already in digital form. Setting up the GIS requires that all of the digitized data are correctly referenced using the same map projection assumptions, but this is handled easily within the ARCView software. The entire process of acquiring the data and setting up the GIS only took a couple of days, most of which involved processing the raw street data.

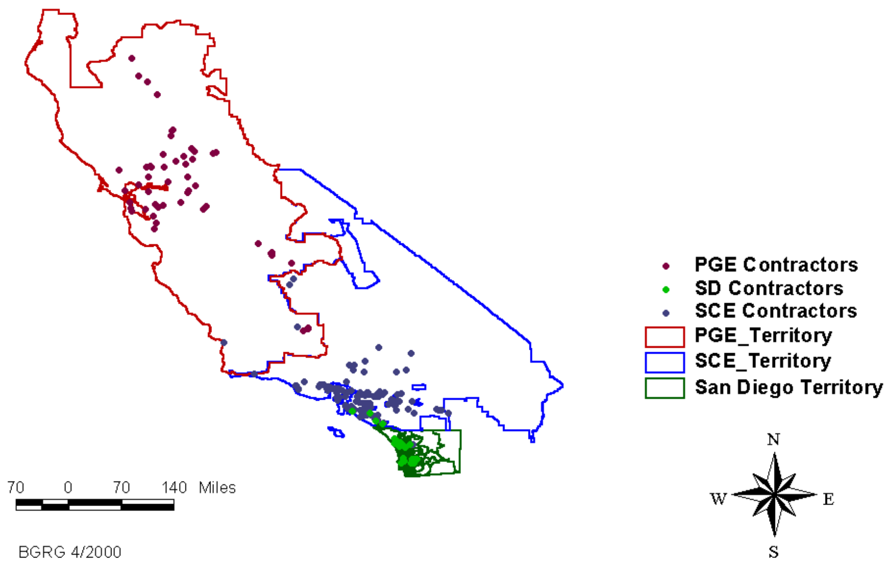
- **Boundaries**: The utility boundaries were obtained for the respective utility companies.
- **Street Segmentation**: The street data were 1996 US Census Tiger digital line graph data downloaded from the ESRI ArcData Online web site. It is 1996 US Census Tiger digital line graph data. With this information established, the GIS can pinpoint most street addresses to a specific block and side of the street. The entire GIS contains 2,343,185 separate line segments
- **Census Data**: The census data 1990 US Census Block Tiger Data downloaded from the ESRI ArcData Online web site. Census blocks are the smallest unit of aggregation of the census data and include number of households, average income, census block ethnicity, and age of housing stock. There are 5,869 census tracks in the utility territories studied in California.
- **Participant Contractor Data**: A list of contractors, who had qualified for the program was obtained from the participating utilities in the form of database files recording the contractor name, street address, type of work performed (e.g. HVAC, windows, walls, plumbing, ducts). There were 210 contractors included in the database.
- **Participant Households**. A list of participants, measured as those households that have completed the voucher form and returned it to the utility, was obtained from each utility. The exact location of each household is geocoded. Also included in this database are the measures included and the amount of incentives earned.

Case Study 1: Contractor Coverage Analysis

One pressing issue confronting planners and evaluators is the need to know what degree of coverage there is for participant contractors over the service territories. By coverage we mean how many contractors are available to service the underlying households in an area. Program managers had a rough idea of the coverage by examining the counties that contractors indicated they worked in. This gave the program managers the knowledge that some areas of their service territories were without qualified contractors. This crude assessment did not provide information on how many contractors served a particular area, nor did it provide the number of contractors as a percentage of the number of available households. There are approximately 1750 HVAC contractors serving the 10,000,000 residential single family and mobile home market in California, or approximately one contractor per 5000 single family households, (see Wirtshafter Associates, Inc., 2000a). The program needs to determine a similar coverage ratio for qualified contractors in the program. The following case study illustrate how we developed our coverage estimates.

As a first step, we prepared a map in which the location of each contractor's home office is plotted. Figure 1 illustrates the results of this exercise, which provides a visual product similar to the program managers' rough assessment. The map clearly outlines areas of the state where no contractors have their headquarters, but it does not really answer the question as to what areas lack contractor coverage.

