Exploring Urban Activity Patterns Using Electric Smart Meter Data

by

Saul Kriger Wilson

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master of Science in Urban Studies and Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Author . . . .
Department of Urban Studies and Planning
August 26, 2016

Certified by . . . .
Carlo Ratti
Professor of the Practice
Thesis Supervisor

Accepted by . . . .
Associate Professor, P. Christopher Zegras
Chairman, MCP Committee

Exploring Urban Activity Patterns Using Electric Smart Meter Data

by

Saul Kriger Wilson

Submitted to the Department of Urban Studies and Planning
on August 26, 2016, in partial fulfillment of the
requirements for the degree of
Master of Science in Urban Studies and Planning

Abstract

This thesis uses electricity consumption data from household and enterprise-level smart meters in County B, Country A, and Turin, Italy, to explore temporal and geographic variations in urban energy consumption and thus urban activity. A central question is whether electricity consumption patterns vary between different economic sectors, across space, and between different days of the week and times of year.

This data shows clearly that Country A activity patterns are roughly similar across all seven days of the week, whereas Italian electricity consumption declines markedly on weekends, particularly Sundays. In general, and particularly in Italy, this thesis shows strong seasonality to electricity consumption, with clearly identifiable seasons and high correlation in consumption patterns within each season.

This thesis focuses on user type variation in Country A, where although certain patterns are more widespread in some sectors than others, there is significant overlap between pairs of sectors. Hence this thesis is able only to classify land use between residential and industrial sectors, and is unable to classify land use to a meaningful degree of accuracy by analyzing electricity consumption. It is, however, possible to detect geographic variation: urban and industrial centers consume a higher percentage of their electricity on weekdays and during regular work hours than rural areas.

In addition, the impact of various special occurrences on urban behavior is probed. This thesis provides measurement of the impact of various holidays on economic activity, using electricity consumption as a proxy. Large (industrial) consumers are generally much more sensitive to holidays than small (residential) consumers are, except during the summer months in Italy. In general, consumption declines on a single holiday are highly correlated with consumption declines on other holidays. Furthermore, using observations at 15-minute intervals, I attempt to measure the short-term behavior shifts caused by daylight savings time's start and finish.

Thesis Supervisor: Carlo Ratti
Title: Professor of the Practice

Acknowledgments

My colleagues and advisors at Senseable City Lab were both helpful in shaping the research questions addressed here and valuable for the community they provided this past year. Carlo Ratti provided guidance without which this thesis would not have come to fruition, and was a huge help in resolving last minute hiccoughs. The project began as a joint undertaking with Gregory Falco, whose companionship even after his departure was invaluable. Paolo Santi provided guidance and feedback on various drafts, and his organization of the Urb sub-group provided a welcome opportunity to solicit input from colleagues. Behrooz Hashemian both provided needed advice on the details of the project and some community, particularly through the Machine Learning Working Group he ultimately formed. Wonyoung So and Chaewon Ahn helped out with visualizations and provided some constructive feedback on my own ugly graphics. Fan Zhang provided essential mapping help. Umberto Fugiglando helped make up for my woeful ignorance of all things Turin. Marguerite Nyhan, Emanuele Massaro, Dimitrios Milioris, and Daniel Kondor provided both substantive comments and office companionship. Last but certainly not least, I am deeply indebted to Erin Baumgartner and Sandra Wellford for their help through many an administrative headache—which I probably turned into migraines for them.

Occasional discussion with Ettore Bompard and Huang Tao at Politecnico di Torino helped clarify some questions regarding the Italian data.

Several others provided much needed friendship and occasional input. My parents, Steve Wilson and Norma Kriger, as well as Dan Borgnia, Lucas Orona, and Sasha Wrobel all had to listen to more than their fair share. Zhao Jinhua, Hong Yu-hung, Tan Zhengzhen, Cressica Brazier, and Cui Hongyan furnished much needed outlets for my academic interests, as well as good company.

It came to our attention at the tail end of this project that the data from Country A was somewhat sensitive. As such I have removed all reference to its location. As this was done after the thesis was otherwise complete, it may have led to several awkward paragraphs or sections. My apologies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With the spread of smart meters, research into electricity consumption patterns has become increasingly popular. To date, research has focused on residential consumption and the relationship between household characteristics and consumption patterns. In this thesis, I broaden the scope of electric smart meter research to include industrial consumption. Furthermore, rather than emphasizing the relationship between household characteristics and consumption patterns, I ask how electricity consumption varies over industries and space and how it reacts to various external stimuli such as holidays and weather patterns.

Electric smart meters measure electricity consumption at granular levels, providing detailed consumer-level information on electricity consumption at intervals measured in minutes or seconds. Previously available data aggregated at the city level did not allow research into how behavior varied between different users or types of users. And data at the consumer level aggregated over time–such as traditional electric meter readings–lacks the details of how behavior changes from day to day or hour to hour, about which smart meters provide unique insight.

In some areas, electric smart meter data is collected at a large scale [1]. If ultimately released to researchers, this has the potential to facilitate thorough comparisons of behavior in different cities and between cities and suburban or rural areas. Local variation in work patterns, which likely are highly correlated with electricity consumption, are a particularly appealing direction of research. Furthermore, smart meter data provides a way to detect and quantify changes in these patterns over time.

This thesis makes a modest start towards answering these questions. Working with moderately sized datasets spanning six months in one Country A county and slightly over a year in one Italian city, I ask how patterns of consumption vary between types of users, across space, and between these two distant locations.

1.1 Contribution to Electric Smart Meter Research

To date, the focus of research using electricity smart meters has been residential use patterns, particularly the relationship between household characteristics and residential use. Beckel et al use several machine learning techniques to estimate household characteristics from various statistics derived from daily consumption data [6]. McLoughlin et al approach the inverse problem, using self-organizing maps to group households with similar electric consumption patterns, then conducting logistic regressions to see how well household characteristics predict electric consumption patterns [21]. Kavousian et al explore regression models that estimate the impact of household characteristics on minimum, maximum, average, and the max-min range of daily electricity consumption [15]. All three work with household survey data, and all three focus exclusively on residential meters.

Another line of research—one that has generally proceeded without much large scale data—focuses on identifying individual activities within a metered household using smart meter data. This generally depends on extremely high frequency smart meter data. This branch of research, in turn, splits in to two (overlapping) subsets: one focus is on uncovering these activities [22, 26], while another focuses on covering them back up to protect privacy [13, 10, 26]. Neither is of particular relevance to this research project, which uses data at time intervals large enough to pose significant challenges to any such undertaking.

In considering meters from across the spectrum of electricity consumer types, I diverge from this limited body of electric smart meter research, which has focused exclusively on residential smart meters. While interesting and useful, such work only touches on part of how cities work, ignoring the workplace and associated behavior patterns. More to the point, it ignores the largest electricity consumers.

The existing research on smart meters also ignores geographic variation, of utmost importance to grasping urban systems. Electricity consumption is tied to inherently spatial activities, and as I show in this thesis, electricity consumption does vary across space. Urban areas respond to weekends differently than suburban or rural areas; they also respond differently to some holidays.

Hence this research expands the literature on electric smart meters by incorporating non-residential meters and geographic analysis. This approach also better facilitates scaling up if large datasets on electricity consumption become available in the future. It is simply not feasible to conduct mammoth detailed household surveys, so approaches to analyzing data that are self-contained to the big data collected by smart meters have practical advantages.

1.2 Data

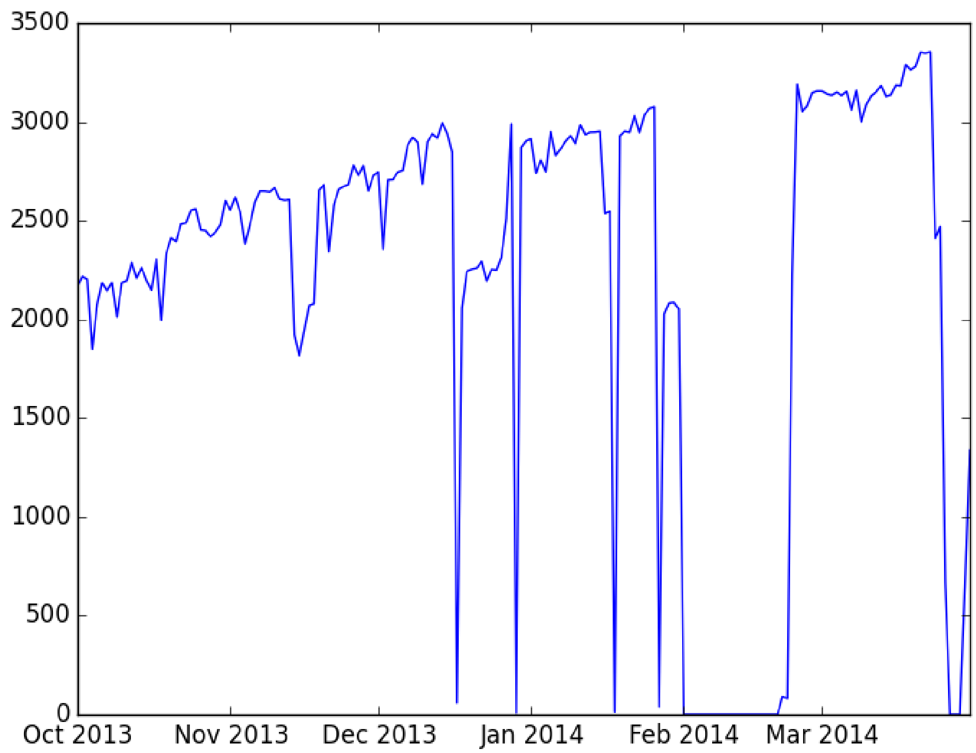
This thesis is based on data from both Country A and Italy. Although the Country A data covers a shorter period of time and is significantly noisier, it includes remarkably detailed information about the rate category and industry type of each meter. The Italian data, on the other hand, covers a much longer period of time but lacks the information about consumers that the Country A data includes. Most importantly, the provenance of the Italian data is much more clear.

The data sets and hence cleaning routines differ slightly for various portions of this thesis, the details of which I describe in the appropriate chapters. Below is an outline of each dataset and data cleaning which applied to all analysis in this thesis.

1.2.1 Country A

The Country A electricity consumption data was collected at fifteen-minute intervals from 4,057 high voltage electricity meters in County B, Country A. Although the data was collected between 1 October 2013 and 31 March 2014, meters have on average only 97 days of clean observations; for much of the month of February, no meters are reporting (see Figure 1-1). The data includes information on the category of the customer that is used to set electricity tariffs,[1] as well as more specific industrial or commercial subcategories; for most meters, there is an associated address. The meters are *not* a representative sample of all meters in the county.

Figure 1-1: Working Meters per Day (Country A)



The 4,057 meters were chosen by cleaning a slightly larger dataset. I remove meters for which the rate category was not listed. As the raw data reported cumulative rather than incremental consumption, I remove those meters that reported no cumulative consumption for any point in time, except where it was a missing value immediately between two non-zero readings, in which case the average of those two were substituted. I convert the data to incremental consumption, and removed meter-days

[1]This categorization is not based upon electricity consumption, but is intended to be what the user actually uses the electricity to do.

Table 1.1: Data Quality by Rate Category (Country A)

Rate Category	**Total Meters**	**Median # of Data Days**	**Mean # of Data Days**
Agricultural Irrigation	15	87	87
Agricultural Production	25	103	93
Commercial	306	120	102
Common Industry	652	114	99
Education - K-12	45	127	107
Large Industry	902	116	101
Large Industry - Fertilizer	5	142	140
Non-Industrial	680	115	96
Non-Residential Lighting and Imagery	497	117	101
Residential	930	109	94

with any observation of negative consumption (none of these meters sold electricity back to the grid). I further remove meter-days with little or no consumption, that is where the 96th percentile fifteen-minute interval consumed no electricity or the 100th percentile fifteen-minute interval consumed less than 0.02 kWh (this is because the data has precision of 0.01 kWh). Lastly, to avoid spikes in the data, I remove meter-days where electricity consumption in the 100th percentile fifteen-minute interval exceeds by more than ten times consumption in the 96th percentile fifteen-minute interval. Because the incremental consumption for the first fifteen minute interval of many days cannot be computed, I work with 95 fifteen-minute intervals for each day.

It came to our attention at the tail end of this project that the data from Country A was somewhat sensitive. As such I have removed all reference to its location. As this was done after the thesis was otherwise complete, it may have led to several awkward paragraphs or sections. My apologies.

A Note on Representativeness

The data from County B does not represent all the electricity consumption in County B, let alone all consumers in Country A. Although throughout this thesis, I make reference to "Country A," this should be understood as shorthand for the unrepresentative sample from County B.

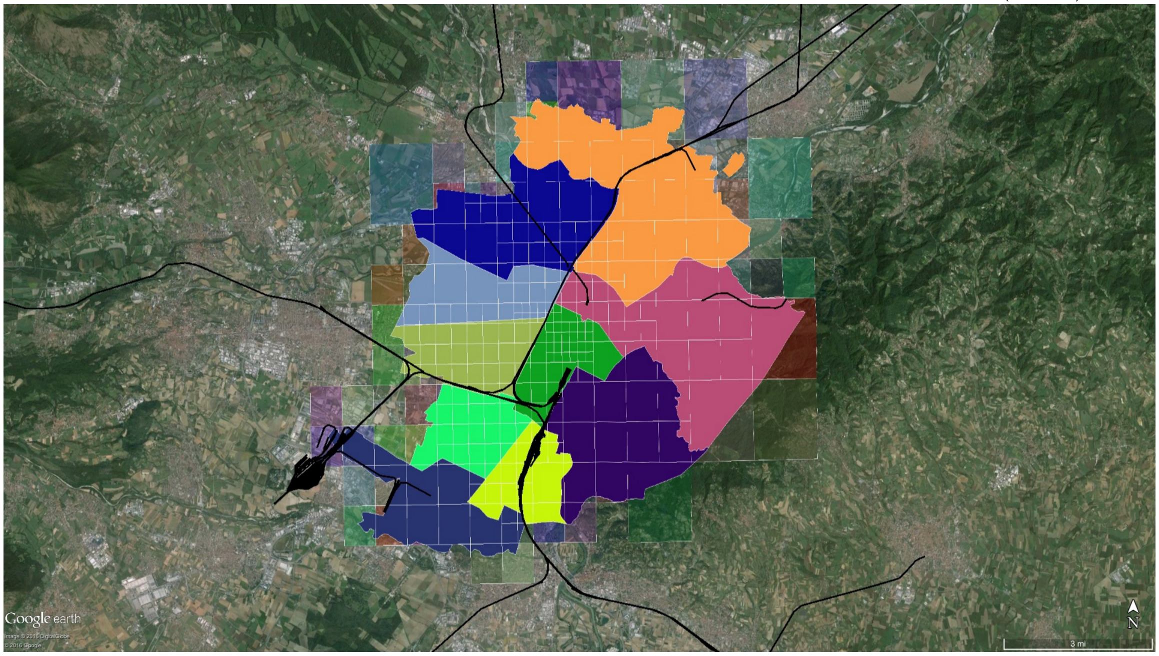
Thus although our small sample of high-voltage meters can hardly be said to "represent" County B, it does give us a broad sense of electricity consumption both in the established urban fringe to County B's northwest, in the new boomtown urban fringe (and construction site) along County B's central north-south axis, and in the peri-urban rural areas to County B's south and west.

1.2.2 Italy

The Italian data for this project comes from IREN, the major local electricity provider in Turin, Italy. This data is best seen as two separate datasets, one covering 2,612 large consumers (those with contracts covering 55 kW or more of power) and another covering 1,595 small consumers (below 55 kW). The first dataset covers the period from 1 January 2015 to 31 January 2016, including all large consumers in Turin. The small consumer dataset is a random sample of small consumers in Turin, covering the period from 1 June 2014 to 29 February 2016. Both datasets consist of fifteen-minute interval consumption data. The data does not include specific addresses or user categories, but does list the contracted power for each meter and the location rounded to a grid. The grid is more detailed in the downtown area of Turin but each square covers a relatively larger area in suburban locations. When presented visually in this paper, rather than using this grid, data is aggregated at the city ward (*circoscrizioni*) level, and each grid block is assigned to the neighborhood in which its center falls. Figure 1-2 shows the original grid overlaid by the wards. (I remain skeptical about the accuracy of certain geographic information in this dataset.)

Compared to the Country A data, relatively little cleaning was performed on the Italian data. First, all observations of negative consumption were removed. Second,

Figure 1-2: Map of Turin with Grid and Neighborhood Boundaries (Italy)



Note: Credit to Google Earth Pro for this map.

days during which consumption was negligible but highly concentrated were removed: if the 90th percentile of fifteen-minute consumption intervals was zero and the maximum was greater than 0.02 kWh, then the meter-day was excised. Then, as with the Country A data, large spikes were removed. This was accomplished by removing meter-days where the 90th percentile fifteen-minute consumption interval was less than 1% of the maximum consumption.[2] Lastly, only meters for which contracted power and geographic information was available were included.

Table 1.2: Data Quality by Electricity Consumption (Italy)

Rate Category	**Total Meters**	**Median # of Data Days**	**Mean # of Data Days**
Large Consumer (>55 kW)	2612	362	317
Small Consumer (<55 kW)	1595	602	515

Shape files for Turin *circoscrizioni* come from AperTO Torino [23].

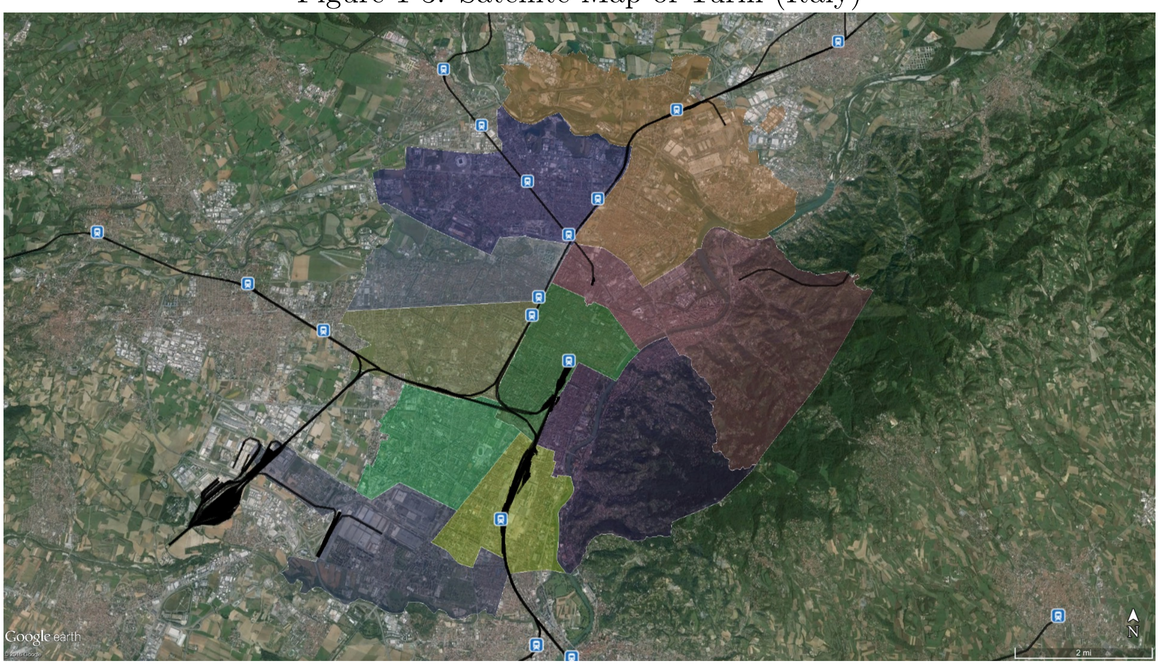
[2] A different standard was used for the Italian data than the Country A data because we sought to use the weakest standard that excluded outliers that distorted our clustering, for example by yielding single member, noisy clusters.

A Note on Representativeness

The large consumer data from Turin, as a complete sample, is we trust representative of Turin's large consumers. It is not necessarily representative of the Turin region's large consumers, let alone Italy's large consumer's as a whole, and references to "Italy" throughout this thesis should be read as shorthand for Turin. Nonetheless, it is not unreasonable to believe that this data provides insight into patterns of consumption that are likely not radically dissimilar elsewhere in northern Italy.

The small consumer data from Turin is a random sample. Again, it is not likely representative of Italy as a whole, although the patterns are likely reminiscent of those elsewhere.

Figure 1-3: Satellite Map of Turin (Italy)



Note: Credit to Google Earth Pro for this map.

Turin is home to a mix of residential, commercial, and industrial consumers. Turin's downtown is the central dark green-colored neighborhood in Figure 1-3. Its adjoining neighborhoods are predominantly residential. The northernmost and southernmost neighborhoods shown on the map are relatively industrial. This provides a nice mix of consumer types with which to work.

1.3 Outline of Thesis

This first chapter outlines the data used in this thesis. As well, it specifies some of the limitations on this data and hence the generalizability of the results from the succeeding chapters. In short, our Country A data includes geographic and user category information, but is not representative and is drawn from a suburban county of a major metropolitan area; hence we cannot draw conclusions about truly urban areas of Country A, nor can we claim our results are representative even of the given county. We can, however, trust that our data is relatively more complete with respect to large enterprises. Our Italian data is representative within Turin and covers a relatively long time period, giving us confidence in the results we glean from it; however, we do not have precise geographic or user category data, restricting our ability to perform fine-grained analysis.

Chapter 2 proceeds to compare Country A and Italian electricity consumption over time and space. Our primary finding is that Country A consumption patterns pay little heed to weekends, whereas Italian consumption patterns—particularly for larger consumers—suggest significantly less productive activity on weekends. We further show that these consumption patterns vary geographically: weekday consumption tends to be particularly high relative to weekend consumption in suburban areas, while the two are more comparable in rural and downtown areas. We note, as well, that consumption patterns are highly seasonal, and that these seasons are often clearly bounded, with high day-to-day correlation within seasons. Lastly, we observe that in Italy electricity consumption is relatively stable from year to year, although consumption during the summer holiday months may have increased.

Chapter 3 attempts a classification of land use using electricity consumption curves, using only the Country A data. Attempting several machine learning techniques, we find that electricity consumption patterns can classify land use between industrial and residential purposes with tolerable accuracy, but is ineffectual at classifying into more fine-grained categories. We ascribe this failure to classify to the similarity between various user categories' consumption patterns observed in Chapter 2: although, overall, average consumption patterns differ between user categories, performing clustering on each category yields clusters that are quite similar to those in other categories.

