"Watts" Where, and Why? Using GIS to Identify Energy Efficiency Opportunities

Richard Crowley, DNV GL
Whitney Brougher, National Grid

ABSTRACT

Geographic visualization and analysis of data currently collected by utilities and energy efficiency program administrators is a powerful and underutilized tool for identifying energy efficiency opportunities. Energy efficiency program tracking and energy consumption billing data capture geographic information ranging from utility service territory down to individual meter locations. However when these datasets are analyzed, more often than not the spatial information is either dropped from the analysis or treated as a categorical bin for organizing and displaying tabular data. The spatial element of the utility data is either ignored, or presented in a way that obscures valuable insight into the trends and geographic differences within utility customers. Program administrators may be missing out on geographically discreet opportunities to better serve hard to reach efficiency markets.

This paper presents the preliminary results of how GIS was used by the Massachusetts Program Administrators to develop an enhanced understanding of their energy efficiency data by spatially evaluating program penetration, participation trends, energy savings, and customer evolution across years. These outputs provide the utilities with a new look at the geographic evolution and dynamics within their efficiency offerings, and can also enhance existing analyses by presenting a more comprehensive picture of the efficiency landscape than a sole table or chart can offer. Through the application of geospatial analysis software, utilities can potentially improve their targeted offers to customers, increase the adoption of energy efficient technologies, and more effectively evaluate what is happening, where it is happening, and why it is happening.

Introduction

Geographic information systems have experienced growing utilization in the energy sector around asset management and identification, grid analytics, and renewable energy systems (ESRI, 2007; ESRI 2010; Simmins et al. 2012; Lopez et al. 2012). Deployment of GIS for energy efficiency is comparatively less utilized, and existing work has largely been focused on applying spatial analytics for identifying residential energy efficiency opportunities due in part to low and no cost robust national datasets with demographic and building profile characteristics including the US Census Bureau's American Community Survey, state level tax databases, and the Energy Information Administration's Residential Energy Consumption Survey (Kelsven 2012; Min, Hausfather, and Lin 2010, Crowley and Quinlan 2011). Publically available robust commercial and industrial energy efficiency data is comparatively harder to access and often available only at coarse geographic levels that may not be sufficient for the locally discreet geographic customer variability within the commercial and industrial landscape – a less homogenous population than the residential built environment. Utilities and energy efficiency program administrators are in a prime position to leverage their internal data and geographic information systems to understand how geographically discreet regions and locations consume energy and participate in energy efficiency program. This report uses data provided by the Massachusetts Program Administrators (the PAs) from their electric efficiency program tracking and billing databases and the GIS software packages ArcGIS and QuantumGIS to demonstrate how DNV GL and the PAs are leveraging GIS analysis to help the PAs better evaluate their commercial and industrial energy customers.

The first step of analyzing the PA data is to actually get it into a working spatial format, meaning tabular data must be brought into the geographic space. The PA customer's site address information was imported into an ArcGIS geodatabase and geocoded using a composite address locator built by DNV GL as shown in Figure 1; derived from several different spatial reference files to provide a high level quality of matched locations[1]. This tool attempts to match PA provided addresses starting at the address level and terminating at the zip code level at which point the location is declared unmatched.

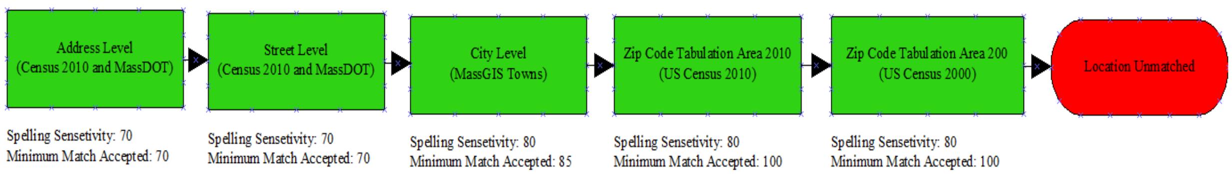

Figure 1. DNV GL composite address locator process flow and minimum acceptance criteria.

We chose to use a composite locator as it allows each individual locator element to keep a high acceptance score while also ensuring that a large number of project points are coded to the appropriate location on the map for use in analysis. An additional benefit of the custom composite locator is that the same tool can be used for time series and cross project data to ensure consistency in how points are placed for location based analysis.