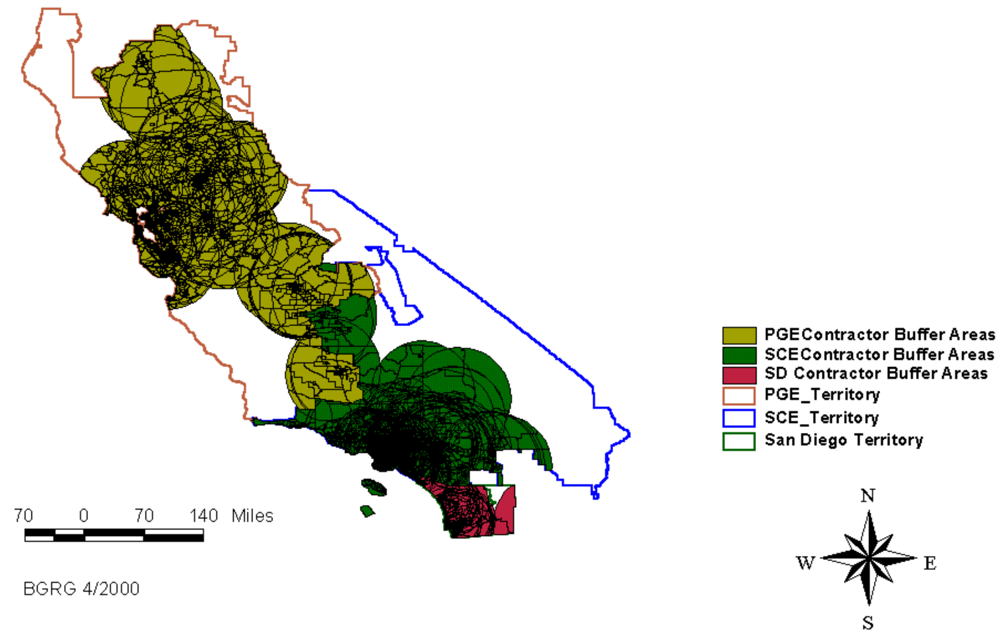
Figure 1: Location of RCP Qualified Contractors



To answer this question, we then wanted to create a map that showed the range of area that each contractor actually covered. Each contractor has a self-determined service area. This area is somewhat fluid in that it shrinks and expands based on current workload and size of prospective job. Contractors could if asked outline the range of their service area, and we see this as a valuable exercise in the future, but at the time of the RCP study no such information was available. We knew from a baseline survey of contractors that on average they define their business range as 60 miles. This figure was clearly too high for urban areas, but our baseline data did not have enough data points to disaggregate the results to specific urban areas. We finally settled on using a sixty-mile buffer for rural/suburban areas and thirty miles for San Francisco, Oakland, Los Angeles, San Diego, and Orange Counties. One feature of the ARCView software is the ability to create a buffer area around a point. Figure 2 shows the Contractor Range created by using the 60-mile buffer except in the urban areas where a 30-mile buffer is used.

As Figure 2 indicates, there are still large areas of the state without any contractor coverage. While Figure 2 can effectively illustrate areas with no coverage, it cannot effectively quantify how much coverage is available, and more importantly how does the coverage that is there compare to the number of households that potentially could be participants. We create the Contractor Coverage Potential to answer this last query.

Figure 2: Areas of State Covered by Using 35-Mile Urban and 60-Mile Rural Travel Limits for Contractors



As we define the real issue, the objective of the analysis is to determine what the probability is that a contractor could service a household. The probability of a specific contractor servicing a household is shown in Equation 1.