Chapter 4 explores the impact of holidays and time changes on electricity consumption. The main conclusion is that industrial or large scale consumers do differ markedly from small scale or residential consumers in their response to such stimuli. Large consumers adjust consumption downward during holiday periods by large percentages, whereas small consumers generally continue to consume at their normal pace. The exceptions are summertime holidays in Italy: these are associated not only with large drops in large consumers' consumption, but also in substantial drops in small consumers' consumption. In general, consumers that reduce consumption for one holiday are much more likely to do it for another.

After brief conclusions in Chapter 5, two appendixes test potential correlates of electricity consumption at more aggregate levels. Appendix A looks at the relationship between weather, seasonal variations in the length of the day, and electricity consumption. Controlling for day of week, a simple linear regression model performs remarkably well using these variables to predict aggregate daily electricity consumption for large consumers, and well enough for small consumers to warrant consideration. Appendix B asks whether our data can be used to test the widely held hypothesis that night lighting is a good proxy for electricity consumption and thus for economic activity. We find that extremely high electricity consumption tends to occur primarily in well lit areas, but that well lit areas and poorly lit areas in many cases consume similar amounts of electricity.

Chapter 2

Temporal Patterns of Electricity Consumption

We begin our exploration of electricity consumption by asking what patterns of consumption prevail in different industries and geographic areas. Of particular interest is how consumption varies over the average week: Is consumption on weekdays radically different from weekends? Is most electricity consumed during work hours, in the evening, or late at night?

We find both variations and similarities across space and industries in the answers to these questions. Generally, industrial and large consumers are more likely to consume more electricity in the morning, and in Turin, much less electricity on weekends. Residential and small consumers are more likely to consume the most electricity in the evening. There are, moreover, geographic patterns of weekday and weekend consumption distinctive to both Turin and County B that belie any simple generalization.

2.1 Literature Review

In exploring time of use of electricity, we build on the standard methods used in the field. Time-of-use analysis for smart meters has generally been performed in one of two ways: through regression analysis and k-means clustering. Regression analyses are particularly well suited to those studies with household characteristic data, generally collected through surveys. This allows for correlational studies between consumption in homes and the existence of specific appliances, etc. [15, 6]. k-means clustering is also a widely used method for clustering household characteristics from energy consumption. While it has yielded strong results, it is extremely sensitive to how the data is prepared and relatively subjective as the researcher defines the appropriate number of clusters.

In using k-means clustering to explore service usage patterns across a region, we build as well on work using cell phones and internet use. Reades et al. show that different functional areas have different patterns of cell phone usage [27]. Further, Calabrese et al. find that k-means clustering on MIT wireless network use patterns results in one cluster that is primarily residential, although other clusters do not classify with high accuracy [3].

2.2 Data

For this exercise, we consider data primarily in two forms: as a weekly consumption curve and as the ratio of weekday to weekend consumption. For the former, we average meters across all available weeks, remove those with no data for certain days of the week, then rank the total consumption of meters and map them onto a sigmoid cumulative probability function to avoid outlier consumption figures. This results in 3,740 usable meters for the Country A data and 4,170 usable meters for the Italian data. For the weekday-to-weekend ratio, we remove meters if for any time of day the ratio does not exist and, to avoid outliers, remove those for which the fifteen-minute interval with the maximum ratio exceeds thirty. This results in 3,405 usable meters for the Country A data and 3,653 usable meters for the Italian data.

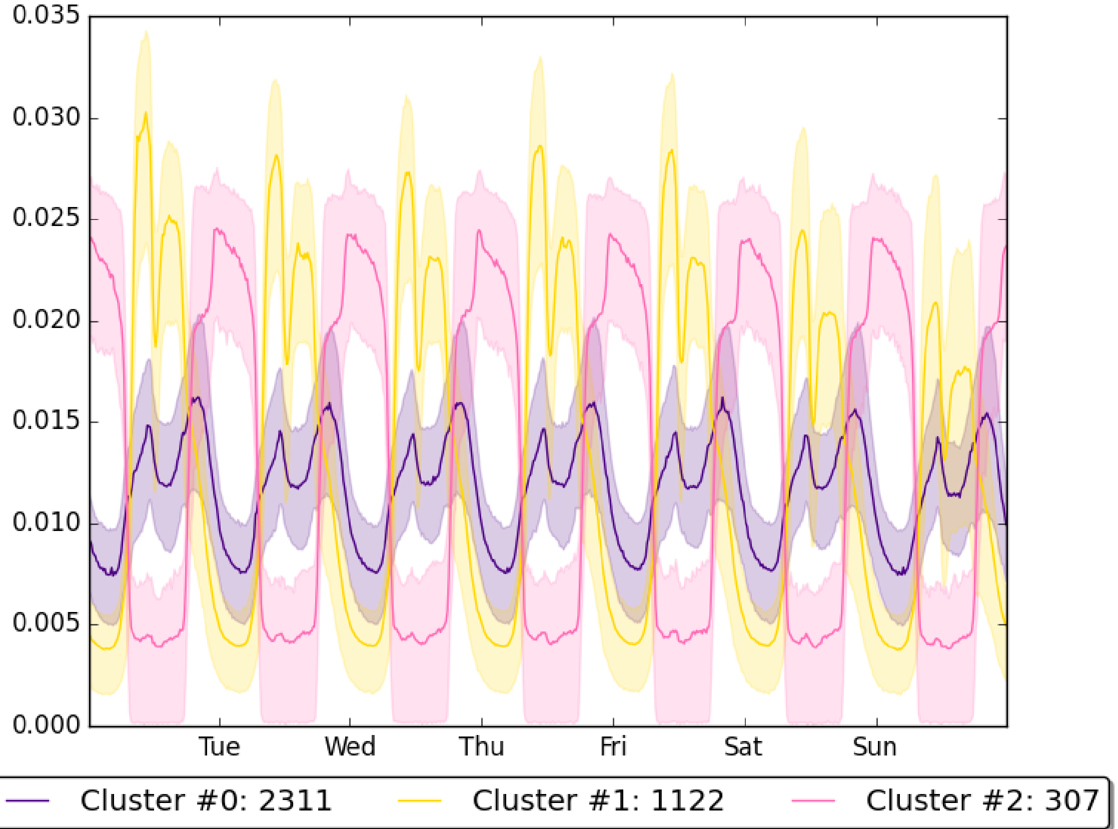
2.3 Clustering on Consumption Curves

2.3.1 All Rate Categories

Country A

Working with our weighted data, we perform spectral clustering and k-means clustering with between 2 and 10 clusters.[1] The Davies-Bouldin index indicated that the "tightest" clustering for the Country A data was k-means clustering with 3 clusters, as presented in Figure 2-1.[2]

Figure 2-1: k-Means Clustering on Average Weekly Consumption, Scaled to Sigmoid (Country A)



Note: Shaded areas denote the interquartile range.

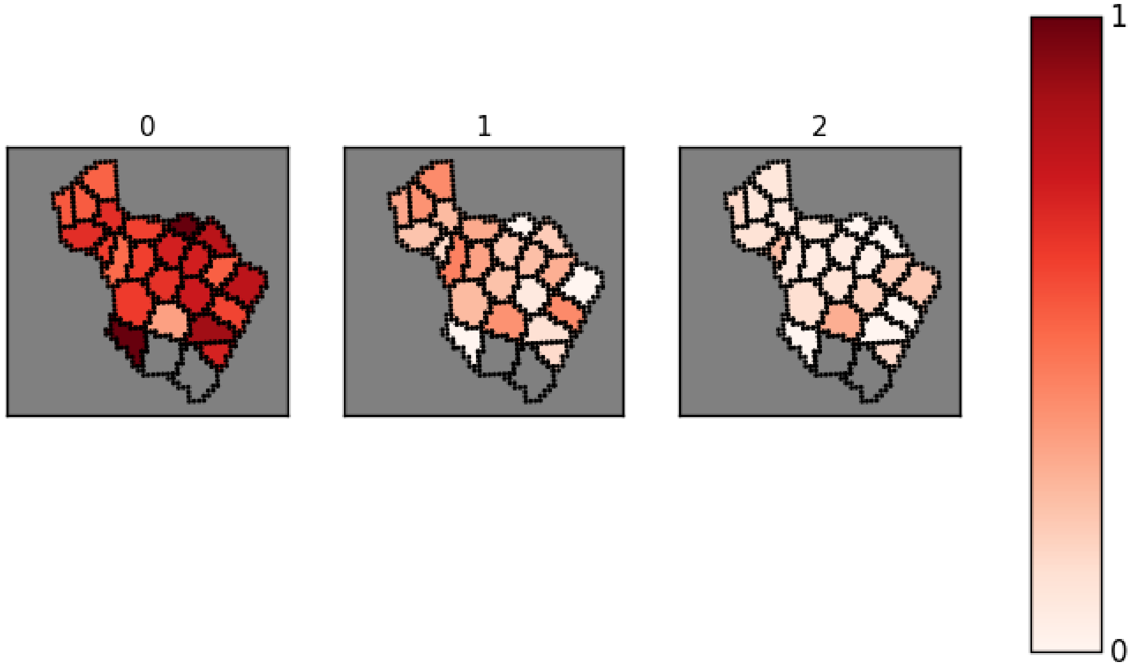
The dominant pattern (Cluster #0) is roughly constant across all days of the week,

[1]For information on spectral and k-means clustering, see [19] and [20], respectively. DBSCAN clustering and hierarchal clustering using the normal Euclidean metric were also attempted for the Country A data, but tended to give a single large cluster and a good deal of noise.

[2]See [7] for details on the Davies-Bouldin index.

with a late morning local maximum in consumption, followed by an early afternoon dip and an evening global maximum in consumption. This pattern is particularly prevalent among residential consumers, probably reflecting a pattern of high electricity consumption before they leave for work and after they return home. However, commercial users also are overwhelmingly in this cluster, and indeed all rate categories have at least a plurality of their meters in this cluster. Looking at more fine-grained sub-categories for non-residential users, we see that services, real estate, education, and construction are particularly dominated by this category—unsurprising, as they are generally industries that operate all day and rarely late at night. A particularly high percentage of meters in towns to the rural south and east of the county are in this cluster, but it is well represented in the county seat and the growing urban suburbs as well (see Figure 2-2).

Figure 2-2: k-Means Clustering on Average Weekly Consumption, Mapped by Town



Note: Darker colors denote towns with a higher percentage of their meters in the given cluster.

The second largest cluster (Cluster #1) exhibits a strong weekday and daylight bias, with a steep drop in consumption at lunchtime and almost no nighttime consumption. This cluster is best represented in the industrial rate categories, as well as the non-industrial rate category (in which, puzzlingly, some forms of manufacturing are included). Hence it is unsurprising that, looking at sub-categories, this cluster constitutes a majority or almost a majority of meters for several forms of manufacturing: chemical manufacturing, food manufacturing, furniture manufacturing, leather manufacturing, machinery manufacturing, and metals manufacturing. Construction material manufacturing, however, show a more nocturnal consumption pattern. Geographically, this cluster tends to include a higher percentage of the meters in towns to the northwest of the county, home to the county seat and the more urbanized areas—this is unsurprising given its disproportionately industrial character.

The smallest but perhaps most distinctive cluster (Cluster #2) is purely nocturnal. This cluster primarily reflects the presence of some streetlights in our sample, and hence is particularly well represented in the Non-Residential Lighting and Imagery rate category to which they are assigned. It does, however, also include a nontrivial (but minority) portion of the construction materials manufacturing sub-category. In some sense, the absence of other industrial sub-categories in this cluster is remarkable: very little is going on at night.

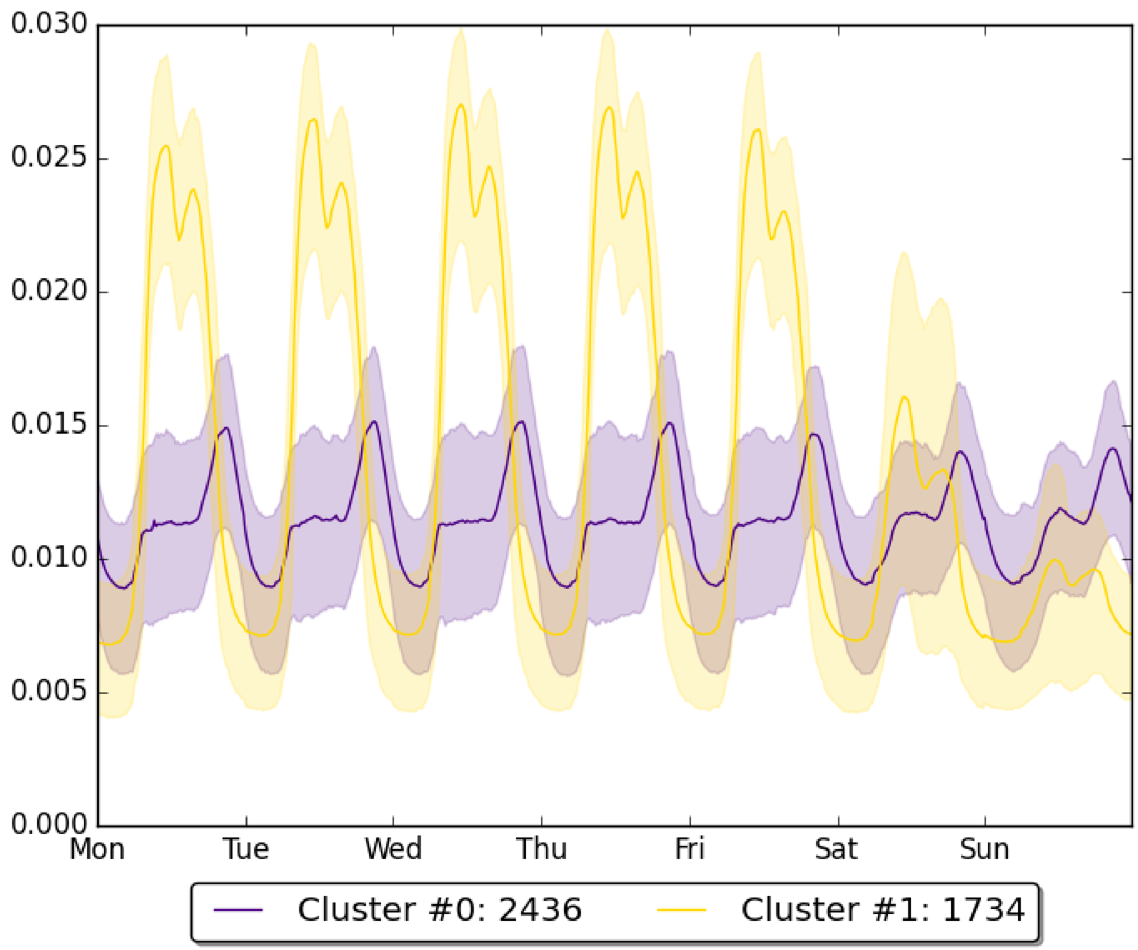
For each meter, we also classify each full week of data into one of the above clusters, to explore if the clusters are relatively stable across time or if meters float from one cluster to another. We find that the daytime and nighttime peaked clusters are quite stable, with almost all their constituent meters classified, for each week, into the respective cluster. However, the cluster with relatively low consumption but an evening peak includes many meters that are classified into different clusters week-to-week. By and large, they vary between the daytime and evening peak clusters, as these clusters are somewhat similar.

Cluster Peak	Meters Solely in Cluster, Considered Weekly	Meters in Cluster
Evening (Cluster #0)	1006 (44%)	2311
Daytime (Cluster #1)	1069 (95%)	1122
Nighttime (Cluster #2)	270 (88%)	307

Italy

Working with the Italian data, we find that the "tightest" clustering is spectral clustering with 2 clusters, as presented in Figure 2-3.[3] These clusters are remarkably different from the Country A data presented above, with much more pronounced differences between weekdays and weekends.

Figure 2-3: Spectral Clustering on Average Weekly Consumption, Scaled to Sigmoid (Italy)



Note: Shaded areas denote the interquartile range.

The larger cluster (Cluster #0) resembles the first of the Country A clusters, which was primarily residential: consumption peaks in the evening at 9 pm, with moderate but steady consumption during the daytime from around 8 am to 4 pm. Late night

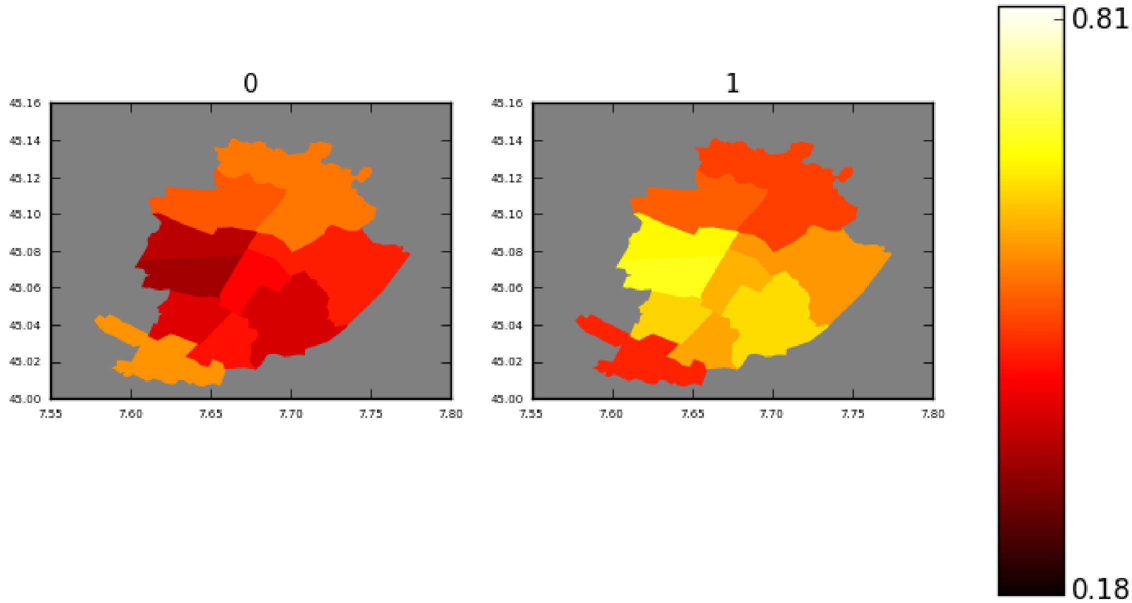
[3] k-means clustering with two clusters yields an almost identical Davies-Bouldin score and almost identical clusters.

consumption is very low, reaching a global minimum at around 4 am. Weekend and weekday consumption are roughly comparable, although there is a slight increase in consumption midday on weekends, and morning consumption is markedly lower. Evening consumption is lowest on Saturdays, followed by Sunday and Friday; it is highest on Wednesday. In general, evening consumption has a long tail, gradually declining into the early morning hours. This cluster predominates in the center city and residential suburbs east and west of downtown.

The smaller of the two clusters (Cluster #1) shows consumption primarily concentrated during the work week. Consumption peaks in the morning around 11 am, dips briefly midday around 1 pm, then reaches a local maximum in the afternoon around 4 pm before plunging in the evening. The nighttime lull in consumption lasts from approximately 8 pm to 7 am, with about 65% of weekday consumption concentrated between 8 am and 6 pm. Consumption peaks on Wednesdays, but each workday is roughly identical, except that Friday afternoon consumption is noticeably lower than afternoon consumption earlier in the week. Saturday consumption, however, is much lower than weekday consumption, and Saturday afternoon consumption is significantly less than Saturday morning consumption. Sunday consumption is negligible. Weekend consumption tends to start later in the morning and conclude earlier. This suggests a regular workplace consumption pattern in which entire enterprises either shut for the weekend (particularly Sunday) or reduce operations substantially. This is notably absent in the Country A clusters. This cluster predominates in the southernmost neighborhood of Turin (where the Fiat factory is) and in the northern industrial areas.

A few cautionary notes are in order. First, we cannot check whether the clusters really correspond to residential and workplace meters, as we do not know much about the users in Turin. Moreover, although the apparently "residential" cluster is only moderately more numerous than the apparently "workplace" cluster, it consists primarily of small consumers, who are underrepresented in our sample. (Recall that we work with a complete dataset of large consumers and a sample of small consumers.)

Figure 2-4: Spectral Clustering on Average Weekly Consumption, Mapped by Neighborhood



Note: Lighter colors denote neighborhoods with a higher percentage of their meters in the given cluster.

Comparative Conclusions

Most striking in comparing the Country A and Italian clustering results is the much larger variation between weekdays and weekends in Italy. In Country A, all three clusters were roughly identical from weekdays to weekends. In Italy, on the other hand, consumers whose peak daily consumption was higher on weekdays tended to have much lower consumption on weekends. Moreover, this finding is unlikely to be a product of poor sampling in the Country A data, as our sample appears to skew towards larger enterprises, which are precisely what is driving the weekend consumption lull in Italy.