In this paper, several different sets of data that where collected at different points in time have been utilized to provide insight into utility data including the 2011 Massachusetts commercial and industrial billing database (Laccetti and Kattel, 2012), the preliminary 2012 Massachusetts commercial and industrial billing and tracking databases (Farland, Kattel, and Crowley. 2013). One of the major benefits of geocoded point data and a true GIS engine over a data visualization dashboard software is the ease of changing the spatial filter based on the question asked – in effect as long as the question can be phrased as "tell me about item X, broken down by Y" where Y is the spatial boundary file, then the data can quickly be reprocessed while maintaining discreet, point level, granularity representing real world conditions. For example, we could have chosen to look at the zip code level, at how consumption differs in rural vs. urban regions or even by a sub region of a PA service territory – so long as the spatial boundary exists and there is point level data, then changes in the study region can easily be handled and remain highly informative of the real world conditions. It is important to note at this point in the process the results are only as accurate as the combination of the data provided, the base layers, and the composite locator acceptance thresholds. It is possible to have matched records that may not be at the correct location. There are several examples included in figure 2 where matched locations appear to be in municipal electric territory; this can be a result of mailing addresses in the raw

[1] GIS base files used to construct the composite locator were collected from the MassGIS Datalayers file listing (http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/layerlist.html), and from the US Census Bureau's TIGER/Line® Shapefiles FTP site (http://www.census.gov/geo/maps-data/data/tiger-line.html).

data being coded as site locations, false positive matches, or even correct matches but a misleading base layer with regard to town boundaries used as a proxy for PA territory. With datasets approaching over half a million records it may not be cost effective to correct these false positives, particularly when they are not obvious ones, as they represent a fractional percentage of the total. One potential option is to reject these matches altogether and leave the resulting locations as unmatched based on the raw data provided; a second is to perform spot check remediation on known outliers. DNV GL used the later of the two approaches in this study.

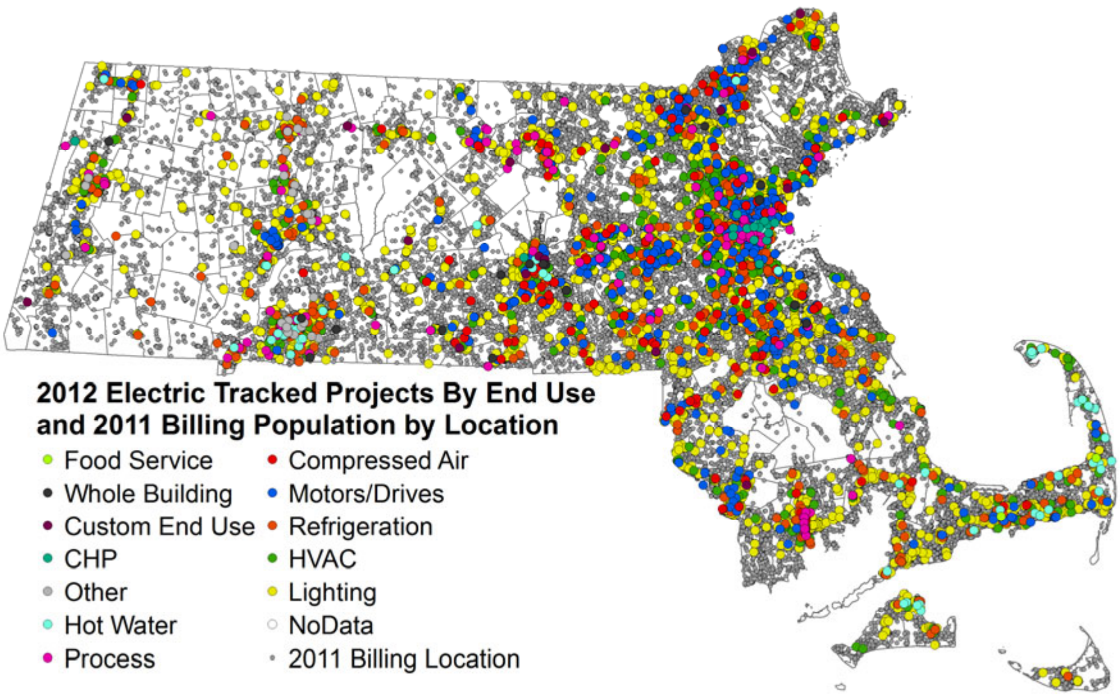

Figure 2. Preliminary 2012 electric commercial and industrial tracked energy efficiency projects by end use, overlaid on 2011 commercial and industrial billing population.

Dot maps are a useful first view of the utility data. With this single map, we are looking at how nearly 400,000 electric commercial and industrial customers and 12,000 commercial and industrial electric energy efficiency projects were distributed across Massachusetts. We can see that as one might anticipate the projects are concentrated around the Boston area and that in the sparser populated western regions of the state we have few projects corresponding to a smaller customer base. While this result intuitively obvious, it is essential to perform these checks on the point level data. In these checks we are expecting to find no surprises but rather confirm that the base layer that will feed into the geographic analyses does not contain any unexpected data considerations. The point maps can also be informative in helping identify potential questions. On close visual inspection we can also see that lighting measures are geographically dispersed, but electric hot water measures appear to show some spatial clustering. We will visit this more in depth later in the paper. One important point to note about this type of map as a first look at the data, and in fact of any GIS map in general, is that the visual output can be highly influenced by decisions made in the map process. For example, in this map the dots are stacked on each other by a hierarchy reading left and down in the legend. This can make it appear as though there were no lighting projects occurring in Boston though in actuality Boston has the highest count and density of lighting projects anywhere in the state, but the sheer volume of overlapping project points in the Boston region visually masks this.