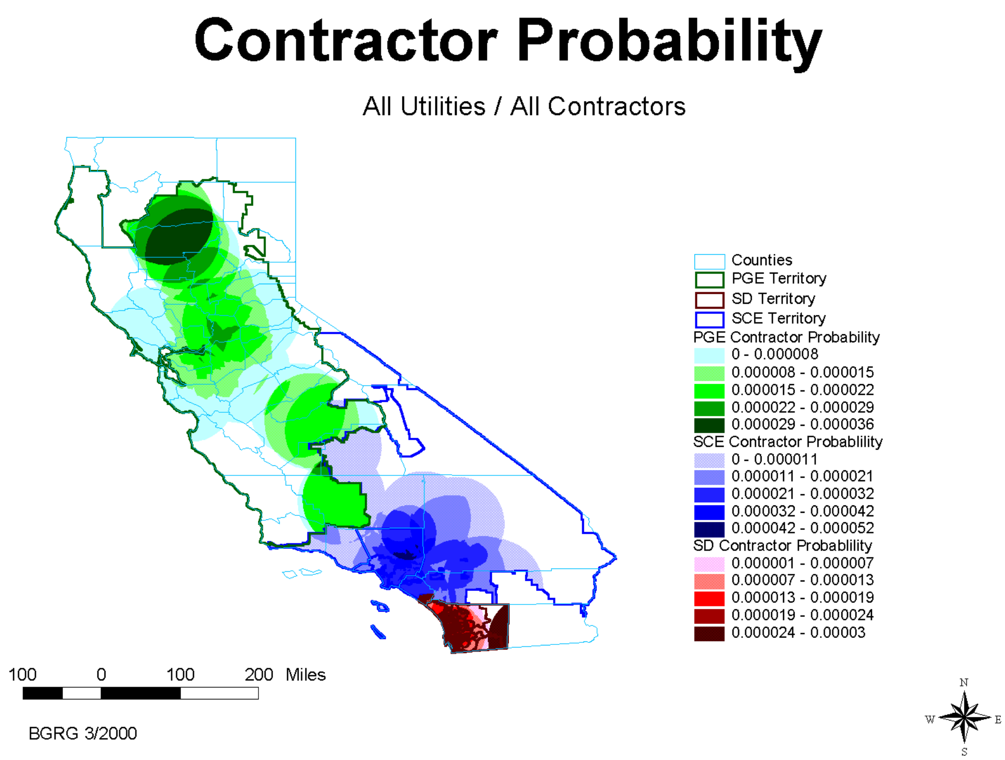
$$\text{Contractor Service Probability by specific contractor} = \frac{\text{no. crews}}{\text{no. of households in contractor range}}. \qquad \text{[Equation 1]}$$

The probability that a household will be serviced by all of the available contractors is the summation of the individual contractor probabilities for those contractors whose service territory overlaps a particular location, as shown in Equation 2.

$$\text{Overall Contractor Coverage Probability} = \sum \text{Contractor Service Probability (Equation 1)} \qquad \text{[Equation 2]}$$

To calculate these probabilities involves generating a dataset with topological structure so that each area within a utility region correctly accounted for probability of households being serviced by contractors within each common intersection of contractor service areas. Effectively, it requires that the database identify all of the unique combinations created when each circle in Figure 2 is overlaid unto all of the other circles. The resulting map of Contractor Coverage Probability is shown in Figure 3.

Figure 3: Contractor Coverage Potential



This is an intensive computational task and it was necessary to transfer the data to a mainframe system using ArcInfo software. Even with this transfer, the overlays had to be done one at a time. Each contractor range was added and the resulting composite map was created. Then the next contractor range was overlaid to the composite map. Each time the composite was added to, the program recalculated the areas of the intersecting polygons and recalculated the probabilities. A master program was written to optimize the execution of these programs.

The resulting geo-spatial data is a new geometry displayed as a complex set of new polygons. These new polygons are formed from contractor service area buffer overlaps, each overlap consisting of a unique combination of intersections. The probability of a household in the unique buffer overlaps being serviced by a contractor is a function of the number of households each contractor has available to service within their buffer area and the number of contractor service areas overlapping.

Still as Figure 3 illustrates, using 1999 program participation data of qualified contractors, most of the state has in fact very low probabilities of contractor coverage. There are no areas with probabilities of 1 qualified contractor per 25,000 households even when including all license types of contractors. The coverage probabilities for the individual license types are generally less than 1 qualified contractor per 100,000 households. There are vast areas of the state that have essentially no contractors serving their areas. Keep in mind that a contractor coverage probability of less than 1/25,000 means it will take at least 25 years to service all of the homes, even if one assumes that each contractors can do 5 homes a day, 200 days per year. To achieve market transformation will require significant increases in the number of participating contractors as the program progresses in maturity.