2.3.2 Clustering within Rate Categories

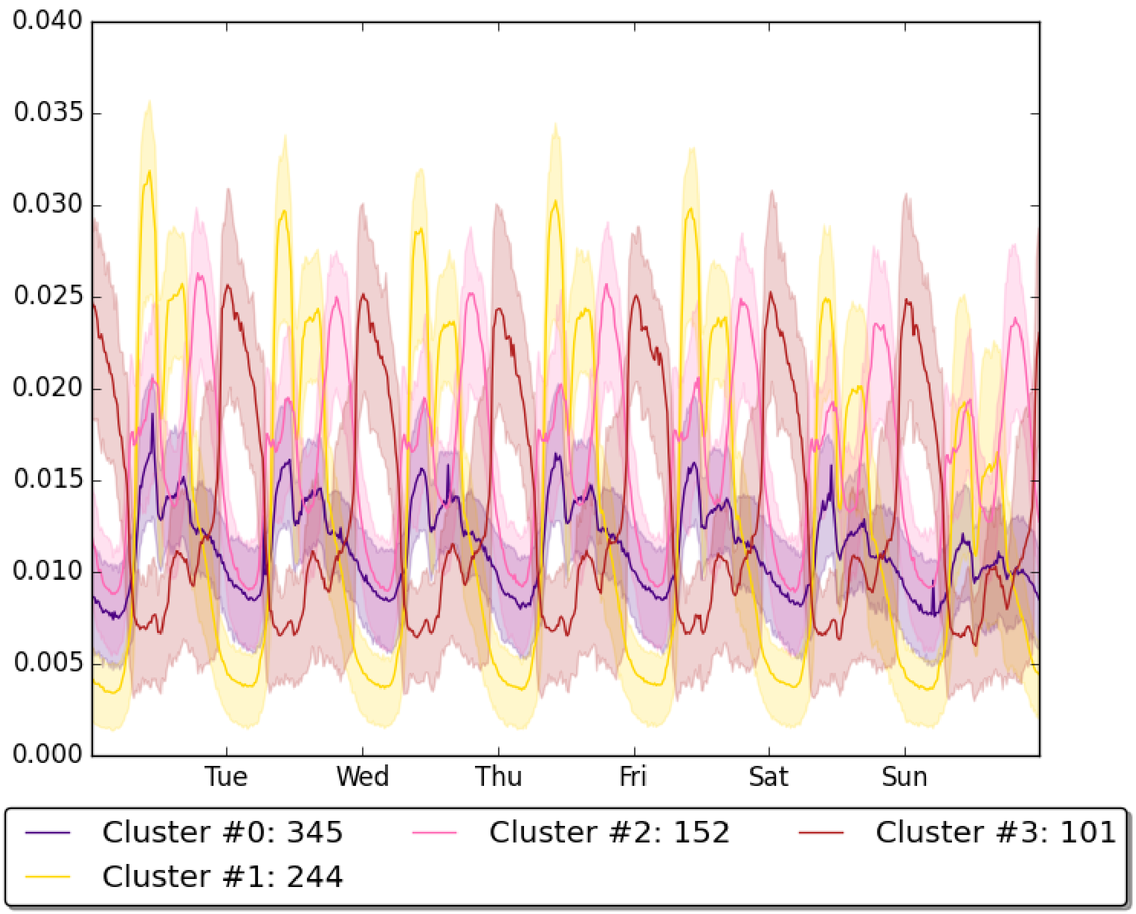
To further explore the geographic distribution of consumption patterns, we perform clustering within categories of users. In the Country A case, we are able to divide users by rate category; in the Italian case, we use contracted power as a proxy for user type.

Country A

We perform k-means clustering on four of the larger rate categories, setting the number of clusters for each by finding the minimum Davies-Bouldin index for between two and four clusters. Our results at this level of detail are already severely restricted by the limited geographic spread of meters within each rate category, not to mention their small number.

Residential meters fall into two clusters, one dominant, evening-peaked cluster and another daytime cluster. The daytime cluster is more predominant in the north of the county, with its higher rate of urbanization; to the extent that any residential customers in our sample reside in the southern part of the county, they fall exclusively in the pattern with some morning and significant evening consumption.

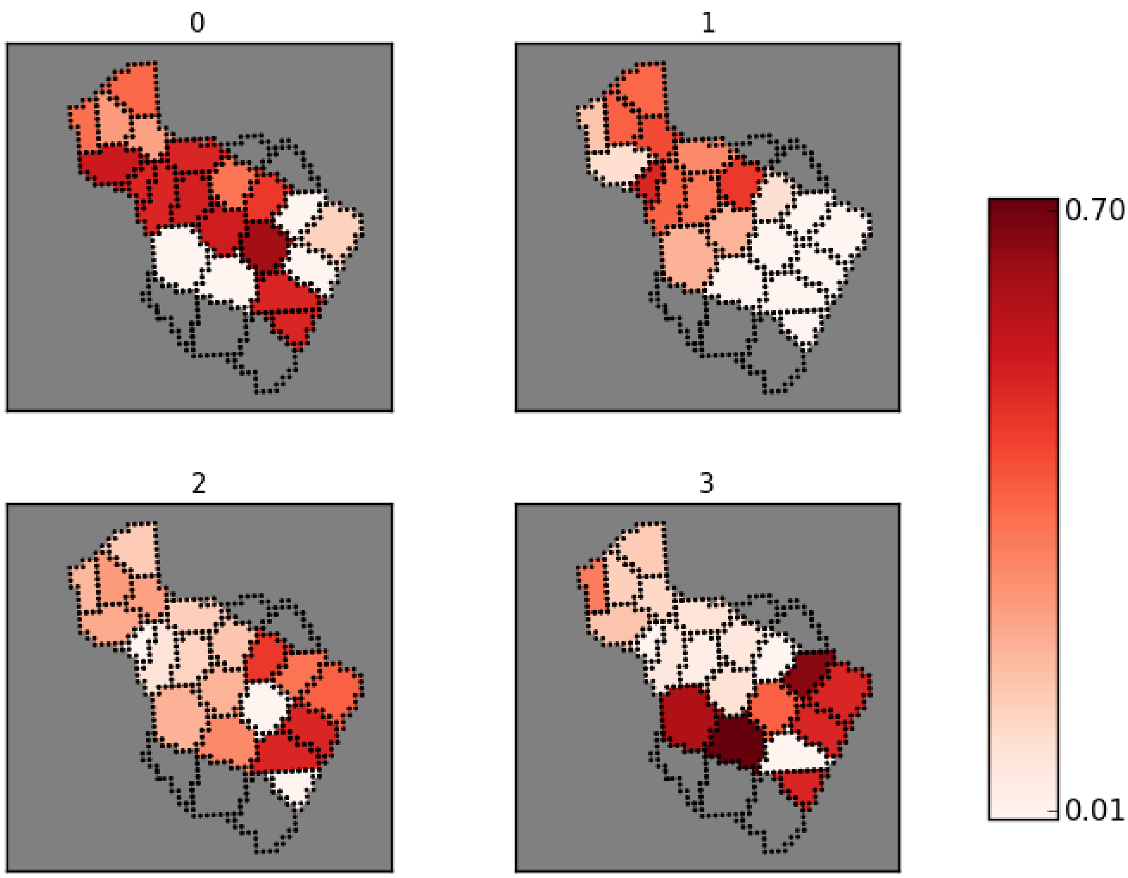
Figure 2-5: k-Means Clustering on Large Industry Users



Note: Shaded areas denote the interquartile range.

Common and large industry display a similar pattern, with more evening consumption in the rural periphery. We form two clusters of common industry meters, one with roughly the same consumption pattern every day, peaking in the evening, and another with more consumption on weekdays, peaking in the morning. Both are widely spread across the county, but the latter is slightly more concentrated near the county seat. We form four clusters of large industrial meters: two with morning peaks (Clusters #0 and #1), one with an evening peak (Cluster #2), and one dominated by nighttime consumption (Cluster #3) (see Figure 2-5). The former two predominate in the urban center and the latter two in the rural southeast of the county (see Figure 2-6).

Figure 2-6: k-Means Clustering on Large Industry Users, Mapped by Town



Note: Red dots outline townships with no users.

Commercial users display an opposite pattern, with late evening and nighttime consumption concentrated around the county seat. After we form four clusters, a low-consumption, early evening peaked cluster dominates. As well, there are a weekday morning-peak cluster, a nocturnal cluster, and a high-consumption evening-peaked cluster. The first cluster is roughly evenly distributed outside of the county seat, but the high-consumption evening-peaked and nocturnal clusters are very lightly represented in the rural parts of the county.

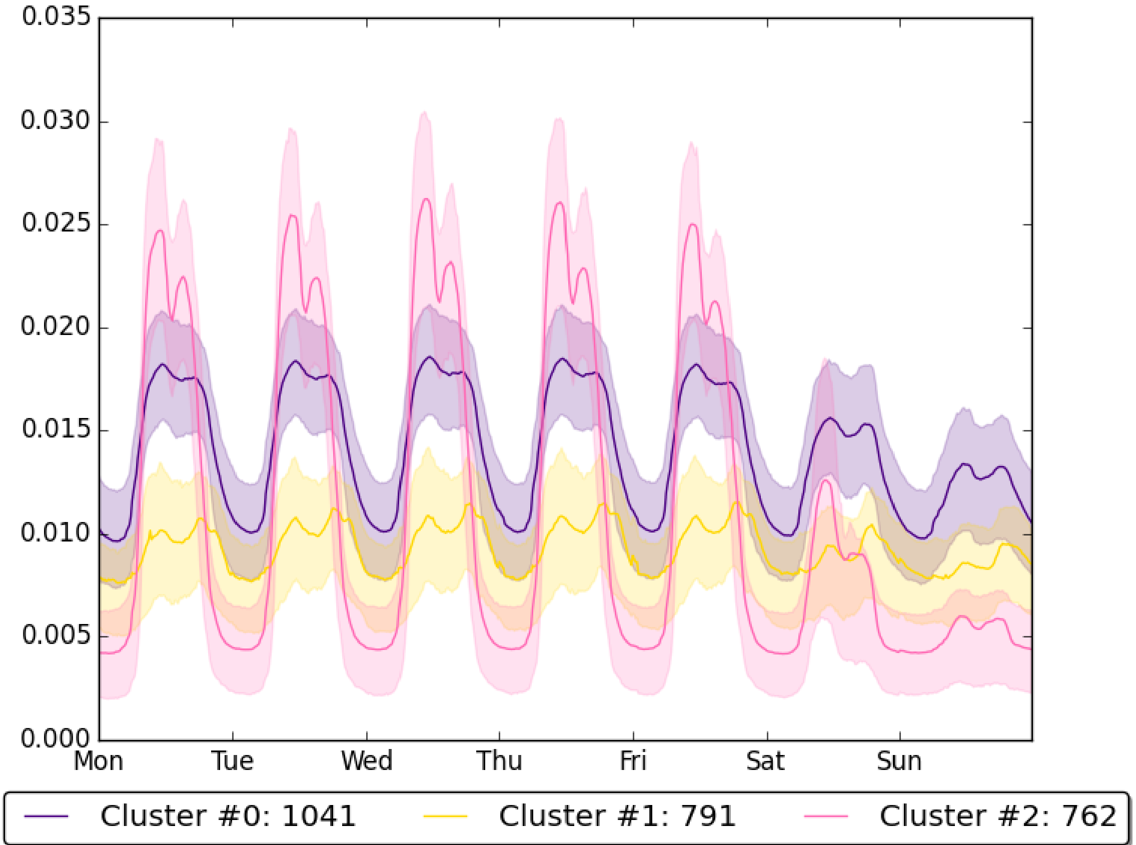
In short, we find that the pattern of more rural users consuming more electricity in the evening holds for residential and common and large industry users. This does not appear to hold true for commercial users. However, these results are particularly subject to our small and unrepresentative sample, and should be interpreted conservatively.

Italy

Since the Italian data does not include user category information, we use the contracted power as a substitute, noting that users with very high contracted power ($>$ 55 kW) are unlikely to be common residential consumers. However, users with low contracted power ($<$ 55 kW) are quite possibly not only residential consumers but also common commercial or retail entities. Nonetheless, this division is the best we can accomplish with the given data and does provide some insight into variations in consumption patterns between larger and smaller consumers. It is worth reiterating that the large consumer dataset is complete: that is, all consumers with high contracted power are included in the dataset.

Performing spectral clustering on these large consumers, the optimal number of clusters is three (Figure 2-7). The smallest of the three clusters (Cluster #2) closely resembles the "workplace" cluster identified in the preceding section. The largest cluster (Cluster #0) also shows slightly higher morning than afternoon consumption with a more mild midday dip in consumption; the daily peak period of electricity consumption (that is, the "workday") is longer for this cluster. Cluster #1 shows the least daytime electricity consumption, with a peak in the evening rather than the morning, and with Sunday consumption almost in line with late night consumption. Interestingly, in contrast to the Country A data, even considering up to nine clusters, there is no cluster with primarily nighttime consumption (although beginning with the sixth cluster there is a small group of meters whose consumption peaks in the late evening).

Figure 2-7: Spectral Clustering on Average Weekly Consumption for Large Consumers, Scaled to Sigmoid (Italy)

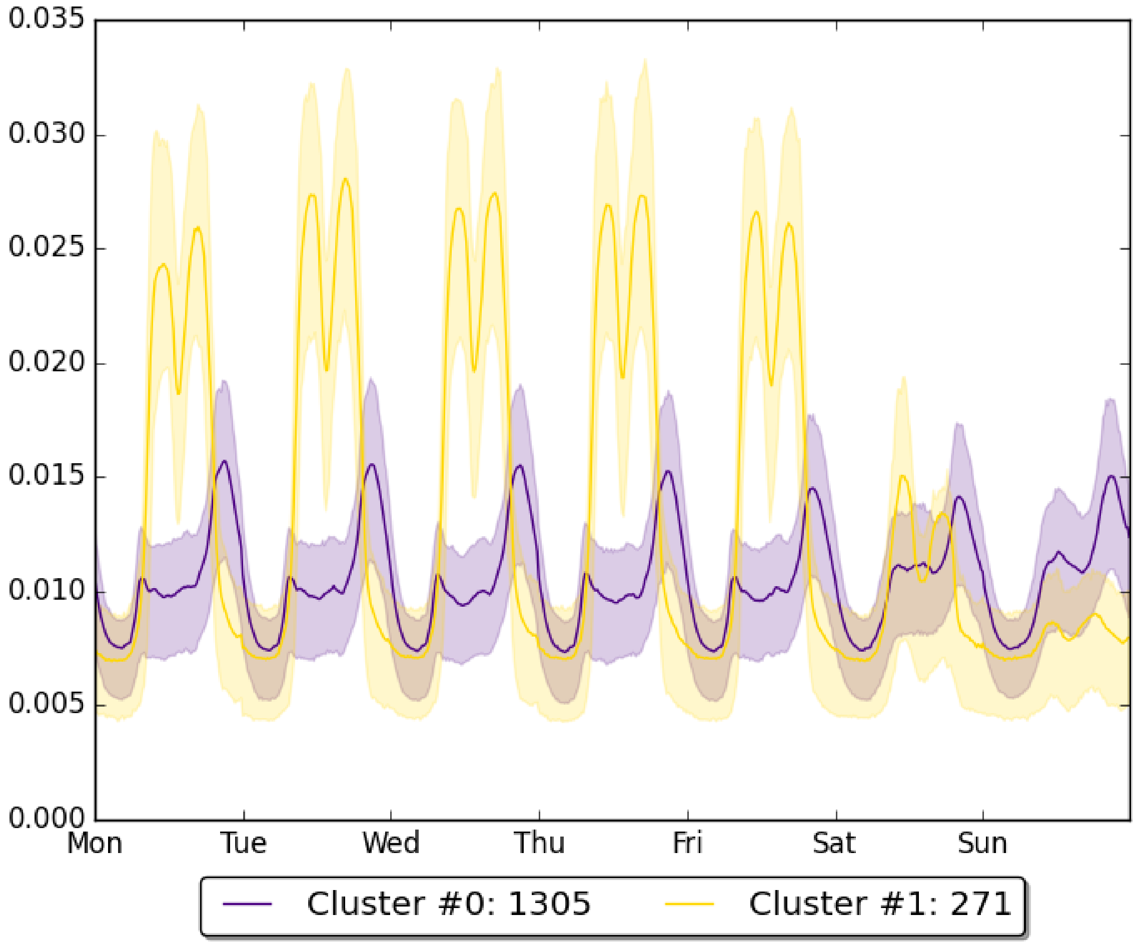


Note: Shaded areas denote the interquartile range.

These large consumer clusters lack clear spatial patterns, although the high workday consumption pattern, Cluster #2, concentrates in the downtown and eastern suburbs, while Cluster #0 predominates in the southern neighborhoods.

Turning to smaller consumers, the "tightest" clustering is spectral clustering with two clusters (Figure 2-8). Here, the vast majority of meters fall into Cluster #0, closely resembling the "residential" cluster from the previous section, with an early evening peak in consumption. This cluster predominates everywhere in Turin, but is less pronounced in the downtown area and the southernmost area (with the Fiat factory, see Figure 2-9). Cluster #1 shows a pronounced weekday work-hours bias,

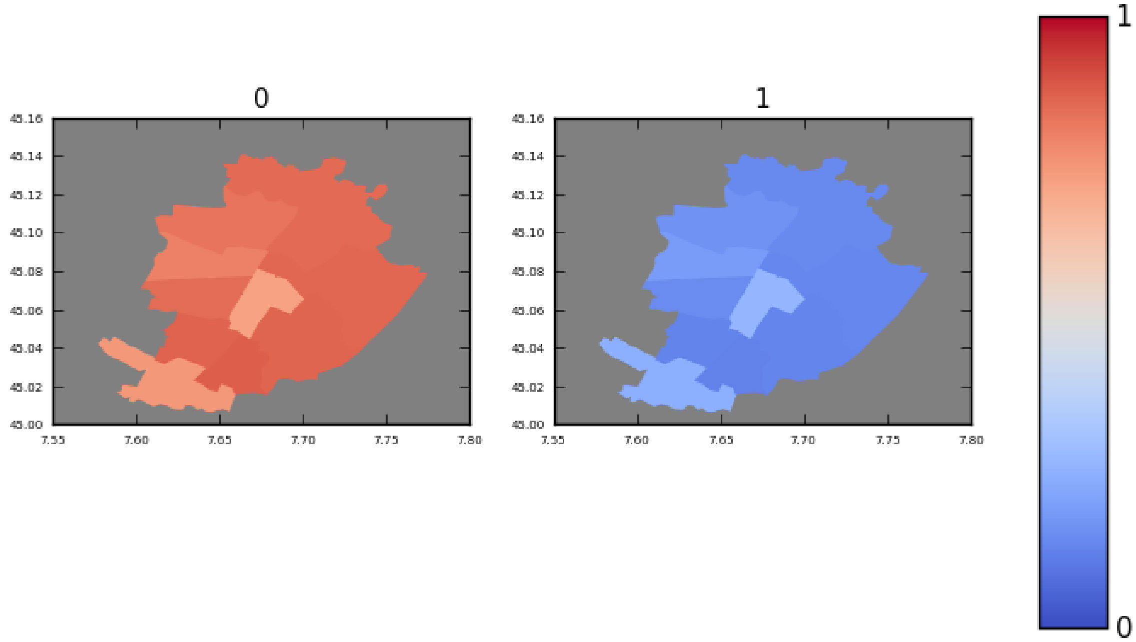
Figure 2-8: Spectral Clustering on Average Weekly Consumption for Small Consumers, Scaled to Sigmoid (Italy)



Note: Shaded areas denote the interquartile range.

with twin morning and afternoon consumption peaks divided by a midday dip in consumption. Saturday consumption is significantly less than weekday consumption (and significantly higher Saturday morning than Saturday afternoon); Sunday consumption is negligible. This cluster strongly resembles commercial or perhaps retail consumption patterns, and is particularly predominant in the downtown and southern (Fiat factory) neighborhoods. While the first "residential" cluster (Cluster #0) is relatively similar to the larger of the Country A residential clusters, it lacks an early morning consumption peak and its evening consumption peak is far more pronounced than the Country A one. The second, "commercial" cluster (Cluster #1) is unlike any of the Country A commercial clusters, which are relatively consistent all seven days of the week.

Figure 2-9: Spectral Clustering on Small Consumers, Mapped by Neighborhood (Italy)



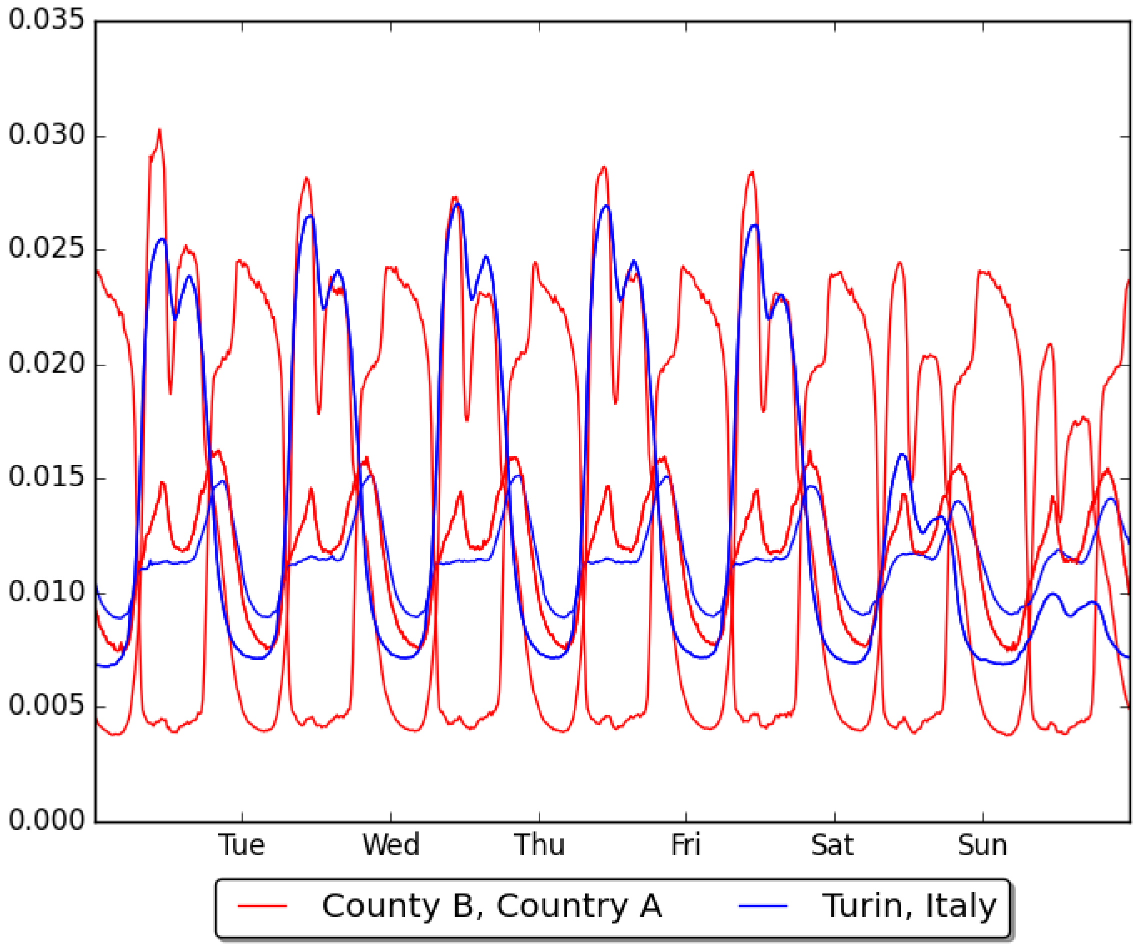
Comparative Conclusions

Again, the most striking contrast between Italy and Country A is the far greater extent to which Italian workplaces rest on weekends, as visible in Figure 2-10. A secondary finding is that Italian workplaces are less nocturnal than Country A ones, and indeed our best guess as to which meters are commercial in Italy suggest that they maintain remarkably restricted hours, limited to regular work hours on weekdays–unlike their Country A counterparts, which are more likely to operate in the evening.