The question becomes then how can we take this wealth of data that we have created and use it to provide valuable insight. Given that we are working with commercial and industrial efficiency data a good starting point was to understand where are the highest raw regions of energy savings to see if anything surprising could be found. Figure 3 is the preliminary snapshot of where raw reported commercial and industrial energy savings were highest. Not surprisingly, the raw amounts of energy saved correspond closely with the regions of high commercial and industrial population densities.

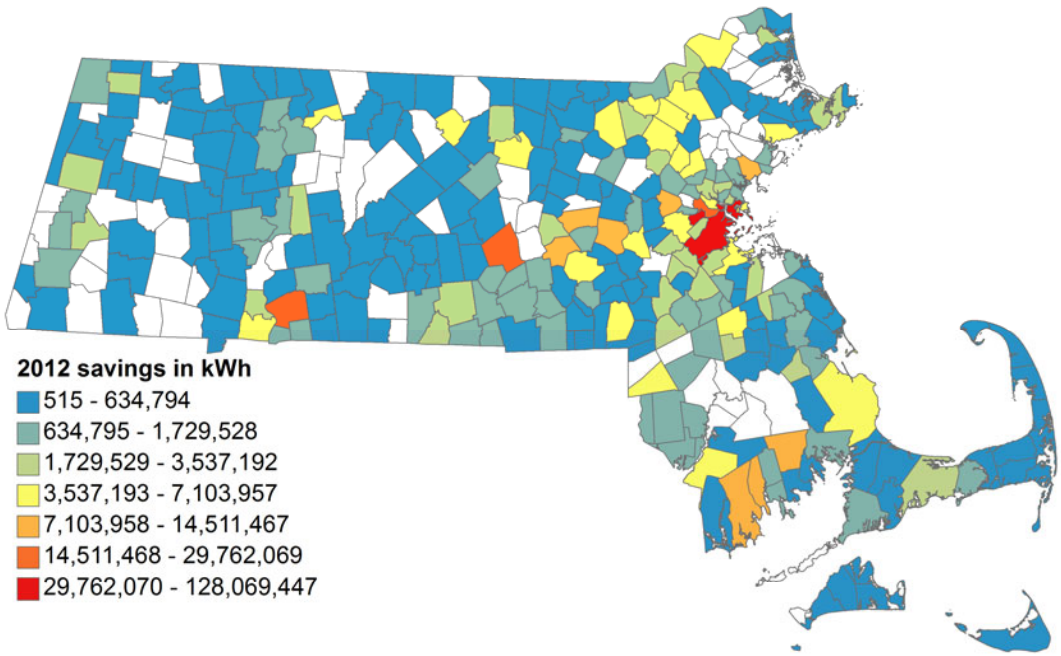

Figure 3. 2012 kWh savings broken down by the natural break classification scheme[2].

This is an interesting view of the data, but perhaps not the new insight we are looking for. The sum savings, not surprisingly, seem correlate pretty strongly with population; and so perhaps understanding savings as a percent of town level energy consumption could be more insightful and minimize population effects. Areas of lower population will have lower savings, but also lower consumption. Figure 4 presents the preliminary savings normalized by town consumption; diagonal lines indicate savings in towns that are primarily served by municipal electric providers.

[2] Natural breaks (Jenks) are one of the packaged classification schemes provide in the ArcGIS software. This scheme seeks to maximize changes between classes while minimizing variance within the classes. For more information please see http://resources.arcgis.com/en/help/main/10.1/index.html#//00s50000001r000000