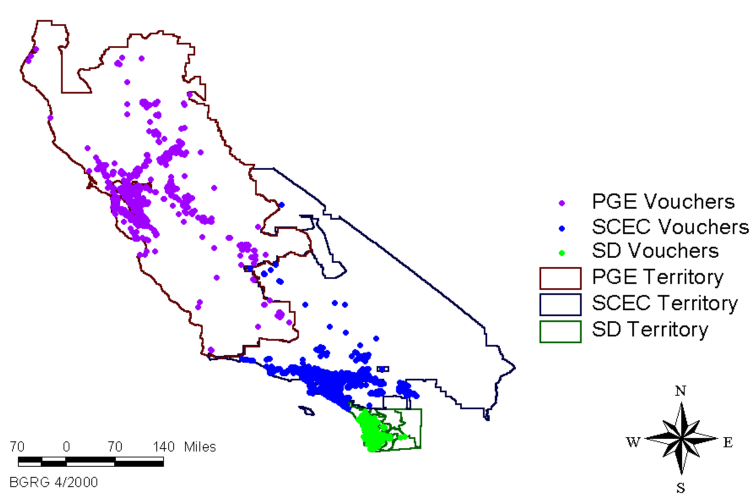
The results also indicate a concentration of contractors around the specified training sites of the utilities. The data shows that the program needs to expand its training/program approval workshops out into other areas of their service territories.

Case Study Two: Voucher Distribution

A second issue of concern to the program managers is the question of which residential customers are participating in the program. Many DSM programs have been criticized because they tend to attract mostly better educated and wealthy households. Program managers are thus sensitive to this issue and need intelligence regarding who is participating. Unfortunately, the data necessary to make such an assessment is not typically available to program managers until after evaluations are completed. Utilities do not generally know the income or ethnicity of their customers unless they have entered into a payment arrangement, so the program managers generally must wait for survey results to determine program participant characteristics.

Because the GIS can pinpoint the exact location of each home that participates in RCP (participation is marked by the return to the utility of a completed incentive voucher), the opportunity exists to superimpose the participant data onto the underlying census data. We do this by summing incentive dollars by census tract and then using standard database queries to relate incentive dollars to average income, and ethnicity. The exact location of the each participant household is shown in Figure 4.

Figure 4: 1999 Voucher Locations



The voucher data can be overlaid with census data to determine the types of households that are participating. This gives planners and evaluators an approach for determining ethnicity or income level that is more reliable than post-facto survey responses. In this analysis we have aggregated the census blocks into four quartiles, based on the median income of the census block. Table 2 gives the results of the data aggregation.

Table 2: The Distribution of Program Benefits by Census Track Median Income Quartile

	Census Track Income Quartile	Median Income of Entire Quartile	Number of Vouchers	Total Incentive Dollars	Vouchers/ Household
PG&E	Lowest	$25,463	365	$63,015	0.00356%
	2nd	$34,212	251	$51,218	0.00211%
	3rd	$42,023	315	$61,261	0.00262%
	Highest	$53,914	246	$42,127	0.00223%
SCE/SoCalGas					
	Lowest	$26,050	1073	$161,560	0.00907%
	2nd	$35,652	530	$85,022	0.00382%
	3rd	$43,383	505	$95,920	0.00393%
	Highest	$54,218	572	$119,516	0.00460%
SDG&E					
	Lowest	$26,646	36	$5038	0.0145%
	2nd	$35,888	98	$14,855	0.0371%
	3rd	$43,484	173	$30,034	0.0704%
	Highest	$54,613	173	$25,648	0.0707%

The results of Table 2 show that for SCE/SoCalGas there is marked skewness towards support of the lowest income quartile. This is largely the impact of the single largest contractor, who has done over half of SCE/SoCalGas's jobs, and targets the mobile home community. This emphasis on the mobile home market puts the RCP program at least for SCE/SoCalGas in a unique position of having their voucher distribution favoring the lower income areas. SDG&E which has focused mostly on air-conditioning services has a perceptible bias towards areas with higher median incomes.

The GIS gives an opportunity for further examining issues of ethnicity. In Table 3, the distribution of census tracts by race is examined. The census tracts within each income quartile have been divided based on whether or not a voucher was redeemed by any households in that census tract. Using 1990 census data for percentage of population that is white, the race distribution of participant homes can be generally compared to the race distribution of non-participant areas. The results demonstrate that census tracts with higher percentages of white persons are more likely to have had participants in RCP. This is particularly true of the lower income census-tract quartiles. This indicates that RCP's positive record in reaching lower-income households is not as positive in reaching non-white households.