2.3.3 Clustering Days over Time

As a last clustering exercise, we cluster the mean consumption curves of each day in our sample. This allows us to compare days–for example, do days in the summer consume more than days in the winter, or do weekends consume in different patterns than weekdays. Our results largely conform with those presented above.

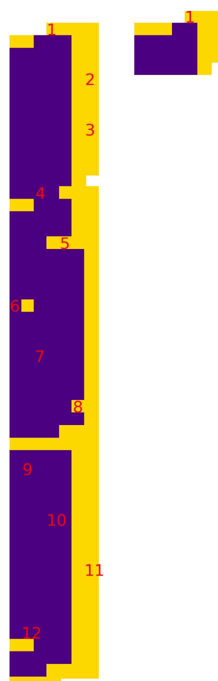
Figure 2-10: Comparison of Country A and Italian Clustering Results



Italy

In general, daily consumption patterns in Italy differentiate primarily between weekdays and weekends and secondarily by season. This is particularly true for large consumers, for which a relatively high mean consumption curve predominates during the week but a lower one is customary on weekends (Saturdays and Sundays in the winter, Sundays from May through mid-August), holidays and several additional days in August (see Figure 2-11). (Considering additional clusters, higher weekday consumption becomes the norm starting in May, with a yet higher consumption cluster taking over in July.)

Figure 2-11: Clustering Consumption by Day, Large Consumers 2015-2016 (Italy)



Note: Numbers denote the month of the year.

A quite similar pattern emerges among small consumers, with generally lower consumption on weekends and holidays. However, consumption goes down in August and late June 2015, and is generally lower in the spring and fall than in the summer and winter.

Country A

A similar approach to our (temporally more limited) Country A data does not separate days into weekends and weekdays (see 2-12). Rather, late December and early January–as well as early March–join a high consumption cluster. Only when looking at four clusters do weekends appear at all.

Figure 2-12: Clustering Consumption by Day, 2013-2014 (Country A)

Note: Numbers denote the month of the year.

2.4 Variation between Weekdays and Weekends

Another approach to differentiating users of electricity is to consider the difference between their weekday and weekend consumption. We begin by looking at the ratio of weekday to weekend consumption. This detects both entities that are only open on weekdays or weekends as well as tendencies to wake up earlier on weekdays, etc. In general, our findings reinforce those from above: Italians take their weekends more seriously than those in Country A. We also note geographic variation within both Turin and County B when it comes to weekend versus weekday consumption.

Country A

As with the full-week consumption patterns, we compare spectral and k-means clustering and find that k-means clustering with two clusters is optimal, as shown in Figure 2-13.

By far the most meters tend to consume about the same amount of electricity on weekdays as weekends. (However, if we look purely at residential customers, we notice a tendency to consume more electricity on weekdays between about 7 and 8 am, likely attributable to a habit of waking up earlier on weekdays than weekends.)

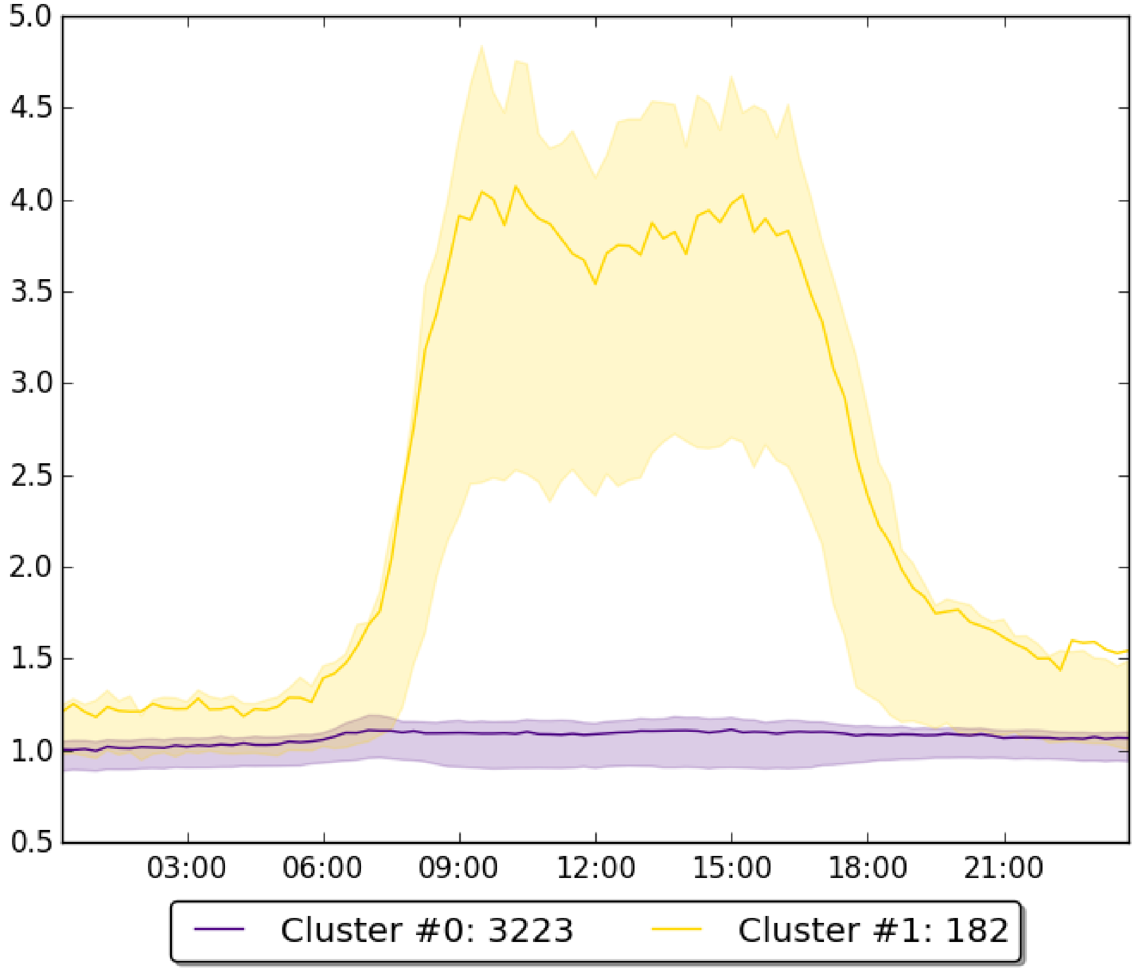
A small minority of meters consumes far more on weekdays. These meters tend to be industrial (or 'non-industrial'), particularly in logistics, chemical manufacturing, machinery manufacturing, metals manufacturing, and manufacturing more broadly. With this sense of what industries take weekends off, it is also interesting to observe that the tendency to take weekends off is largely localized to the relatively urban and industrial area surrounding the airport adjacent to the county seat.

An alternative approach is to simply look at the average ratio of electricity consumption on weekdays to that on weekends in each town. This unsurprisingly shows a similar pattern, as portrayed in Figure 2-14: areas around the airport consume more electricity on weekdays than weekends relative to the county seat or the more rural southeastern parts of the county.

Italy

A pattern very similar to that in the Country A data emerges in the Italian data: most meters consume approximately the same amount of energy on weekdays as weekends, but a small portion consumes far more on weekdays (see Figure 2-15). However, even

Figure 2-13: k-Means Clustering on Weekday-Weekend Ratio, Scaled to Sigmoid (Country A)

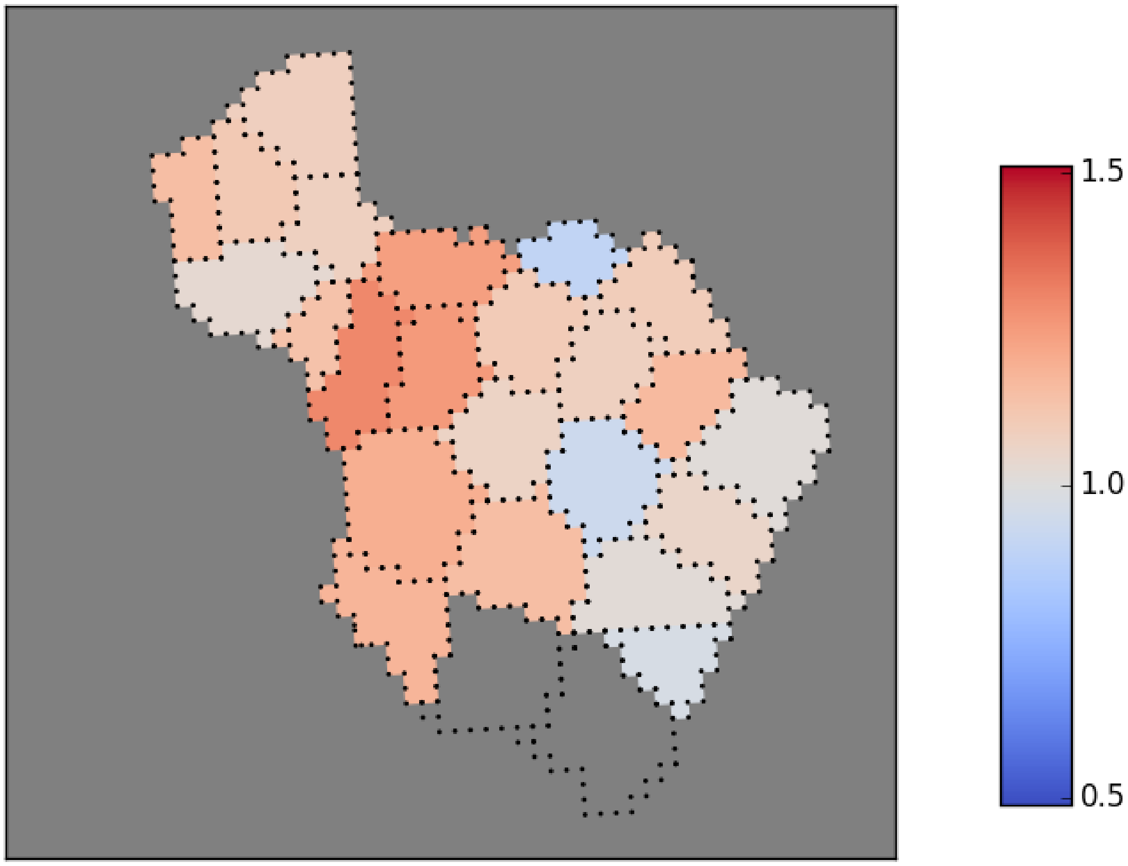


Note: Shaded areas denote the interquartile range.

the large cluster with approximately the same consumption on weekdays as weekends is different: in Italy, this cluster shows a bulge in the morning, suggesting that most users wake up later on weekends or start work later on weekends. Moreover, the smaller cluster exhibiting much higher consumption on weekdays is much larger in the Italian case–20% of our sample meters in Turin as opposed to 5% of our County B sample.

If we instead perform clustering separately for large and small consumers, we find that the main cluster for small consumers actually consumes less midday on weekdays than on weekends, consistent with a residential consumer who goes to work.

Figure 2-14: Average Weekday Consumption/Weekend Consumption, Mapped by Town (Country A)

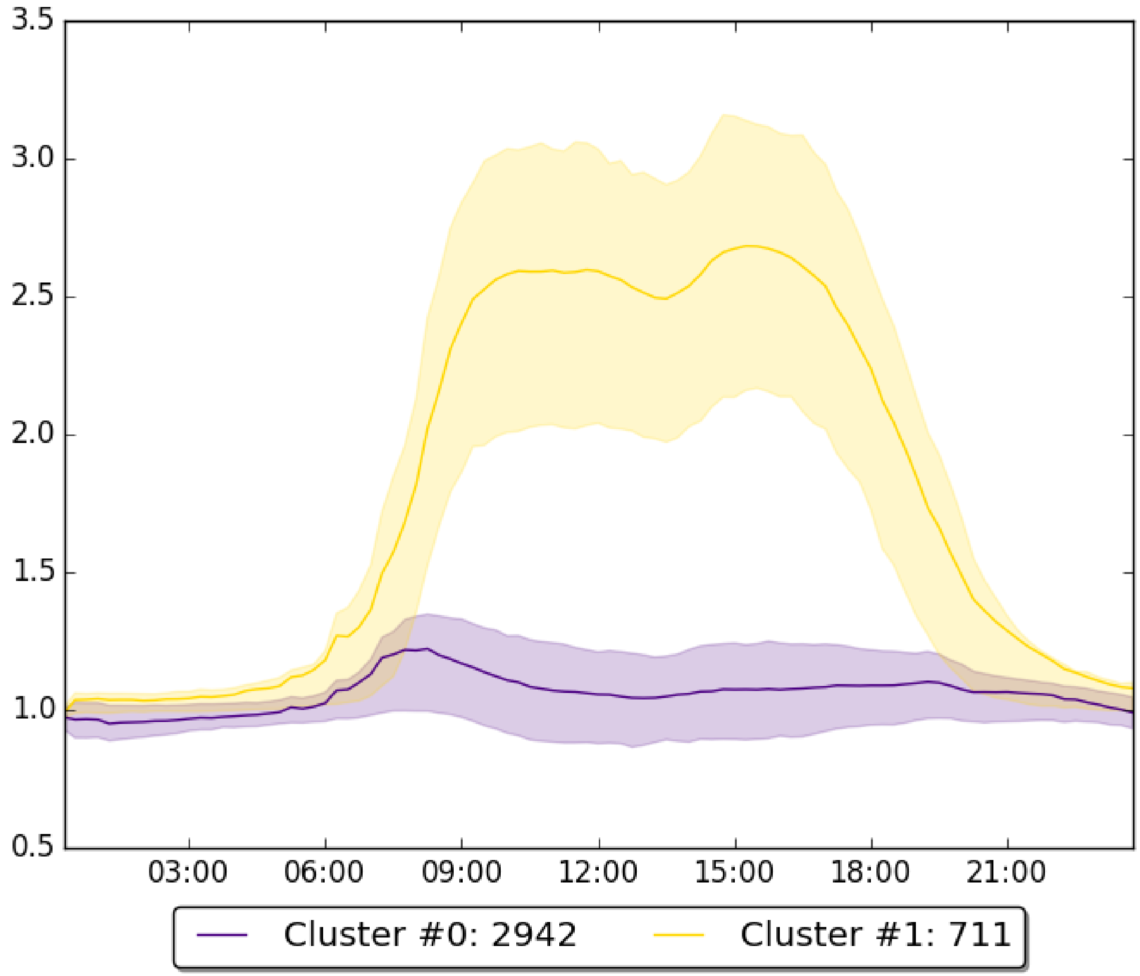


Note: Darker colors denote towns with a higher ratio.

Moreover, they have a quite pronounced increase in consumption on weekdays relative to weekends around 7 to 8 am, consistent with waking up earlier on weekdays to go to work. Looking at large consumers, both of the clusters exhibit a weekday consumption bias.

Generally, as shown in Figure 2-19, large consumers show a much stronger tendency to consume more electricity on weekdays than weekends than do small consumers. This is tempered slightly in the downtown area (and the neighborhood around the Fiat factory), where small consumers (presumably the small commercial entities noted above) consume noticeably more weekday electricity than weekend electricity. Large consumers, on the other hand, generally consume most of their electricity on weekdays, sometimes upwards of 50% more than they do on weekends. This is

Figure 2-15: k-Means Clustering on Weekday-Weekend Ratio, Scaled to Sigmoid (Italy)

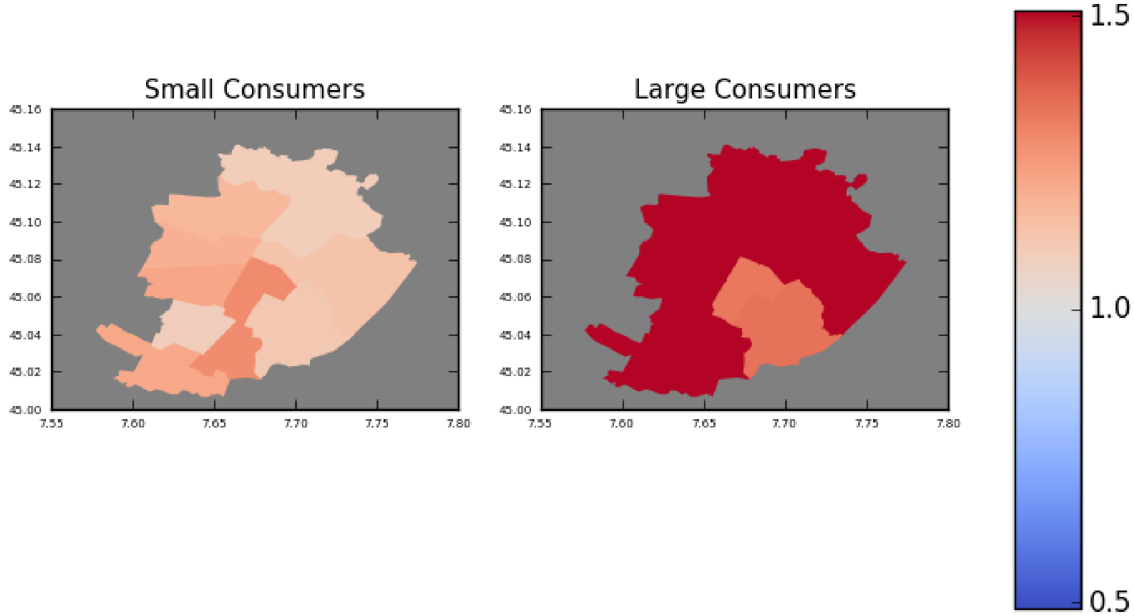


Note: Shaded areas denote the interquartile range.

particularly evident towards the north of Turin.

Weekend rest patterns are not uniform across Saturday and Sunday, as is evident from Figure 2-17. Across the city, small consumers consume about the same amount of electricity on Saturdays as Sundays, including even in the city center–surprisingly, suggesting that few small commercial enterprises that are open on Saturdays close on Sundays. Large consumers, however, consume vastly more on Saturdays than Sundays–in some suburban areas, more than twice as much.

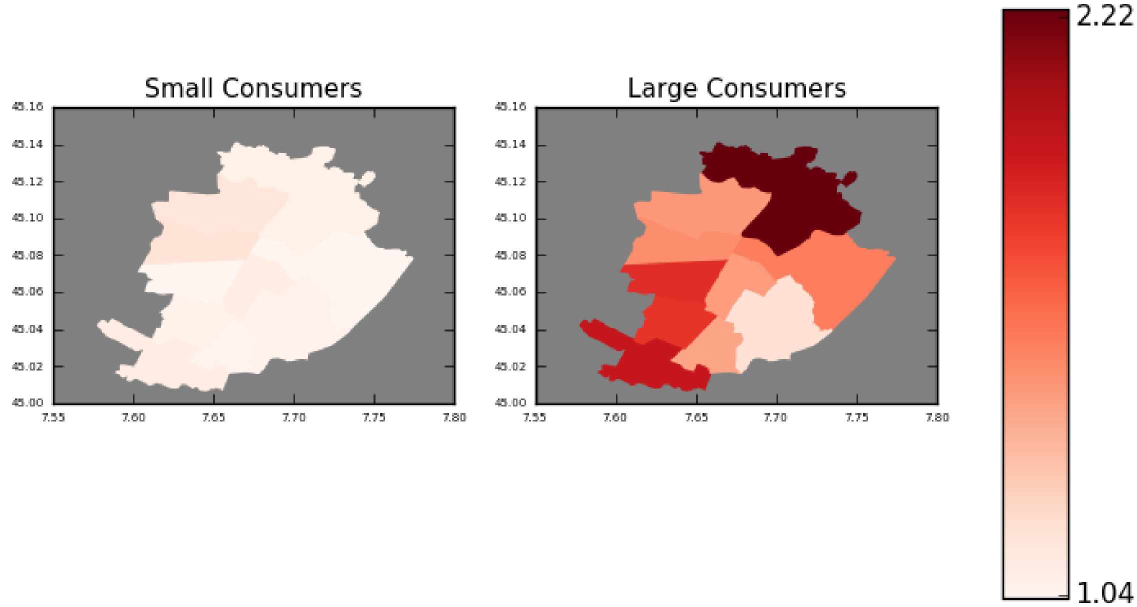
Figure 2-16: Average Weekday Consumption/Weekend Consumption, Mapped by Neighborhood (Italy)



Comparative Conclusions

Again, we find that weekends cause a much more pronounced dip in electricity consumption in Italy than Country A. This decline is particularly evident among large consumers, whose electricity consumption collapses sharply on Sunday. Although small consumers also consume slightly less energy on weekends, the difference is less pronounced.

Figure 2-17: Average Saturday/Sunday Consumption, Mapped by Neighborhood (Italy)



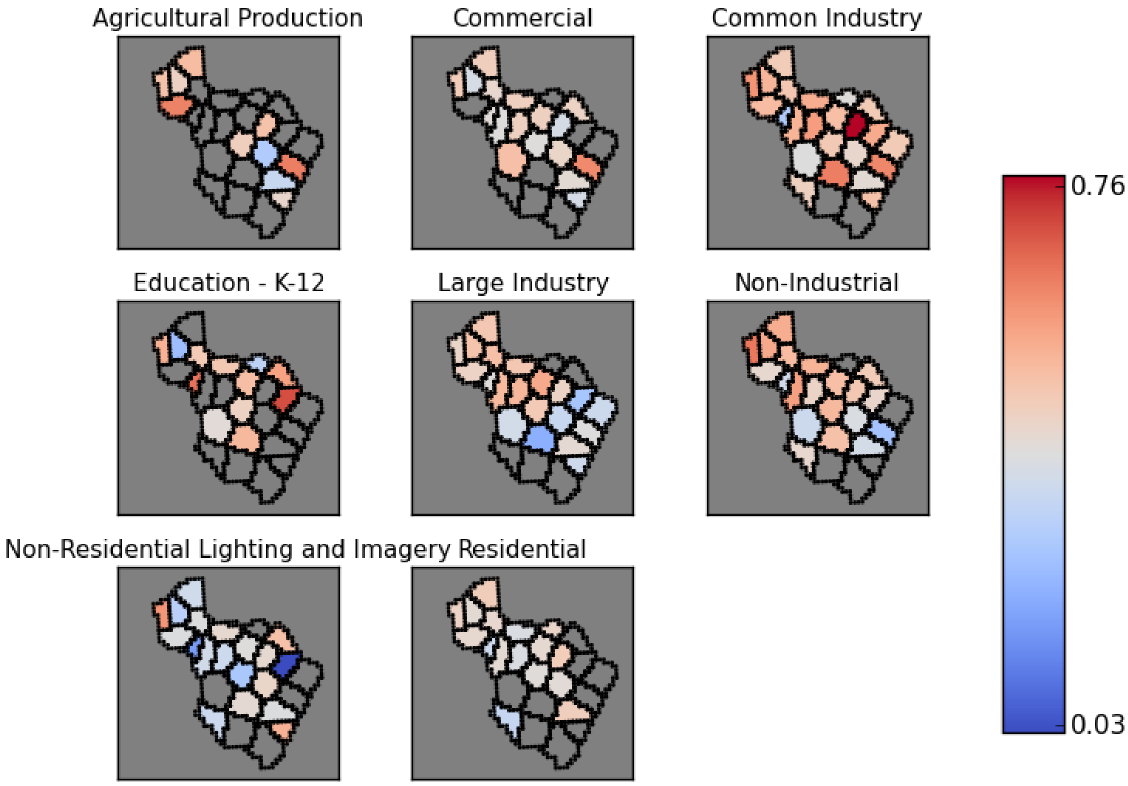
2.5 Variation across Time of Day

To explore the extent to which users maintain regular work hours, we compare weekday consumption between 8 am and 5 pm to total weekday electricity consumption. This serves as a measure of the extent to which production for industry (and household consumption) is concentrated during traditional work hours–that is, it is both a measure of the extent to which economic activity spills over into the evening and the extent to which employees' responsibilities appear to spill over into the evenings.