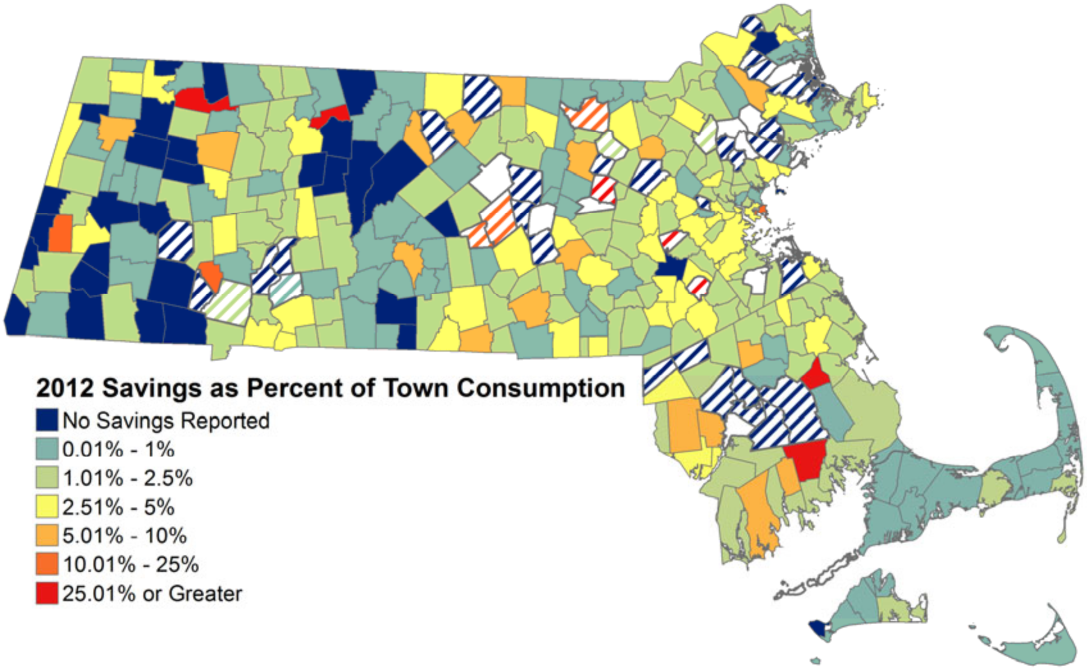

Figure 4. Preliminary 2012 kWh town level savings normalized by town consumption.

With this view, we are presented with a better understanding of which towns, despite low total savings, actually experienced a large percent savings relative to the town consumption. In the northwestern part of the state we see two towns in particular that have preliminary 2012 savings in excess of the 25% of the town level 2011 energy consumption. This insight can be used to select the sites in these towns to better understand what type of business they are, the measure that was installed, and if or how this measure might relate to other areas in the state.

What if we would like to know which towns preformed above or below the mean savings? The map in figure 4 is useful for allowing us to evaluate normalized savings for one town against another, but it can be difficult at face value to understand how these towns relate to one another in the larger state picture. In figure 5 we have removed the towns with no identified savings and for the remaining towns taken the sum of the savings of all projects in each town and normalized it by the sum of the 2011 billing data[3]. The data was also transformed to generate a normal bell curve so that we could look at towns that deviated substantially from the mean energy participation ratio. This view allows us to better understand which areas of the state achieved high savings penetration relative to the state mean. The resulting view tells a much more interesting story. We can see using this metric several of the western towns performed substantially well in terms of energy efficiency, and by referring back to Figure 1 we can also get a feel for what end uses these projects addressed.

[3] 2011 is used as it represents the last full-year cycle before the tracked measures were installed, and the energy savings in the tracked measure database are reported at the annual savings level.

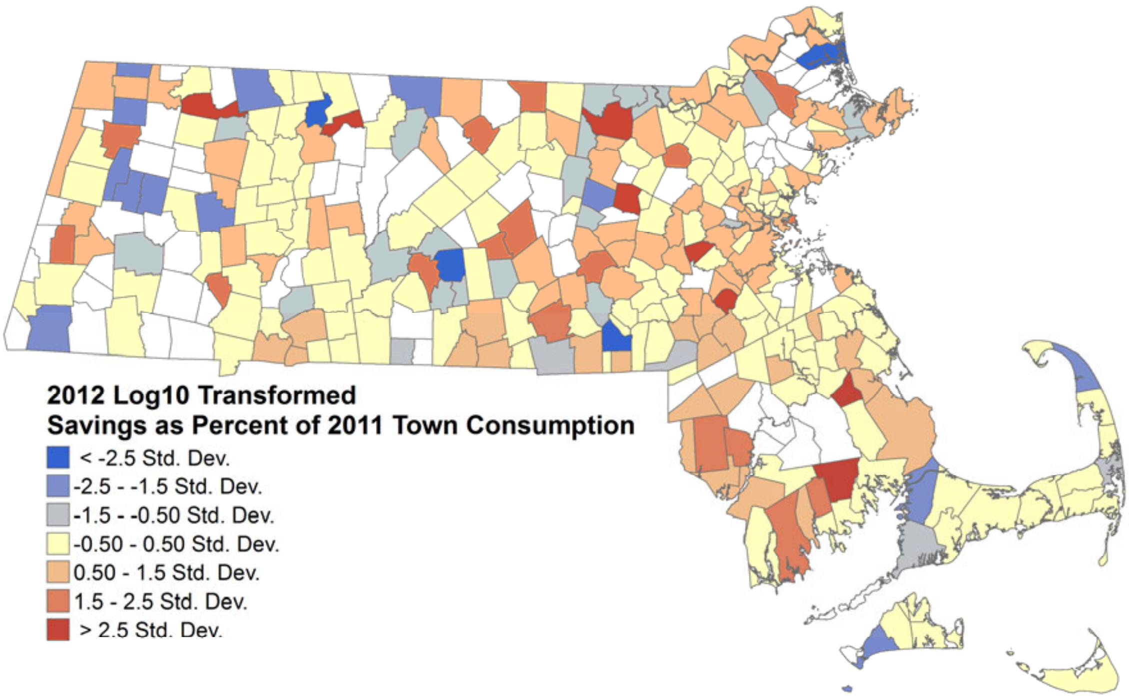

Figure 5. Log base 10 transformed total town 2012 energy savings normalized by 2011 town consumption.