Table 3: Race Distribution Differentiated by Census Tracks with and without Program Participation

	Percent of Entire Quartile Population that Is White	Percent of Population that is White for those Census Blocks in Quartile that Had a Participant	Percent of Population that is White for those Census Blocks in Quartile that Did Not Have a Participant
PG&E			
Lowest	61%	65%	60%
2nd	74%	77%	73%
3rd	76%	80%	75%
Highest	79%	82%	79%
SCE/SoCalGas			
Lowest	45%	69%	41%
2nd	63%	70%	60%
3rd	73%	78%	71%
Highest	81%	82%	80%
SDG&E			
Lowest	61%	66%	59%
2nd	77%	81%	73%
3rd	81%	84%	75%
Highest	88%	87%	89%

Conclusions

The results of the two case studies indicate the power of the GIS in several ways. The creation of maps that locate actual contractor and participant household locations provide a much clearer picture of areas of program activity. These maps illustrate quite clearly that large areas of the state are not participating in the program. The GIS even assists the program managers in understanding why these areas are not participating.

All three utility-run programs relied on one training facility in each service territory. As it turns out, nearly all of the participating contractors are located within two hours drive of one of these facilities. This should not be surprising because contractors have indicated that they are only willing to travel on average sixty miles for a paying job, so it is not unexpected that few will travel more than two hours for a training session.

An interesting exercise that could be performed using the GIS would be find out what the travel tolerances are for contractors going to training courses. To accomplish this, we would measure the actual distances each contractor traveled to get to the training session. We would also establish the travel distances to the training sight of all licensed contractors. Then we could plot the number of those that need travel versus the total number that could have potentially traveled that distance to attend. This will produce a distance decay ratio that could help determine the number of training sites need and their optimal location.

The GIS system also demonstrated real utility in determining the ethnic, racial, and income characteristics of participants. Most programs do not have the ability to measure these factors unless they engage in post-program customer surveys. These surveys only represent a portion of the participant population and may be biased due to non-response issues. The GIS is able to locate accurately the census tract of each participant. By aggregating responses by census tract and then using the known characteristics of each tract, we are able to provide valuable statistics regarding the types of households that have participated. This information can be provided more quickly so that program managers can make program modifications. The GIS can also help in providing detailed information for targeted marketing to underserved areas.

The two case studies only begin to describe the full potential of GIS as a planning and evaluation tool. Now that the GIS is in place, the program managers can use the system to explore numerous other activities. We discuss five more below.

Further Discussion of Other Applications

The addition of climatic data allows the program managers to create their own climatic zones or to analyze the existing participation using the 16 California climatic zones. One particularly pressing issue that is needed is to analyze the cost-effectiveness of the various technologies that are part of the program. As it stands now, neither the program managers nor the contractors know what the paybacks are for the measures that are in the program. The GIS system could establish a payback map for each measure based on the underlying climatic conditions and simple engineering assumptions.

An enhancement to the payback map would be to have a map of overall measure potential. This assessment would require the development of a housing status survey that indicates what measures have already been adopted and what measures still could be potentially installed. The GIS could use these data to create a map showing the local potential for each measure across the state.

A part of the RCP program includes a contractor referral service. Potential participants call the hotline or check an Internet site to find qualified households. Names of local contractors are given to the caller. Right now, the definition of local is at the zipcode level or county level. If each participating contractor drew an outline of their service territory, the caller could be given a more accurate list of contractors who would actually be willing to serve that caller.

The GIS can also be used to implement targeted marketing. Program managers can send information to customers in specific areas where participation is low or where program potential is high. Using the GIS to effect this target marketing will better pinpoint the target audiences.

Perhaps the most important future opportunity for GIS lies in its ability to track market effects. The GIS if used over time can track the development of the market for the program measures. The introduction of new contractors into areas not previously served, and the initiation of participants in new market areas constitute the types of market effects that RCP as a market transformation program are designed to encourage. The GIS can be used to identify these effects both on a strictly spatial basis, and also using the underlying demographic data, or to measure potential data described as the second enhancement above.

This tool would be a vast improvement over today's approach, which essentially treats all potential households as homogeneous.

References

ESRI, 1999a, *ArcView GIS, Version 3.2, 1992-1999,* Environmental Systems Research Institute, Inc., Redlands, CA.

ESRI, 1999b, *ArcView GIS Demo Edition, April 1999*, Environmental Systems Research Institute, Inc., Redlands, CA.

Wirtshafter Associates, Inc., 2000a, *Report of the Residential Contractor Program Evaluation Volume I: Phase I Residential Contractor Program Market Assessment*, Prepared for California Board for Energy Efficiency (CBEE) and Pacific Gas & Electric (PG&E), PG&E Study ID #424C.

Wirtshafter Associates, Inc., 2000b, *PY99 Residential Contractor Program Market Assessment and Evaluation Study: Volume 1: Summary Report*, Prepared for Southern California Edison.