Country A

Unsurprisingly, the urban areas of County B, by and large, are more likely to consume a higher percentage of their electricity during these hours than more rural areas. To a slightly lesser degree, weekday evening consumption (5 to 10 pm) shows a rural bias. However, we caution that these observations can only be made with regard to the full set of users. The most rural communities are in some cases altogether missing residential customers, so we can say little about behavior patterns at much granularity.

Figure 2-18: Percentage of Total Weekday Consumption Occurring Between 8 am and 5 pm, Mapped by Town (Country A)

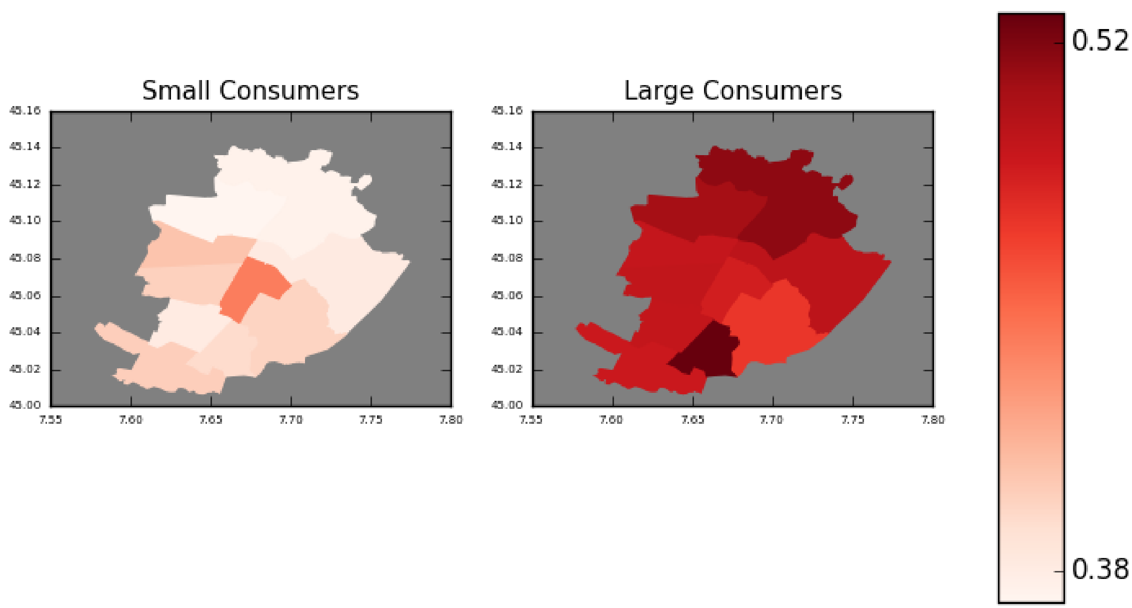


Still, it is interesting to observe variation across user categories, as in Figure 2-18. We observe first that consumption is particularly concentrated during regular work hours for Non-Industrial, Common Industry, Education, Large Industry, and Agricultural Production users, for whom 47 to 50% of weekday consumption occurs between 8 am and 5 pm. Unsurprisingly, a very low percentage of Non-Residential Lighting and Imagery consumption occurs during the daytime (<40%), as this category includes street lights. Although the pattern under which rural areas tend to consume less of their electricity during work hours holds for some of the rate categories, it is not universally true–particularly with regard to Common Industry users. Again, it bears repeating that our distribution of meters across some areas is particularly sparse and unrepresentative.

Italy

In Italy, by contrast, variation across geography and contracted power is clear. Small consumers generally consume a relatively low portion of their weekday electricity during work hours, mostly around the 38% we would expect if electricity consumption were uniformly distributed across the day. In the downtown, this is slightly less pronounced, with a relatively higher portion of weekday electricity consumption occurring between 8 am and 5 pm–again, likely the small commercial enterprises we believe to be included in our small consumer sample. Large consumers are markedly different, with about one half of weekday electricity consumption occurring during work hours.

Figure 2-19: Percentage of Total Weekday Consumption Occurring Between 8 am and 5 pm, Mapped by Neighborhood (Italy)



Comparative Conclusions

The functional distinctions between the urban core and the suburban areas of Turin is more pronounced than in County B. However, the underlying patterns are roughly the same: the urban core in both tends to consume more electricity during work hours than does the periphery–the rural areas of County B or the suburban areas of Turin. Moreover, generally, large consumers and industries are more likely to concentrate their consumption during regular work hours, although again this is much more pronounced in Italy.

2.6 Variation across Time of Year

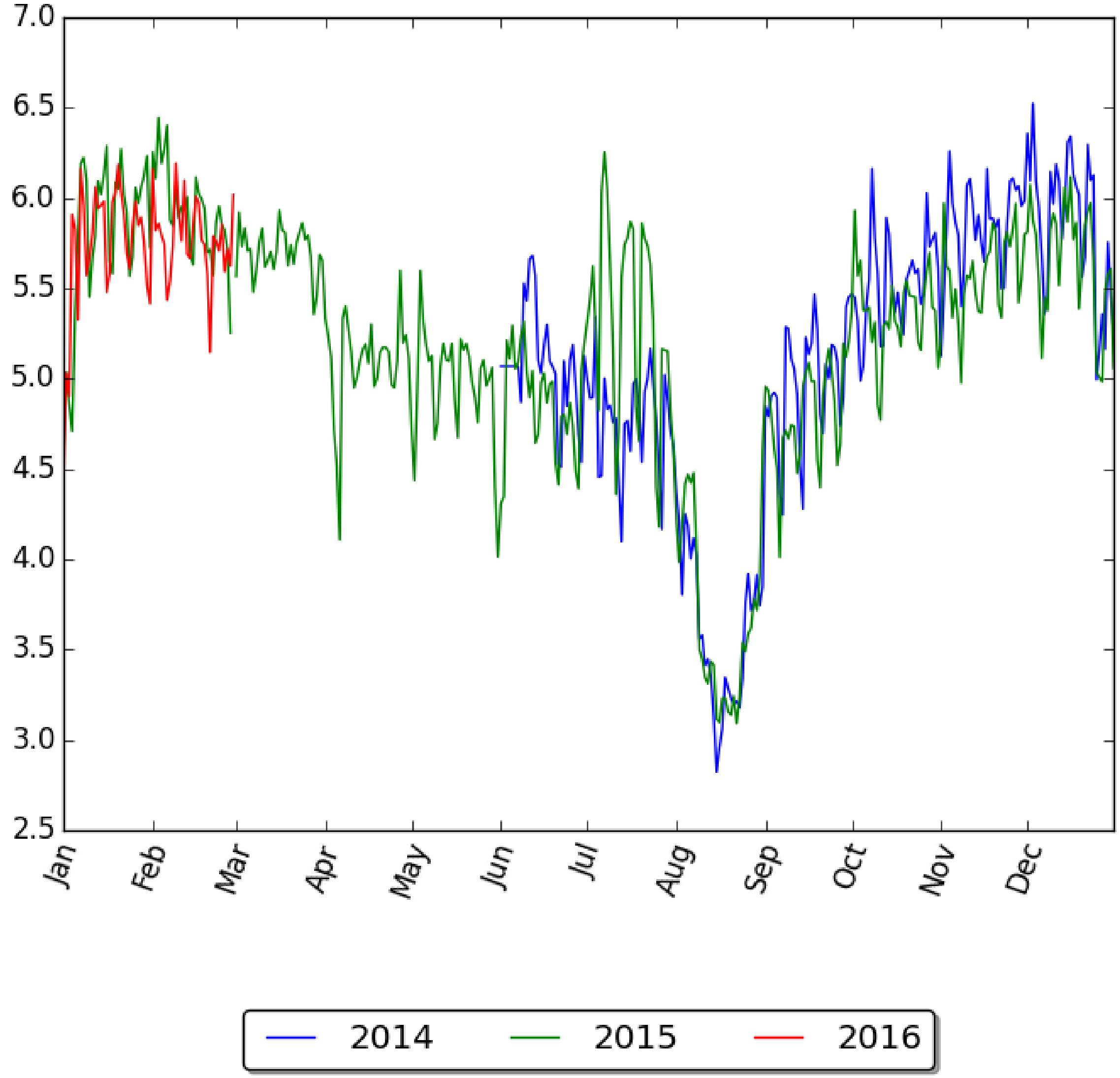
Although our focus is on smaller timescale variations in electricity consumption, it provides some insight to explore the annual cycle of electricity consumption and correlations between consumption on different days. It is immediately clear that although electricity consumption displays strong weekly periodic behaviors, it is also closely linked to time of year. Unfortunately, as our Country A data only covers several months (and very few meters are working on even a reasonably high percentage of days), we are in a better position to analyze the Italian data. Still, we include the Country A data as it displays interesting seasonal patterns. In general, the quantity of electricity consumption has long been known to be highly seasonal [11]; we add to this by showing that consumption by users is strongly correlated within but not across seasons.

Italy

For the Italian data, we treat small and large consumers separately. We begin by exploring simple trends in aggregate consumption. For small consumers, we consider those meters for which we have at least 600 days worth of observations (a relatively small subsample of less than 1000); for large consumers, we look at those for which we have at least 300 days of observations.

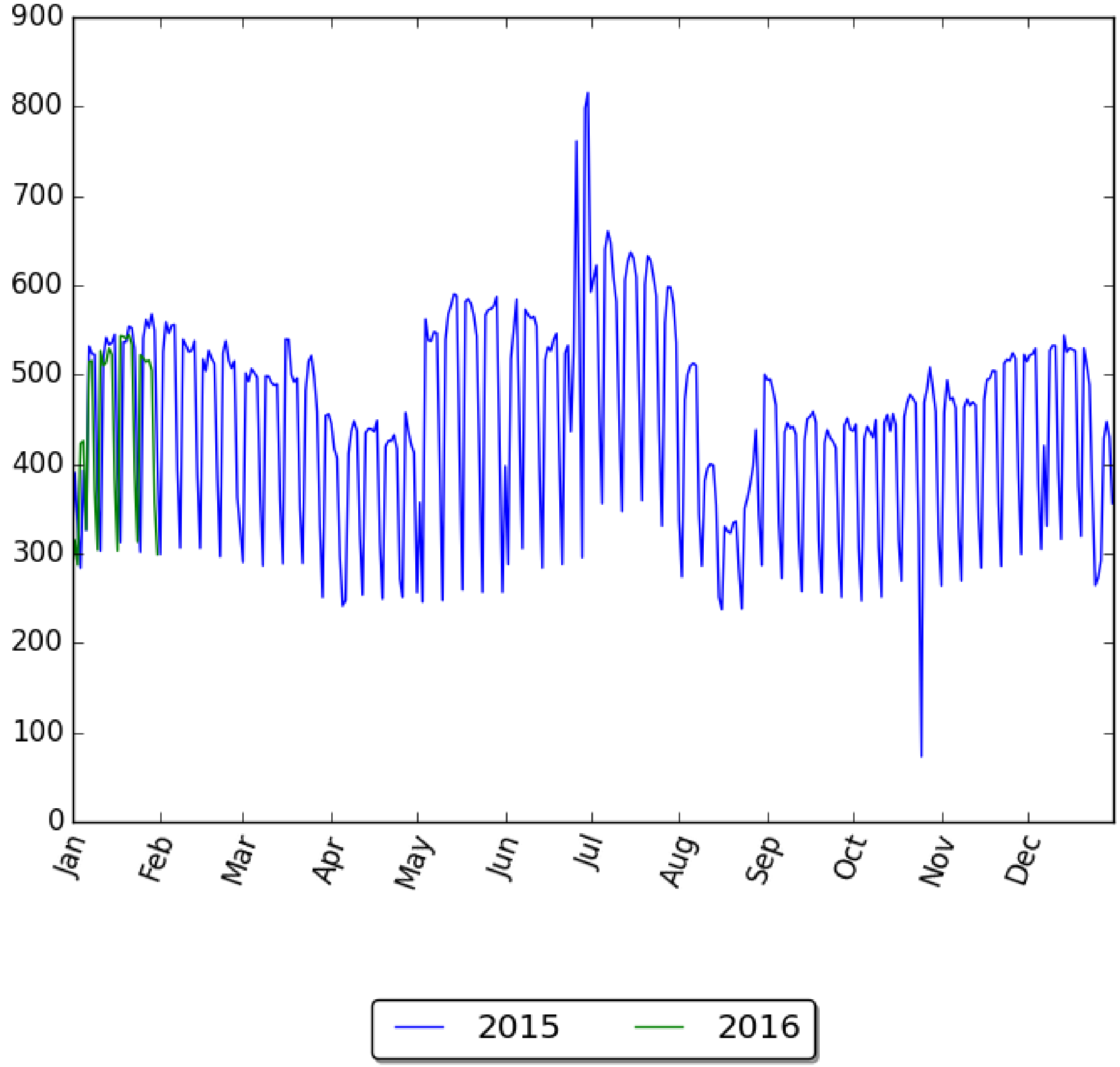
Looking at small consumers, we make the (unsurprising) discovery that electricity consumption in Turin is particularly low during August. Indeed, median weekday consumption is almost half that in July. A more subtle variation is that weekend consumption is markedly lower relative to weekday consumption during the spring and summer than during the winter, suggesting that consumers may be more inclined to take weekend vacations during the warmer months.

Figure 2-20: Daily Electricity Consumption for Small Consumers, by Year (Italy)



As we would expect, electricity consumption by large consumers is much more stable across time, and differences between weekday and weekend (particularly Sunday) consumption are extremely clear. August, however, continues to represent a significant lull in consumption, with weekday electricity consumption declining to roughly the level of weekend electricity consumption in June or July. Summer consumption, meanwhile, is noticeably higher than consumption the rest of the year, particularly around the summer solstice. Indeed, large consumer consumption jumps significantly in the first week of May, just after the May 1st holiday. (We speculate that this might

Figure 2-21: Daily Electricity Consumption for Large Consumers, by Year (Italy)



be due to turning on air conditioning.[4])

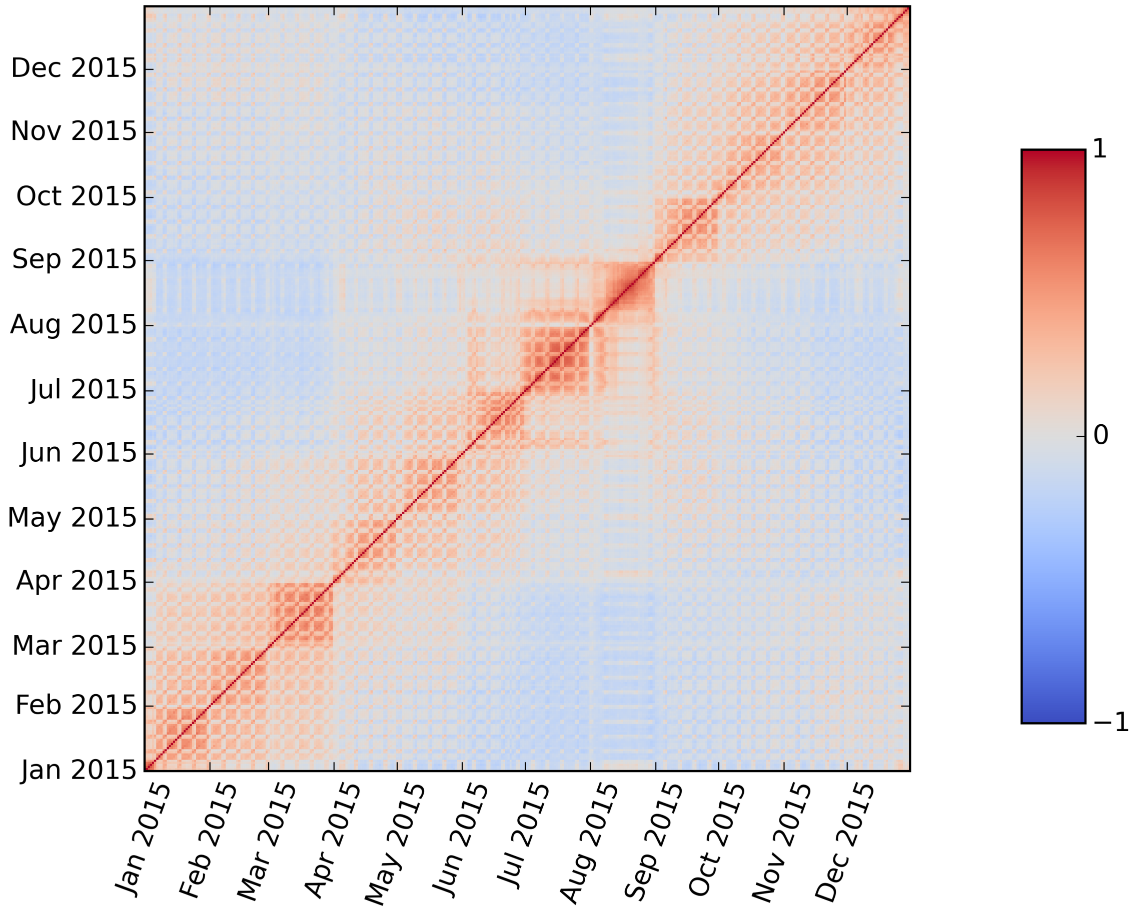
In sum, electricity consumption patterns by large consumers are much more stable across the year than by small consumers. Both, however, reduce consumption markedly in August, in contrast to consumption increases earlier in the summer.

Next, we ask if there is any correlation between electricity consumption on one day and that on another day. For each pair of days in 2015, we calculate the Pearson correlation coefficient between normalized consumption on those meters functional on both days. To normalize consumption for a given meter, we subtract from a

[4]Thanks to Umberto Fugiglando for this insight.

given day's consumption the meter's average consumption and divide by its standard deviation.

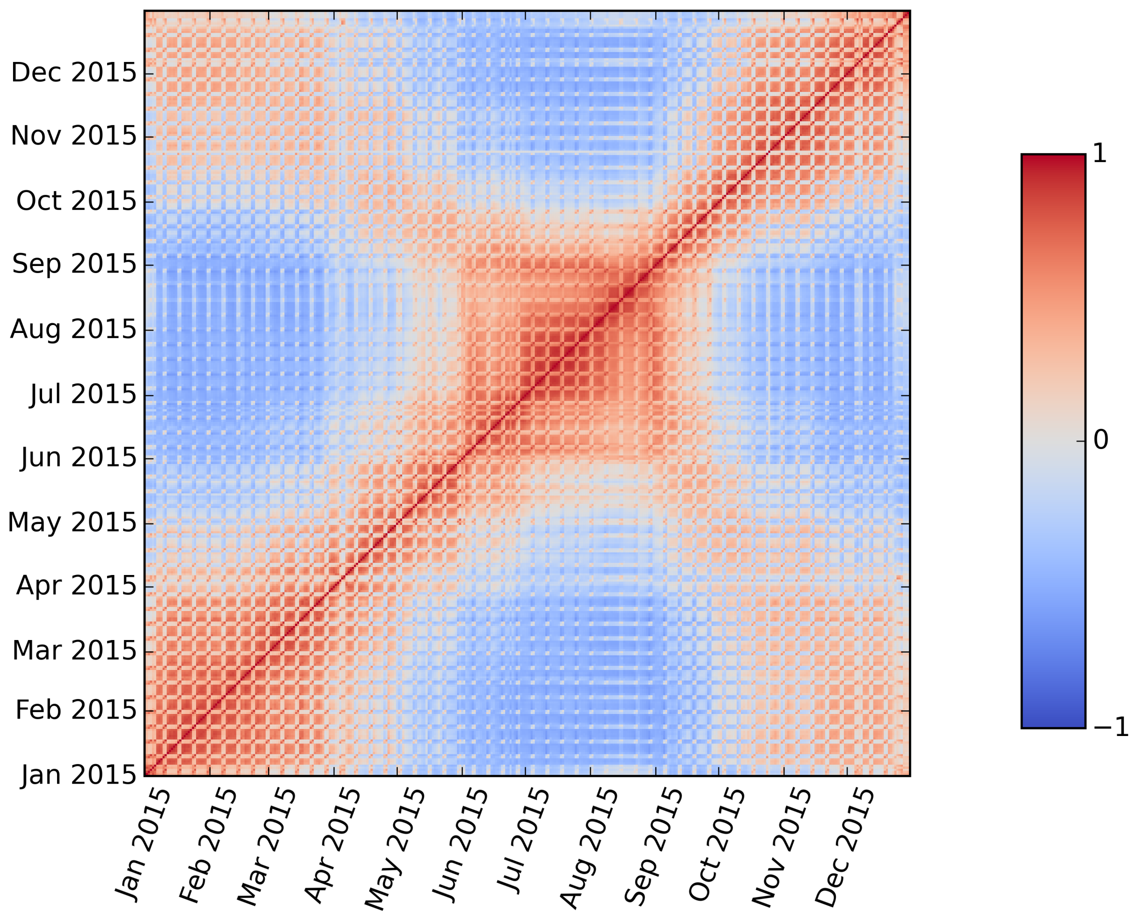
Figure 2-22: Correlations between Daily Consumption for Small Consumers, by Day (Italy)



These correlation coefficients make immediately obvious certain seasonal distinctions. For small consumers (Figure 2-22), there is relatively high correlation between consumption on weekdays between January and the end of March, between April and the end of June, between June and the end of August, and between September and the end of December. Curiously, there is little correlation between January and December consumption (except exactly at Christmas/New Year's), and indeed there is weak negative correlation between winter and summer consumption. Correlation between holidays is relatively week.