Thus far, the focus has been on savings – however this is only part of the picture. A single large customer can substantially change year over year savings ratios, particularly for smaller PAs or towns with generally smaller commercial and industrial facilities. Savings viewed in isolation may provide a misleading picture and the more appropriate question may therefore be what types of participation rates do we see and how do these compare to other regions in the state, and the state as a whole. Figure 6 illustrates the 2012 town level participation rates. In a 2012 report, DNV GL found that statewide the participation rate was 2.7% of the billing population (Laccetti and Kattel, 2012) but at the more granular town level we can see large fluctuations in regional participation. Figure 7 presents the participation rates using the same log transformation described for figure 5 and highlights towns with substantially higher or lower participation versus the 2012 state mean. Identifying participation rates, and in particular identifying areas absent participation, can in and of itself yield enough insight for PA to capitalize on institutional knowledge of their customer base to target efficiency offerings or to direct programmatic staff to focus outreach efforts on low participating regions.

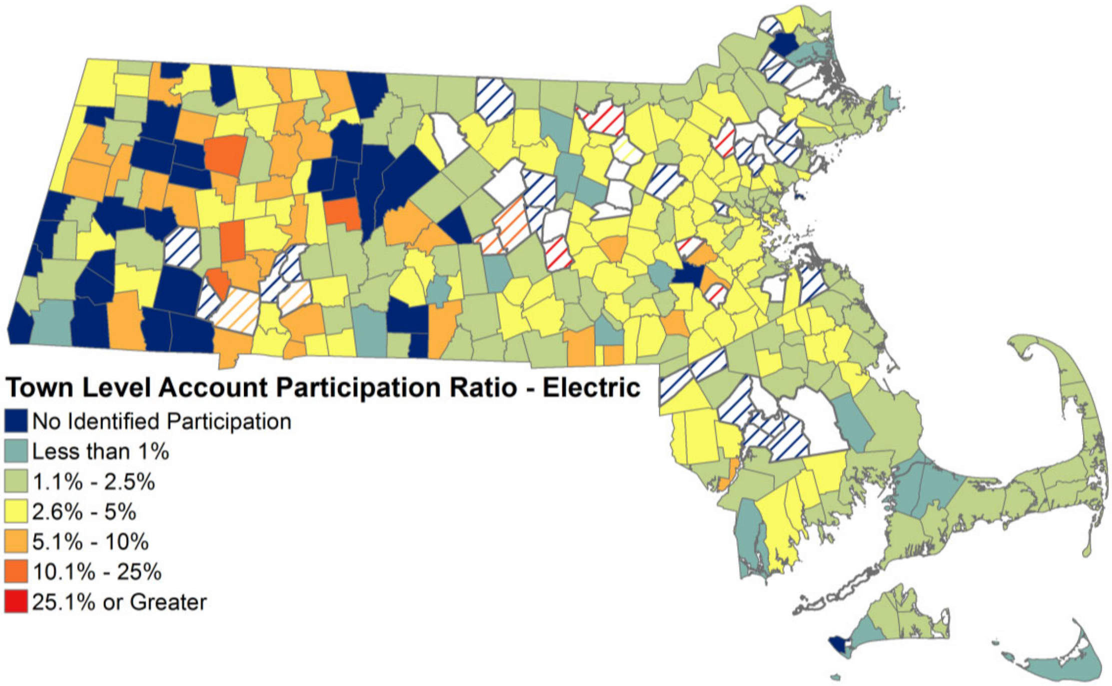

Figure 6. 2012 project participation rates versus 2012 population.