Large consumers exhibit a similar seasonal pattern (Figure 2-23), except that fall

Figure 2-23: Correlations between Daily Consumption for Large Consumers, by Day (Italy)



and spring are correlated, as are early and late winter, resulting in three distinct seasonal groupings: June through September, September through early October in addition to April through May, and mid-October through March. Moreover, correlations to proximate days are generally stronger, while correlations with more distant days are more negative. Holidays continue to only correlate with other holidays in the same season–for example, Christmas correlates highly with the Feast of the Immaculate Conception but *not* with the Feast of St. John.

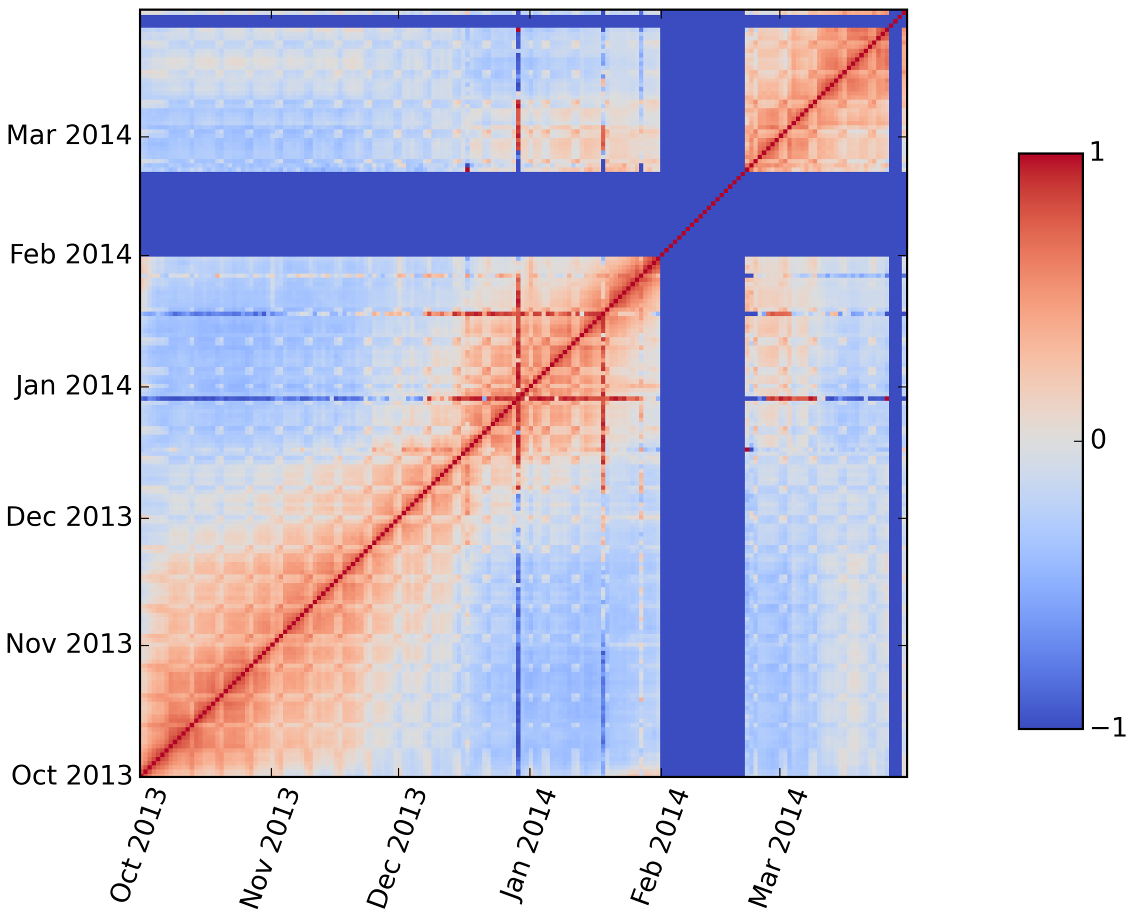
For both large and small consumers, correlations to proximate days are particularly high in July and August, and indeed, changes to consumption patterns for weekends are almost wiped out in August for small consumers.

Country A

Although we are severely restricted in any seasonal analysis of the Country A data because it covers only half a year (during which nearly an entire month of data is missing), it nevertheless displays interesting seasonal patterns.

Again, we ask how consumption between individual days correlates. We find, as in Italy, that consumption tends to correlate with proximate days, although the relatively high correlation across the winter that we observed in Italy is largely gone, perhaps aided by our lack of summer data for comparison in Country A. Although we see a noticeable weekly periodicity to our correlations, it is far weaker than in the Italian cases.

Figure 2-24: Correlations between Daily Consumption, by Day (Country A)

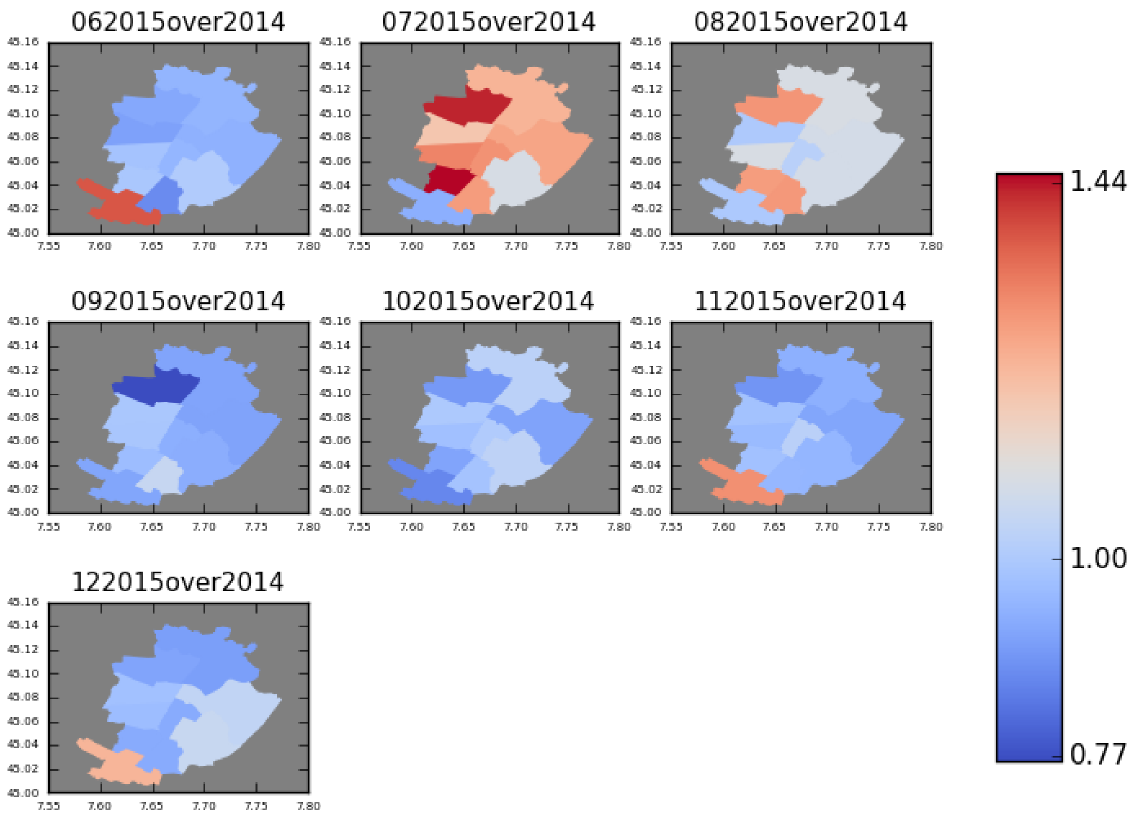


Note: Dark blue areas are missing data.

2.7 Variation across Years

Lastly, we turn to changes in electricity consumption over time. Due to seasonal effects, it is preferable to limit ourselves to year-over-year comparisons, which in turn restricts us to an analysis of the Italian data. Moreover, because our large consumer data includes only 13 months of data, we must consider large and small consumers separately. We compare only meters for which we have data in both years.

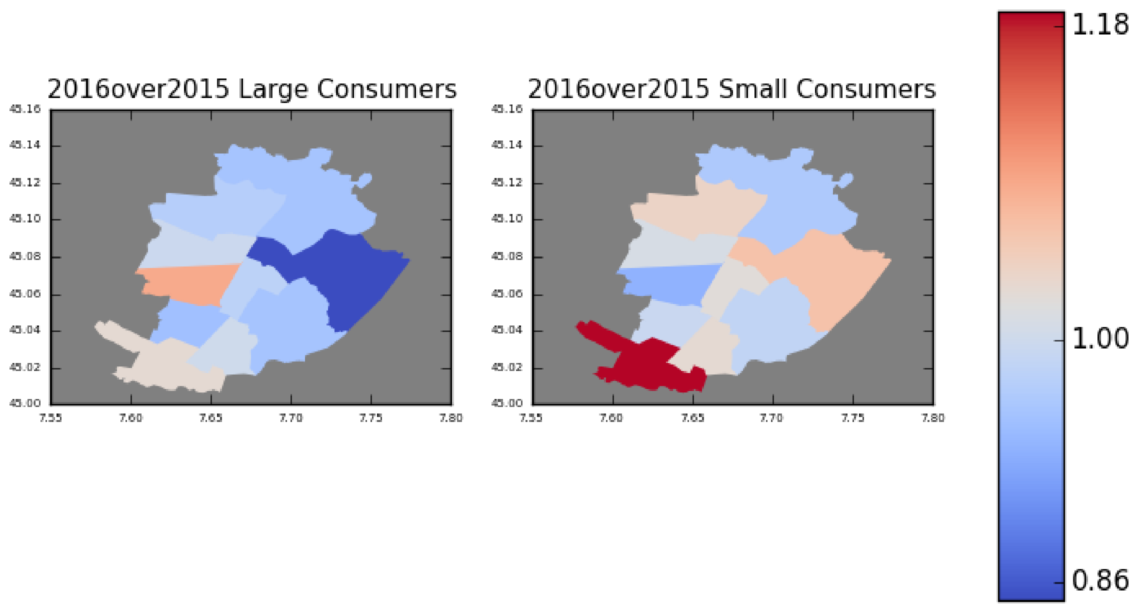
Figure 2-25: Change in Electricity Consumption by Neighborhood for Small Consumers, 2014-2015 (Italy)



Looking first at small consumers, we find that consumption changed little from 2014 to 2015 in most months. However, during the peak of the summer–August and, most particularly, July–consumption increased dramatically in 2015, likely related to unusually high temperatures in summer 2015 [25]. Meanwhile, the only neighborhood to show significant increases in consumption during any of the other months was the southernmost neighborhood, home to the Fiat plant.

For January, we are able to compare 2016 to 2015 consumption for both small and large consumers. Large consumer consumption was generally stable, except for a sharp decline to the east of the city and a sharp increase to the west of downtown. The same western neighborhood experienced a decline in small consumer electricity consumption, while the same eastern neighborhood experienced an increase in small consumer electricity consumption, suggesting an inverse relationship between changes in small and large consumer electricity consumption. For both small and large consumers, the southern neighborhood with the Fiat plant continued to show impressive increases in electricity consumption.

Figure 2-26: Change in Electricity Consumption by Neighborhood, January 2015-2016 (Italy)



In sum, it is hard to draw many conclusions about changes in electricity consumption over time from the limited data we work with here. However, it is evident that there is significant variation across neighborhoods within Turin, and moreover that there is nontrivial variation in change from one month to the next. Combined, this should encourage us to look closely at sub-municipal data on economic growth.

Chapter 3

Can Electricity Consumption Classify Land Uses?

This chapter asks whether we can classify land use using only electricity consumption patterns. That is, we ask whether different categories of land use are associated with distinctive and mutually exclusive electricity consumption patterns. Given the potential scale of electricity smart meter data, if possible, this might permit an efficient mapping of land use characteristics across large swathes of space–a tool that would be particularly useful in places where land use is poorly recorded or records are not publicly available.

Ultimately, we are not particularly successful at classifying land use by electricity consumption data. The findings of the previous chapter indeed foreshadow this result: when we attempt clustering within each user category or industry, we find that residential and industrial users are notably distinctive, but that certain sets of user categories share similar clusters. This severely limits the prospect of land use classification using these consumption curves: if, for a given consumption pattern, a meter could plausibly be common industry or large industry, it is hard to use the consumption pattern alone to classify the meter. Indeed, this is hardly surprising: the rate categories are not set to differentiate by consumption patterns so much as by economic function, and indeed the boundaries between the rate categories appear relatively confusing.

3.1 Literature Review

Land use classification is typically approached using satellite data. This can attain quite high accuracy rates, but is limited in the categories that can be identified: what is going on inside a roofed structure can be harder to detect. Indeed, land use classification using satellite data has been very successful at differentiating different types of vegetation from generic built up areas, with accuracy rates sometimes exceeding 80% [4, 8, 9]. Some more ambitious studies have approached the various uses of built-up land. Chen et al. were able to differentiate single-family residences, multi-family residences, industrial/commercial areas, etc., with 69% accuracy [?], and Lu et al. were able to classify into six categories that separated residential and commercial/industrial land uses with 82% accuracy [18]. Still, both these studies were undertaken in suburban areas, raising questions about their applicability to dense urban environments with mixed-use neighborhoods.

Several other more novel approaches to land use classification have been attempted. Pan et al. were extremely successful using taxi pick-up/drop-off GPS data to classify downtown land use in neighborhoods with high taxi usage and a single land use, correctly classifying 95% of these carefully chosen areas [24]. Considering instead whole cities, Reades et al. show patterns of cell phone usage vary between different functional areas of cities [27]. Developing on this idea, Toole et al. were able to classify 54% of land use using random trees to develop a model based upon cell phone records; excluding residential uses, they could correctly forecast 40% of land use [29]. This study attempts instead to use electric meters to classify land use.

3.2 Data

This chapter exclusively uses data from the Country A sample. This dataset includes details about the rate category of users, which we treat as their "ground truth" land use category. (The Italian data does not include any specifications regarding the user beyond their contracted power.) There are risks to using rate categories as "ground truth:" given large differences in electricity tariffs across different rate categories, it is entirely plausible that certain users are miscategorized. Nonetheless, we operate on the assumption that such miscategorization is not unduly rampant.

In classifying meters, we again use average weekly consumption curves adjusted to fit a sigmoid cumulative probability function, but also consider daily consumption data for each meter. Finally, we use the k-means clustering means for each rate category developed in our analysis of electricity use patterns. We consider only those rate categories for which there are over 200 meters in our sample: Residential, Large Industry, Non-Industrial, Common Industry, Non-Residential Lighting & Imagery, and Commercial; as well, we run our model on a simple binary choice of Industrial (Common Industry and Large Industry, combined) and Residential meters. We further restrict ourselves to those meters with average consumption of at least 0.01 kWh per fifteen minute interval.

In developing our classification model, we also used a measure of the smoothness of electricity consumption, where the smoothness $s_m(t)$ for meter m at time t in minutes is defined as:

$$s_m(t) = \frac{\sum_{k=0}^{3} |c_m(t + (k+1) * 15) - c_m(t + k)|}{c_m(t + 4 * 15) - c_m(t)}$$

where $c_m(t)$ is the consumption of meter m at time t. To facilitate computations, we treat missing values (including cases where the numerator was zero) as zero, and adjust infinite values to a very high number (10^6).

3.3 Classification

Having explored the different consumption patterns characteristic of different rate categories and shown that they do differ in some regards, we turn to classifying meters. Our technical aim is to use the information in a meter's consumption pattern to identify its rate category. More broadly, we are seeking, on the one hand, to create a method for classifying land use and, on the other hand, to utilize our errors to identify meters that may have been assigned the wrong rate category by the utility.

3.3.1 Methods

We divide our data in half, making training and testing subsets. We seek to use the training data to predict the rate category of the testing meters by comparing the consumption pattern of each meter to the consumption patterns of the (labeled) meters in the training data. While it would be ideal to consider consumption patterns over the full study period, it appears that no meters were functioning for all days of the study period. Further work would be necessary to adjust for missing data, and so in the meantime we have focused on three other regimes:

1. A vector of average consumption for each fifteen minute interval in the week, mapped to conform to a sigmoid cumulative distribution function

2. The collection of daily consumption vectors

3. A week-long vector of "smoothness" as defined in the Data section

For (1) and (3), we used a combination of these five methods to classify meters:

- *Nearest Neighbors*, which associates to each vector the modal class of the nearest 45 vectors in the training set, weighted by distance. (This method prefers to cluster, for example, all very large consumers together. While this may be accurate in many cases, magnitude of consumption may overwhelm temporal patterns for these consumers, which risks classifying multiple households on a single meter as industrial, for example.)

- *Dot Product*, which associates to each vector the modal class of the 45 vectors with which its dot product is greatest. (For this method, vectors are normalized to unit vectors, from which we lose information about the magnitude of consumption but which allows us to compare consumption curves focusing solely on what time of day consumption was largest and the temporal variation in consumption.)

- *Decision Trees after PCA*, which finds an optimal tree of binary decisions, each with respect to one variable, to classify the training data, then applies it to the testing data. As one variable (one fifteen minute interval) has little special meaning, we first conducted PCA on non-outlier data reducing our data to seven dimensions, and then performed decision trees on the first several coefficients. Although we present here the optimal outcome after considering several parameters limiting tree complexity to avoid over-fitting the model, this model is the most complicated and hence least elegant. (Due to computational concerns, we were unable to implement this on (3).)

- *Support Vector Machines (SVM)*, using a Gaussian radial basis function kernel. This transforms the data into an infinite dimensional Hilbert space, then finds a hyperplane that classifies the data. (Due to computational concerns, we were unable to implement this on (3).)

- *Dot Product on k-means*, for which we applied the dot product method described above (albeit considering only one element with the highest dot product), but with the intra-rate category k-means clustering averages from the Clustering on Consumption Curves section as the training set. This produces a particularly elegant model as it minimizes the size of the training set. (We applied this only to (1).)

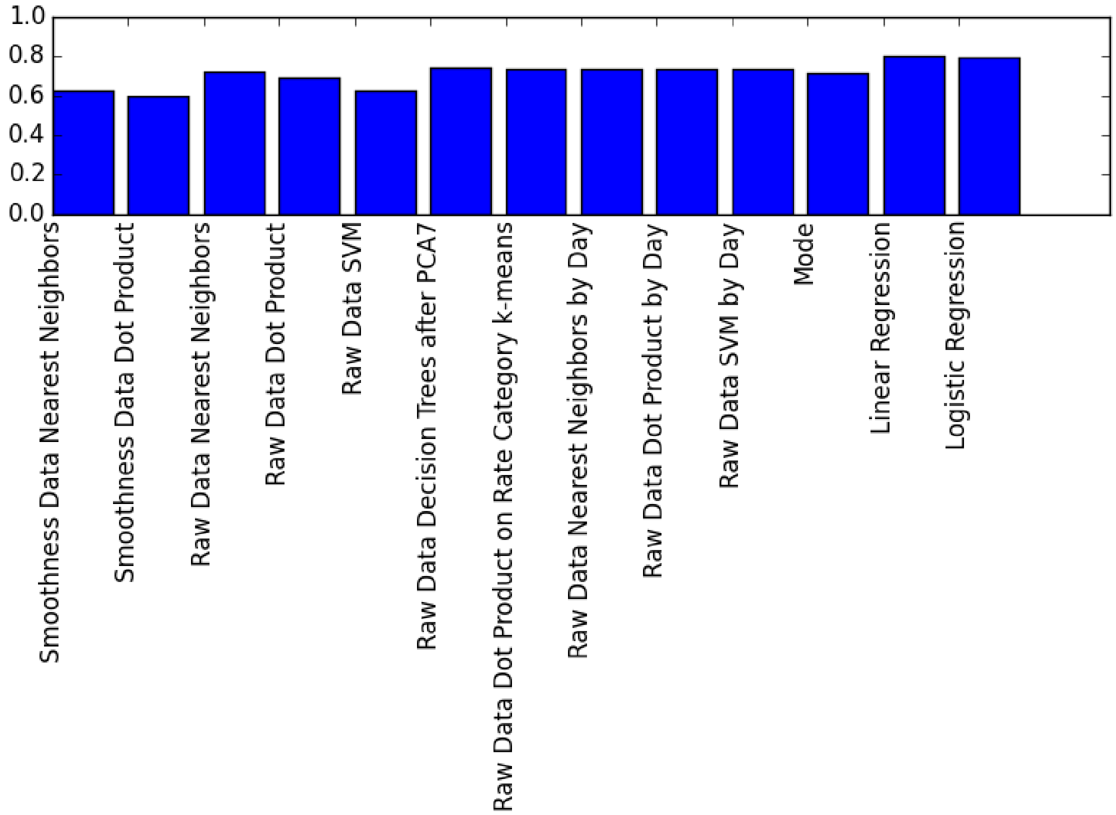
For (2), we compare consumption curves from each day to training consumption curves from the same day, assigning an estimated classification using nearest neighbors (see above). A predicted classification for the meter is then the modal classification across all days. This allows us to take into account potential seasonal variation that could be diminished or lost in the averages considered in (1).

Finally, we constructed three models to combine the above approaches. One simply took the mode of the results from the various methods. For the binary classification between industrial and residential categories, we were able as well to construct linear and logistic regression models (again, dividing our sample in two to form testing and training subsets).

3.3.2 Results

The results, shown in Figures 3-1 and 3-2, were remarkably uniform across various methods. All methods performed substantially better than we would expect from randomly assigning labels. However, variation between methods was quite small.