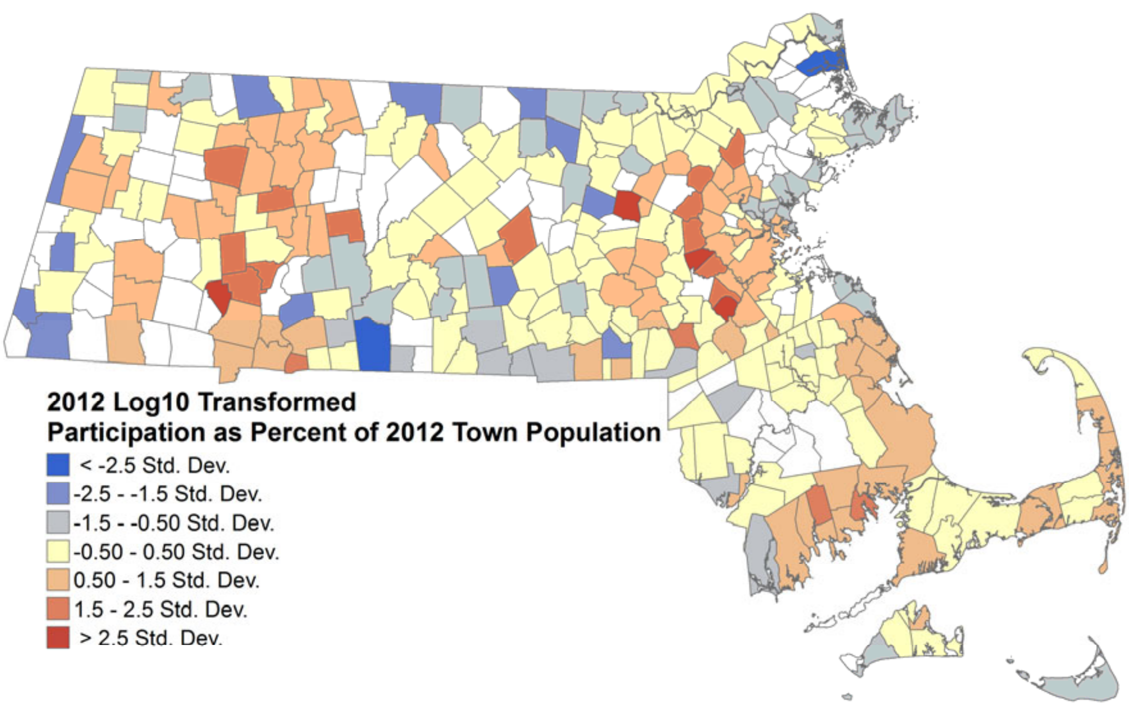

Figure 7. Log base 10 transformed total town 2012 participation normalized by 2012 town population.

These aggregate summary maps presented thus far can help the PAs to focus their outreach and efficiency offerings to specific regions, and serve the added benefit of leveraging the point level spatial data in such a way that billing and consumption data can be presented without compromising specific customer details. Taking the analysis in a different direction, we can also look at the point level data and how it relates to other points on the map. One of the key elements of generating useful geographic visualization is to select attributes that offer meaningful insight and actionable business intelligence for the client. The program tracking data includes details on the project end use (e.g. lighting, hot water), the building type (e.g. office, lodging) the program details (e.g. custom retrofit) and the energy impact of the measure installed – all of which can be leveraged to gain insight into how projects are being implemented across the state.

As an example, a PA may be interested in increasing the participation of buildings that serve primarily as lodging. At the highest level, the PA may already know how many customers they have that qualify as a lodging building type. However, if these establishments are clustered there may be added value to assigning a resource to outreach to this cluster. In the case of lodging there is a reasonably large spread across the state, but with apparent higher numbers appearing around Boston as shown in figure 8– but where are the heavy densities of lodging located at so that the PA can focus efforts on those clusters?

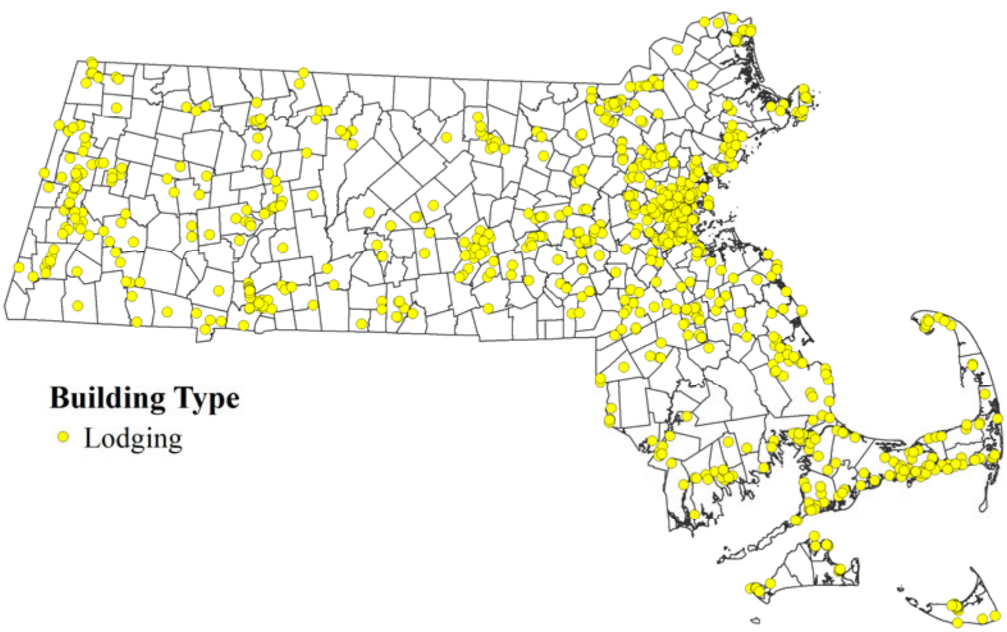

Figure 8. Location of buildings classified as lodging, derived from DNV GL Massachusetts existing building characteristics database (Leahy, 2013).

To evaluate clustering DNV GL used a 1 mile search radius with 1/5th mile raster cells and a 0.1 decay rate from each lodging instance to generate a density map of lodging establishments within a mile of each unique location. Break values were set to identify areas of below average concentrations, deviation around industry mean, and areas where the concentration of buildings exhibited extreme deviation from the mean and was indicative of large, clustered population's relative to the business type statewide. Figure 9 contains two clusters that stand out as greatly exceeding the mean in Boston and Nantucket (the island in the extreme southeast) – to be clear this does not mean that the same raw numbers of facilities are located in these areas; just that per the population file used, the densities in both areas exceeded the mean by over 10 standard deviations. A deeper dive into the tracking data indicated that none of the participants on Nantucket from the 2012 tracked year were classified as lodging facilities. This would indicate that, on the surface at least, there is a potential cluster for targeting efficiency offerings.