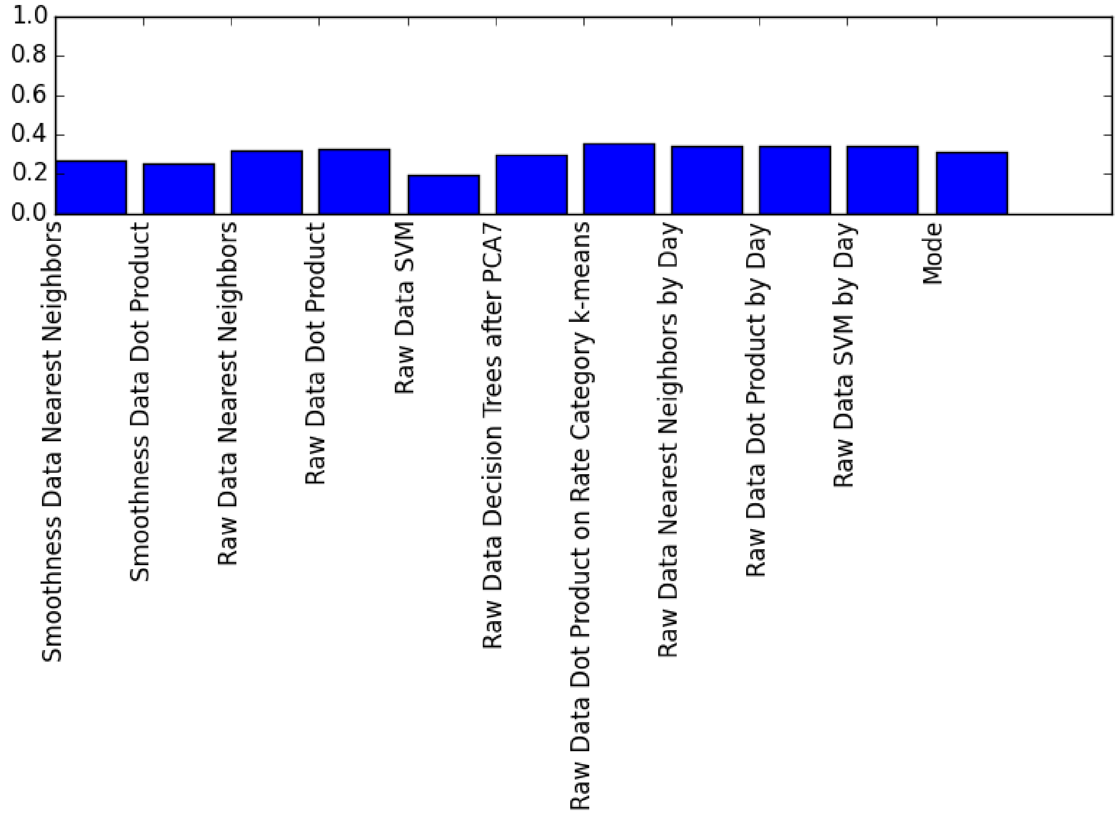
Figure 3-1: Binary Classification Success Rates for Industry and Residential



When applied only to data from the six largest categories, all methods classified between 20% and 35% of meters correctly. If the two industrial categories were combined and compared only with the residential category, all methods estimated between 50% and 75% of meters correctly. Generally, the methods using the smoothness data performed the most weakly, and the SVM method performed quite well. In the binary classification case, the linear and logistic regression models both improved on all the models they took as inputs, yielding around an 80% success rate.

Conducting the same classification on data aggregated to hourly consumption might have smoothed out meaningless dips and jumps in consumption and hence

Figure 3-2: Classification Success Rates for all Large Categories



reduced error, but failed to produce better results.

It is worth noting, however, that slightly better success was attained when we considered only meters where many models concurred in their classification. This method, in particular, suggests a useful way to identify meters that might have been assigned to the wrong rate category.

3.4 Conclusion

In our attempts to use consumption patterns to classify meters into rate categories that were set administratively without direct regard for electricity use patterns, we were far more successful considering only industrial and residential meters than considering a broader collection of categories. On this binary classification question, we were able to correctly classify over 80% of meters using a combination of models.

This suggests that classification is at least feasible to identify outlier meters (which may have been assigned to the wrong rate category), but that significantly more work would need to be done to use electricity consumption patterns for land use classification more broadly. Indeed, it may reinforce the conclusion some have reached about land use classifications: that the current regime of land use classification is in many cases unduly rigid and simply unrelated to the behavior patterns relevant to certain sectors–in this case, energy consumption [5].

Chapter 4

Holidays and Time Changes

In using electricity consumption data to explore local culture and behavior, it can be particularly useful to explore the impact of special occasions on consumption habits. We focus primarily on holidays, but also briefly inquire about the impact of daylight savings time in Italy. We ask two questions: How does the urban system respond to these stimuli? And how long does it take the urban system to return to its normal behavior after a holiday or time change?

We find that large consumers/industry tend to be more responsive to these events, particularly holidays, but are capable of rebounding quite quickly to normal consumption patterns. Small consumers or residential consumers' consumption patterns are sometimes altered, but less frequently and generally to a smaller extent.

4.1 Holidays

Holidays have a significant impact on electricity consumption, most notably through vacations which close enterprises and drive their consumption down. Meanwhile, residential consumption can also be impacted by vacation travel or increased time at home. Generally, we find that holidays lead to much lower electricity consumption, with the greatest decline for large consumers/industrial consumers, for whom closing an entire enterprise dramatically lowers consumption.

Public holidays and particularly the summer school vacation are strongly associated with patterns of tourism and vacationing (e.g., [2]), which explain declines in small consumer or residential consumption. Not only do these officially established holidays vary from place to place, but so do vacationing patterns vary across cultures—as research on French and English Canadians has shown [28]. We show that electricity consumption not only is strongly correlated with these holidays but, particularly as noted above with respect to weekends, may also but be associated with culturally specific behavior patterns.

Our analysis of holidays necessarily focuses primarily on Italy, as Country A celebrates few holidays during our study period. Both Christmas and New Year are internationally celebrated, and hence we explore those in comparison with Italy.

4.1.1 Annual Pattern of Holiday Recognition

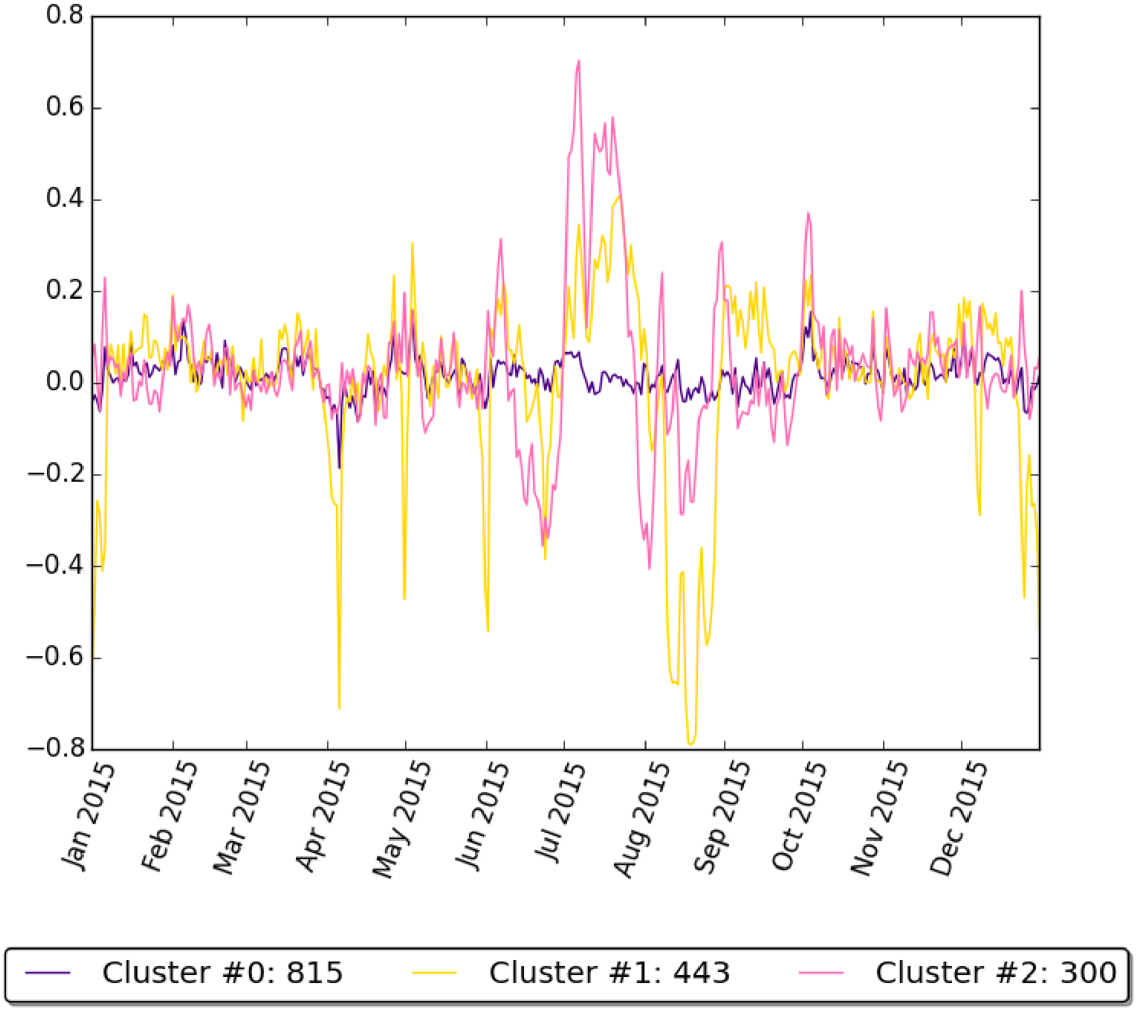
Before delving into the details of individual holidays, we start by looking at overall 2015 patterns in holiday recognition by small and large consumers in Turin. We find that many consumers do not adjust their behavior at all in light of holidays, while others tend to honor every holiday.

We approach this question by conducting k-means clustering on a simplified measure of daily electricity consumption that assigns to each meter-day a 1 or -1 if consumption is more than 20% above or below, respectively, the 2 month running average of consumption for a given meter. Otherwise (or if data is missing for the given day), a 0 is assigned to the meter. Meters are excluded if over 5% of 2015 days are missing. The 20% tolerance was arrived at by experimentation and selecting the lowest tolerance to result in interpretable results. As well, the number of clusters was arrived at by considering the interpretability of the results.

For small consumers (Figure 4-1), we select three clusters, the largest of which exhibits roughly constant consumption throughout the year, with no strong responses to holidays or temperature changes. This cluster is most dominant in the industrial suburbs of Turin. Cluster #1, the second largest, represents the largest consumption dips associated with each major holiday: Epiphany, Easter, Labor Day, Republic Day, the Feast of St. John, the Feast of the Immaculate Conception, Christmas, and New

Year's. It is clear that, among these, Easter and August incur the most widespread declines in consumption. In addition, a significant decline in consumption is evident in August 2015. This is particularly prevalent in downtown and residential areas. Cluster #2, the smallest, is largely constant, but involves a significant increase in consumption in July, when Turin was abnormally hot in 2015.

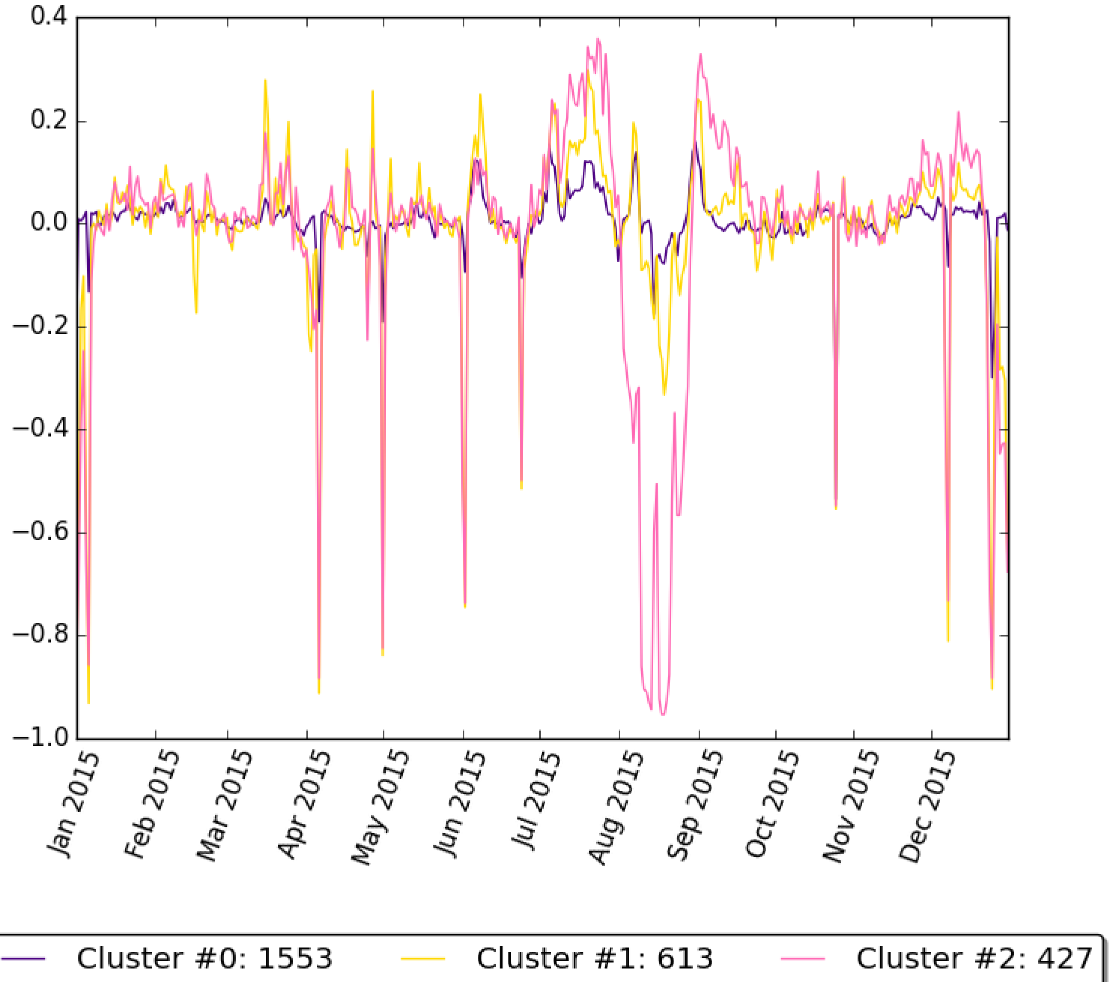
Figure 4-1: k-means Clustering on Days with Abnormally High or Low Consumption, Small Consumers (Italy)



Large consumers exhibit a quantitatively similar but geographically reversed pattern (see Figure 4-2). The largest cluster, Cluster #0, displays roughly constant consumption year-round, albeit with noticeable but small declines on holidays, suggesting that some members of this cluster reduce consumption on some holidays.

Clusters #1 and #2 both show strong declines in consumption on holidays, with near universal participation in all major holidays (except the Feast of St. John). The two clusters diverge primarily in that Cluster #1 involves only a small dip in August consumption, whereas Cluster #2 includes a mammoth decline, suggesting almost all members of this cluster reduced consumption by over 20% in August 2015. Unlike for small consumers, the holidays were recognized more in the suburban areas than in the downtown–that is, Cluster #0 was most prevalent downtown, and Clusters #1 and #2 were more prevalent but not predominant in the suburbs.

Figure 4-2: k-means Clustering on Days with Abnormally High or Low Consumption, Large Consumers (Italy)



We conclude that declines in consumption for holidays are highly correlated. Consumers who reduce consumption for some major holidays are likely to do it for all or at least most major holidays. Most consumers, however, exhibit remarkably steady consumption throughout the year. The geographic distribution of these consumption patterns is reversed for small and large consumers; the former acknowledge holidays with consumption declines more in the downtown area, the latter more in the suburbs.

4.1.2 Internationally Important Holidays

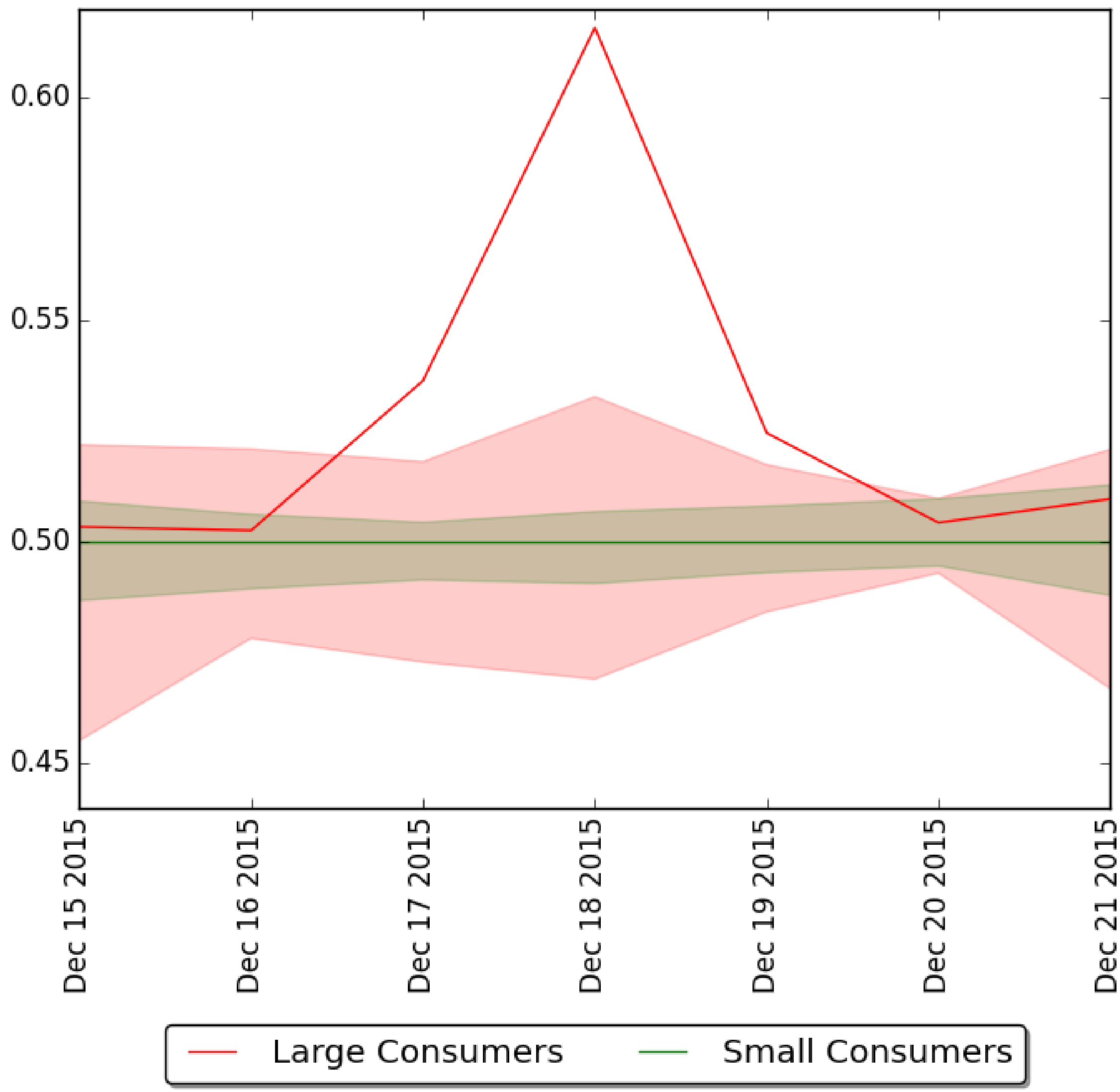
We compare the impact on consumption of two holidays across both Country A and Italy, providing a window into the different responses of the two societies to similar stimuli. We find that New Year's leads to sharp consumption dips among large consumers/industry in both societies, but not among smaller consumers/residential customers. Christmas leads to a drop in consumption in Italy and an *increase* in Country A.

Christmas

Country A and Italian observations of Christmas are radically different. Unfortunately, the proximity of Christmas to New Year's complicates our analysis. Thus although we generally compare a week with a holiday to the next week, in exploring Christmas we compare the week before Christmas to the week of Christmas.

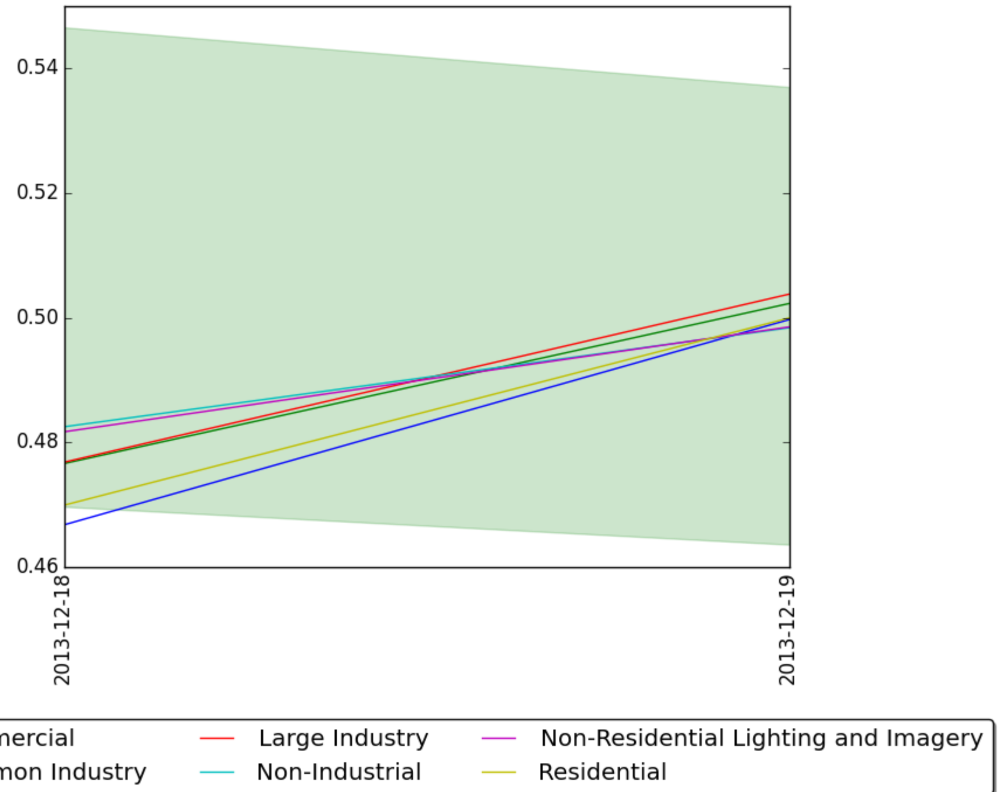
In Italy, we find that small consumer consumption the week before Christmas is almost identical to that the week of Christmas, including on Christmas day itself (Figure 4-3). Looking more closely, we note that although the vast majority of small consumers consume about the same amount the week of Christmas as the week before, a small subset, concentrated downtown, reduces consumption markedly either on Christmas day itself or for the entire week. For large consumers, on the other hand, consumption the week before Christmas is substantially higher than on the week of Christmas. This is particularly true for Christmas day itself: large consumer consumption 7 days before Christmas day is over 60% higher than on Christmas day itself. Large consumer consumption declines are particularly prevalent in the suburbs, and sometimes extend for a day or two before or after Christmas itself.