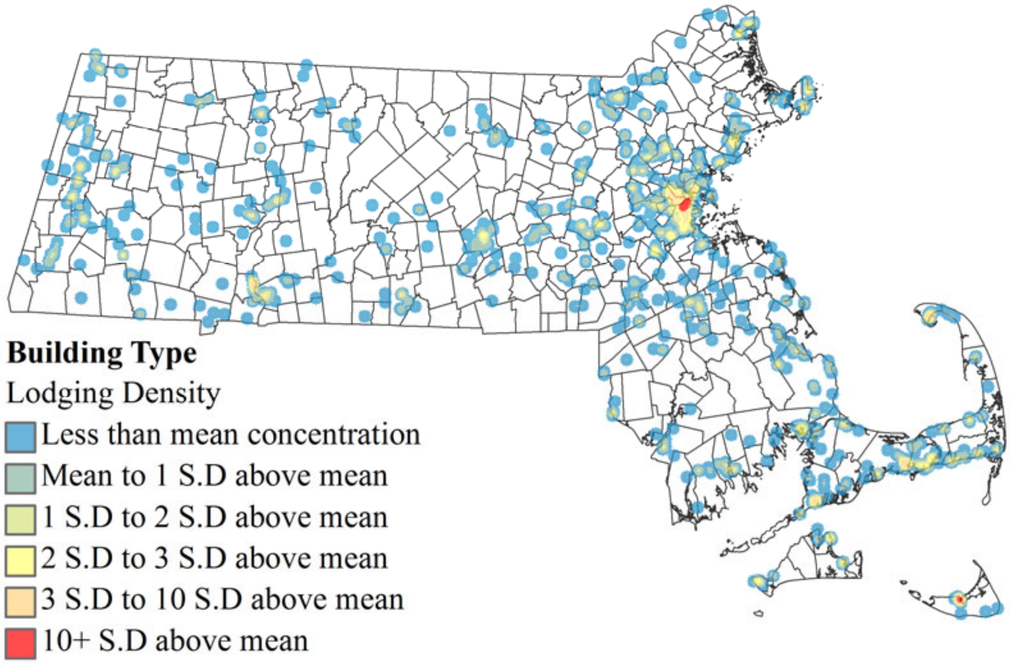

Figure 9. Lodging facilities by relative density as classified by standard deviation above statewide mean.

The next step in this assessment is to try and understand what types of offerings are most likely to be relevant to the target population – what end use would be good to focus on if we wanted to target the hotels on Nantucket. With tabular data, we can look at what types of measures lodging facilities are most likely to install, and suggest that the PAs focus their efforts on those. However, with the GIS capabilities we can also look at more crosscutting end uses and ask what types of end uses commercial and industrial customers are generally focusing on in the Cape Cod region, and are there any that appear to be more unique that would merit consideration for a targeted program offering. As an example DNV GL evaluated a series of project end uses to determine if there were any potential trends. Electric hot water measures stood out as particularly interesting candidate for consideration as an offering to Nantucket given electric hot water measure for commercial and industrial customers do not have a high geographic distribution across the state, and the Cape Cod region was one of the more populated one for this electric end use. Figure 10 shows the project density for electric hot water measures at a 1 mile search radius with 1/5th mile raster cells and a 0.1 decay rate in the Cape Cod region. To further emphasize areas of extremely high concentration by project counts (rather than just deviations from the mean) DNV GL has utilized a hillshade[4] built of the project density counts.

[4] A hillshade is a continuous surface that uses a z-value and an illumination from a specified point on the compass to generate a shaded relief layer that can be used to enhance the visualization of the z-values. For technical details please see: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/How_Hillshade_works/009z000000z2000000/

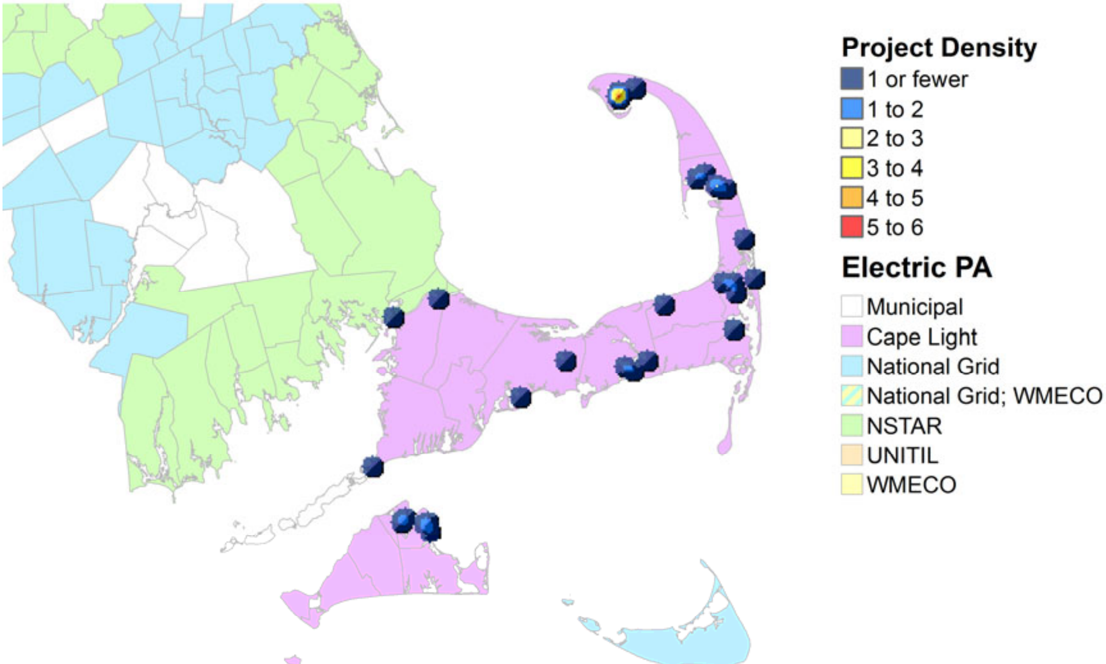

Figure 10. Electric hot water energy efficiency project density from 2012 tracked projects.

There is a particularly intense cluster identified in Provincetown. This particular cluster, and more generally the clusters in the cape region, likely correspond in part to there being no natural gas service provider in several towns, and so commercial and industrial facilities are likely to use electricity for their water heating needs. With this context, the project counts make intuitive sense and can help the electric provider to potentially focus efforts on customers in this area that were not part of the clusters but may be interested in high efficiency water heaters – including that cluster of lodging facilities from figure 9 that did not show up in the 2012 tracking data for any measures at all.

Conclusion

In this paper we have presented several different and accessible ways in which DNV GL is using GIS software to help the Massachusetts PA's gain insight into their efficiency programs. Though the utilization of point level data it is possible for the PAs to dissect their market landscapes though a variety of spatial filters, including town level and proximity to other points as was demonstrated in this paper, while still preserving individual customer confidentiality. By leveraging the geographic component of commercial and industrial billing and tracking data, PAs can generate more information from that data they have already invested in collecting and analyzing and can gain a greater insight into what is occurring with different regions of their service territory. This knowledge can translate into an enhanced geographic understanding of how large sets of billing and tracking data concisely fit together, a deeper insight into which regions are over or under performing, and even a starting point for considering what types of end uses and customers are prime targets for new efficiency offerings and can assist the Massachusetts PAs to continue offering the targeted innovative programs and measures that have made Massachusetts the most energy efficiency state in the United States in 2013.

Acknowledgements

DNV GL would like to thank the Massachusetts PA's for providing the raw data used throughout this report and for their feedback throughout the analysis process. We would also like to thank the anonymous reviewers for their feedback in improving earlier drafts of the paper.

References

Crowley, R., Quinlan, P. 2011. *2011 North Carolina Clean Energy Data Book.* http://energync.org/assets/files/NCSEA%20_2011_CEDB_COMPRESSED_FOR_WEB_low_res.pdf

ESRI (Environmental Systems Research Institute) 2007. *GIS Enhances Electric Utility Customer Care.* http://www.esri.com/library/whitepapers/pdfs/gis-enhances-electric.pdf

ESRI (Environmental Systems Research Institute) 2010. *GIS Best Practices – GIS for Renewable Energy.* http://www.esri.com/library/bestpractices/renewable-energy.pdf

Farland, J, Kattel U, Crowley, R. "2013 MA-CIEC Enhanced Customer Profile Project". Confidential client database. SAS Binary Database.

Kelsven, P. 2012. *Leveraging Data Mining and Geographic Information Systems to Gain Energy Efficiency Market Intelligence.* http://www.aceee.org/files/proceedings/2012/data/papers/0193-000377.pdf

Laccetti, R., Kattel, U. 2012. "2012 MA-LCIEC Customer Profile Project". Confidential client database. SAS Binary Database.

Lopez, A., Roberts, B., Heimiller, D., Blair, N., and Porro, G. 2012 *U.S. Renewable Energy Technical Potentials: A GIS-Based Analysis.* http://www.nrel.gov/docs/fy12osti/51946.pdf

Min, J., Hausfather, Z., and Lin, Q. 2010. "A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States" *Journal of Industrial Ecology 14 (5): 791-807.*

Simmins, J., Lyon, B., Sarfi, R., Tao, M. 2012. *Monetizing the Geospatial Information System (GIS): The Value of GIS Data Quality for Electric Utilities. An EPRI Smart Grid Assessment Report.* http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=000000000001024303