Figure 4-3: Median Ratio of Electricity Consumption by Contracted Power (Italy), Christmas Time: Previous Week/(Previous Week + 22 to 28 December 2015)



Note: Shaded areas denote 5th to 95th percentile.

Figure 4-4: Median Ratio of Electricity Consumption by User Type (Country A), Christmas Time: (25 to 26 December 2013)/(Previous Week + 25 to 26 December 2013)

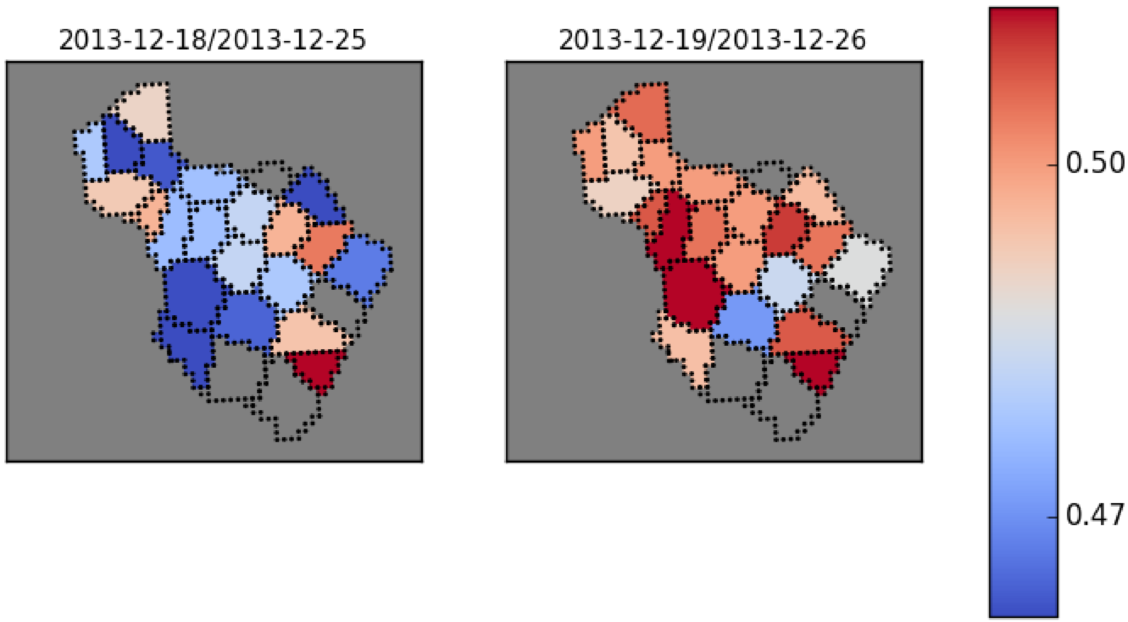


Our data from Country A is much more sporadic, so we are only able to compare Christmas Day and 26 December (Boxing Day, if you prefer) to consumption seven days earlier (Figure 4-4). Whereas consumption on Boxing Day is almost identical to that a week earlier, consumption on Christmas Day is noticeably higher than that a week earlier, by about 10-15% depending on industry. This pattern holds true not only for various industries but also for residential areas, educational institutions, and so forth.[1]

Increased consumption on Christmas was not uniformly distributed across County B. It was particularly pronounced in the county seat and in the central north-south axis of the county, with almost no change in the airport area or the most rural fringe

[1] It is worth noting that there were no particularly unusual weather events these four days. Christmas was slightly chillier than Boxing Day or 18 December, but 19 December was in fact even chillier relative to both Christmas Day and Boxing Day.

Figure 4-5: Map of Ratio of Electricity Consumption by Town (Country A), Christmas Time: (25 to 26 December 2013)/(Previous Week + 25 to 26 December 2013)



of the county.

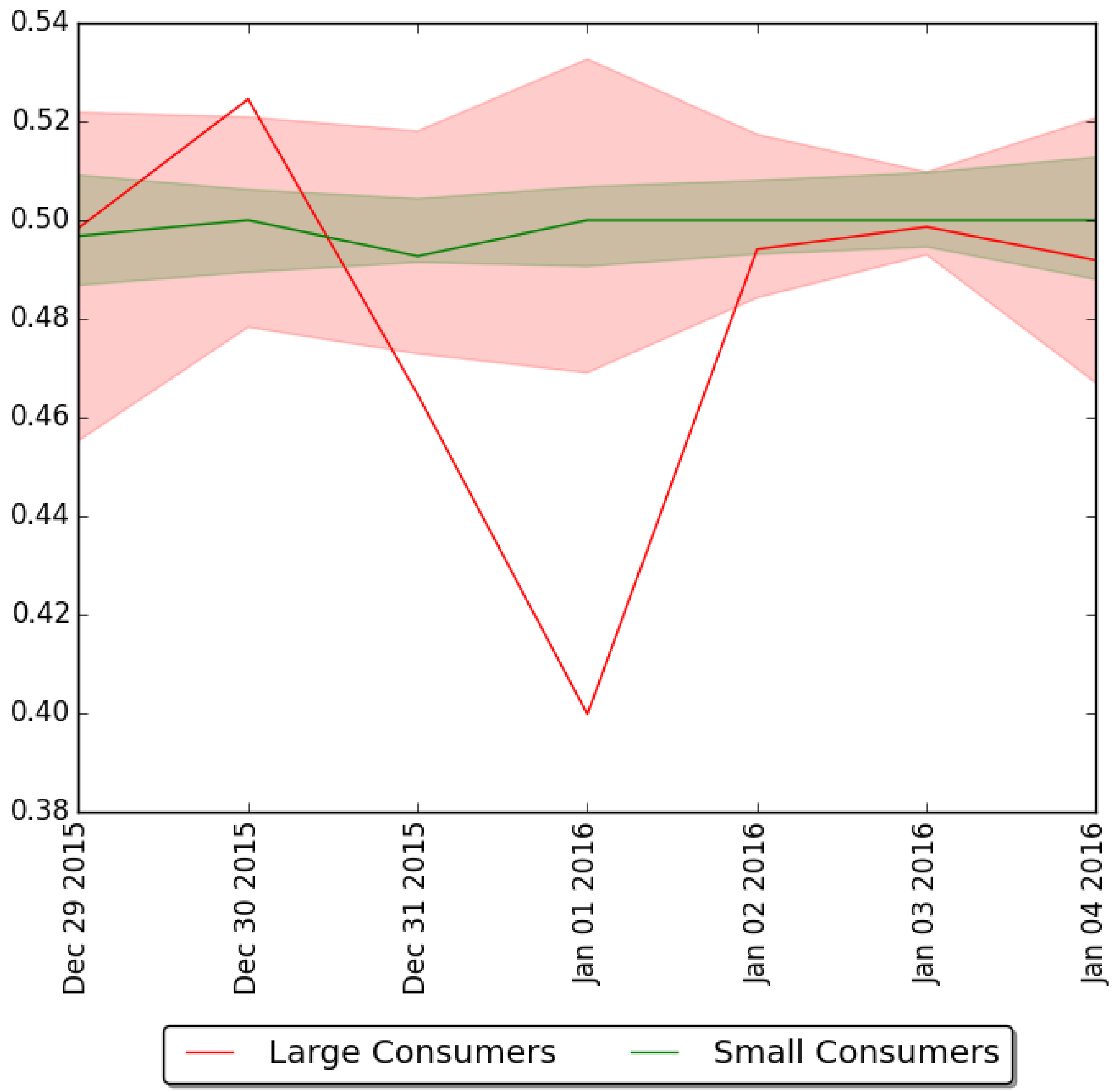
Although it is hard to say much about Christmas observations that we did not already know from anecdotal observations, this data confirms that Italian enterprises take Christmas quite seriously, dramatically cutting back operations, whereas Country A 'celebrations' of Christmas lead to more consumption and no evident decline in economic activity as proxied by electricity consumption.

New Year

New Year (January 1) is officially observed in both countries. Moreover, responses to New Year are quite similar in both countries: work places consume less electricity, while living places' electric consumption is not impacted.

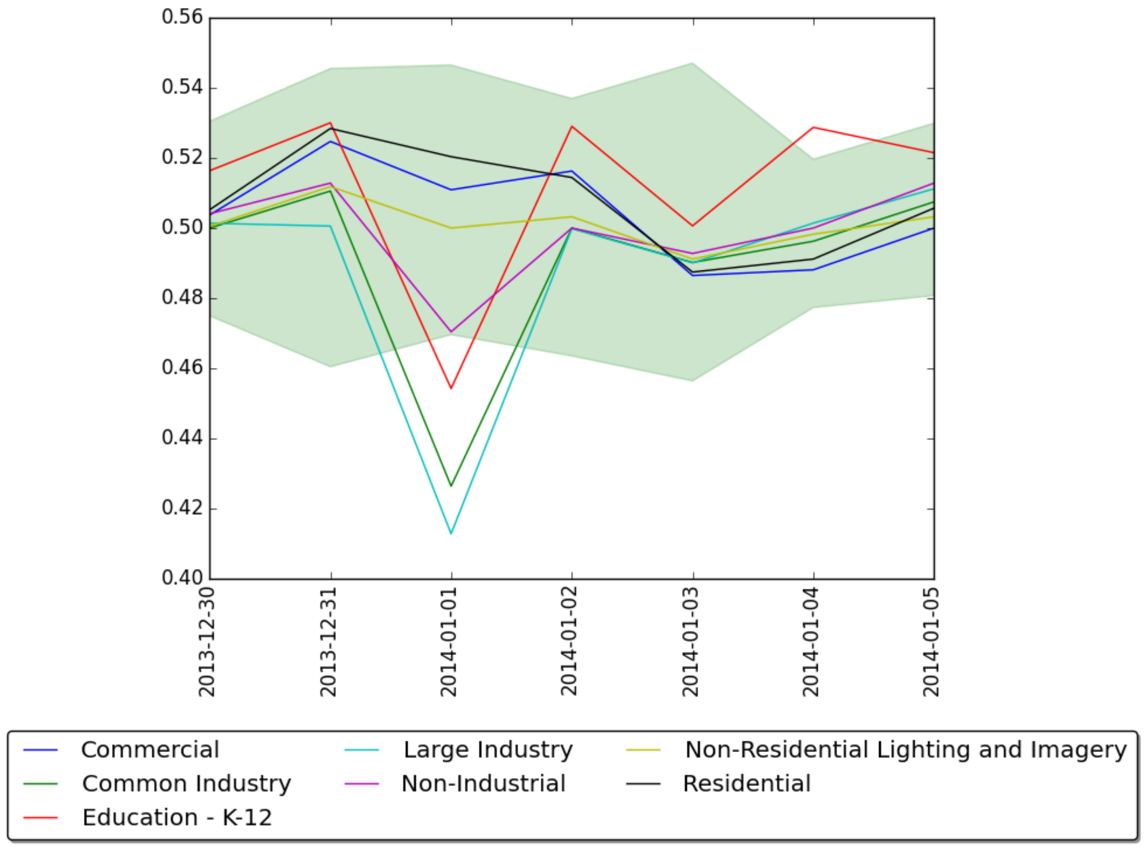
As with Christmas, electric consumption in Italy for small consumers is roughly steady through the week of New Year's relative to the preceding week. For large consumers, on the other hand, consumption is much lower on New Year's Eve and far lower (by over 30%) on New Year's Day; however, 30 December has noticeably higher consumption than 6 January, when Italy shuts down in honor of Epiphany. For

Figure 4-6: Median Ratio of Electricity Consumption by Contracted Power (Italy), New Year's Week (29 December 2015 to 4 January 2016)/(29 December 2015 to 4 January 2016 + Next Week)



large consumers, this pattern is largely uniform across all parts of Turin, although consumption declines slightly less in the downtown area. For small consumers, on the other hand, there is much more geographic variation. Consumption in downtown declines dramatically on New Year's Eve and New Year's Day, not recovering until 4 January, perhaps again reflecting the small commercial enterprises we believe may be included in this sample.

Figure 4-7: Median Ratio of Electricity Consumption by User Type (Country A), New Year's Week (30 December 2013 to 5 January 2014)/(30 December 2013 to 5 January 2014 + Next Week)



In Country A, on the other hand, consumption remains steady on New Year's eve, in fact increasing slightly, before dropping between 15 and 30% on New Year's day for industrial and educational entities. Residential customers, on the other hand, actually consume slightly more electricity on New Year's day and 2 January than the next week. Other user categories remain roughly steady in their electricity consumption after New Year's. Consumption increases on New Year's Eve are weakly concentrated near the county seat, but the distribution of declines on New Year's Day is not easily summarized.

In sum, workplace holidays associated with New Year lead to less consumption by large consumers/non-residential consumers in both Country A and Italy. Smaller consumers or residential consumers continue to consume at roughly the same pace during New Year's, suggesting that they either do not spend much time at home or that their presence at home does not drive an increase in electricity consumption.

4.1.3 Italian Holidays

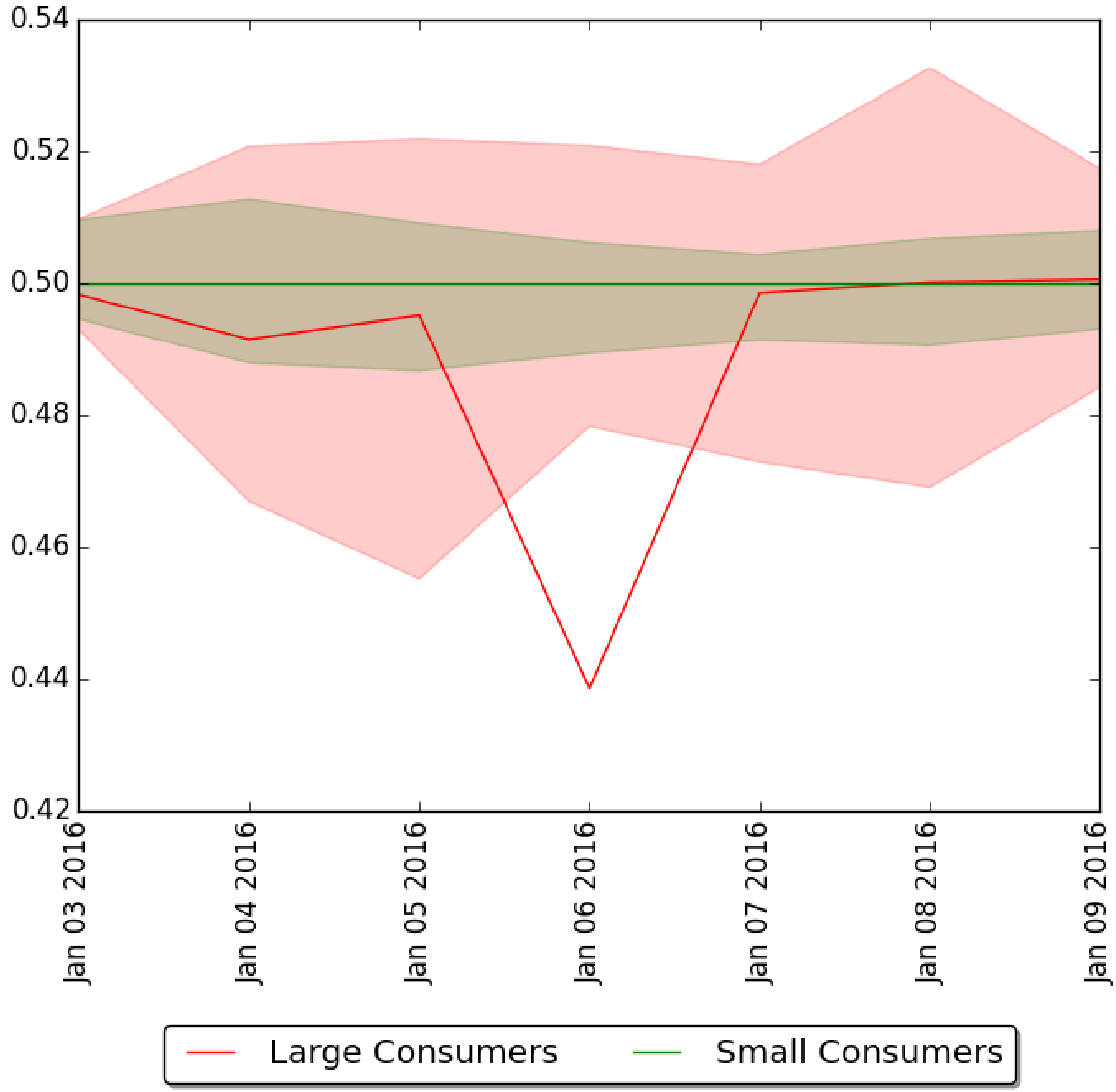
The Italians celebrate more holidays and we have a longer period of data to analyze. These holidays fall into two categories: (1) fall and winter holidays to which large consumers are highly responsive but small consumers are not and (2) spring and summer holidays that impact both small and large consumers. We speculate that this seasonal variation is due to a greater tendency to use holidays an excuse to leave town during the summer.

A Typical Winter Holiday: Epiphany 2016

A typical pattern of sharp holiday declines for large consumers and unaltered consumption for small consumers holds true for Epiphany, the Feast of the Immaculate Conception, and Labor Day. To avoid undue repetition we focus on Epiphany 2016.

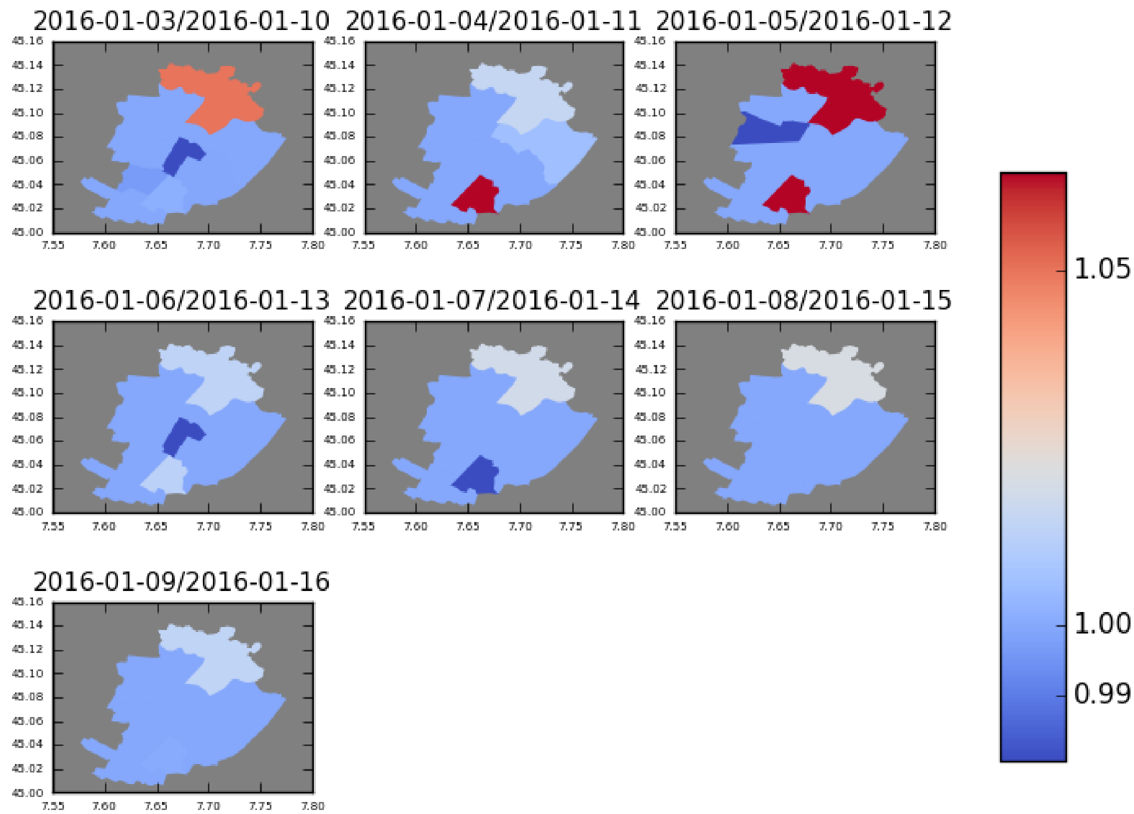
As Figure 4-8 shows, small consumer consumption tends to stay remarkably constant through these holidays, suggesting that even if more people stay home, they do not consume much more electricity. In fact, the majority of small consumers exhibit a slight, 5-10% increase in consumption, offset by a minority of about one third of consumers whose consumption declines markedly either for several days running up to Epiphany or only on Epiphany itself. Large consumers, however, consume much less electricity on the day of Epiphany, with small spill-over effects to the previous two days. We do not observe large counterbalancing increases in electricity consumption before or after these holidays. Moreover, drops in large consumer consumption are

Figure 4-8: Median Ratio of Electricity Consumption by Contracted Power (Italy), Epiphany 2016 Week/(Epiphany 2016 Week + Next Week)



consistently around 25%. However, almost half of large consumers actually maintain steady consumption patterns, while a minority of about 15% reduces consumption from 4 January through 6 January, and another 38% or so reduce consumption on Epiphany only (see Figure 4-11).

Figure 4-9: Map of Ratio of Electricity Consumption by Neighborhood for Small Consumers (Italy), Epiphany 2016 Week/(Epiphany 2016 Week + Next Week)



Geographic analysis shows that these patterns are largely uniform outside of the downtown area (Figures 4-9 and 4-10). The downtown area, however, runs generally counter to the trend. Among small consumers, consumption drops slightly in downtown not only on Epiphany but also on the preceding two days. This, we speculate, is due to the presence of non-residential customers in the small consumer category. More intriguingly, the consumption drop among large consumers downtown is much less pronounced than elsewhere in Turin. These geographic patterns broadly apply to the Feast of the Immaculate Conception and Labor Day, as well, although the variation in downtown consumption by small consumers is less for the Feast of the

Figure 4-10: Map of Ratio of Electricity Consumption by Neighborhood for Large Consumers (Italy), Epiphany 2016 Week/(Epiphany 2016 Week + Next Week)

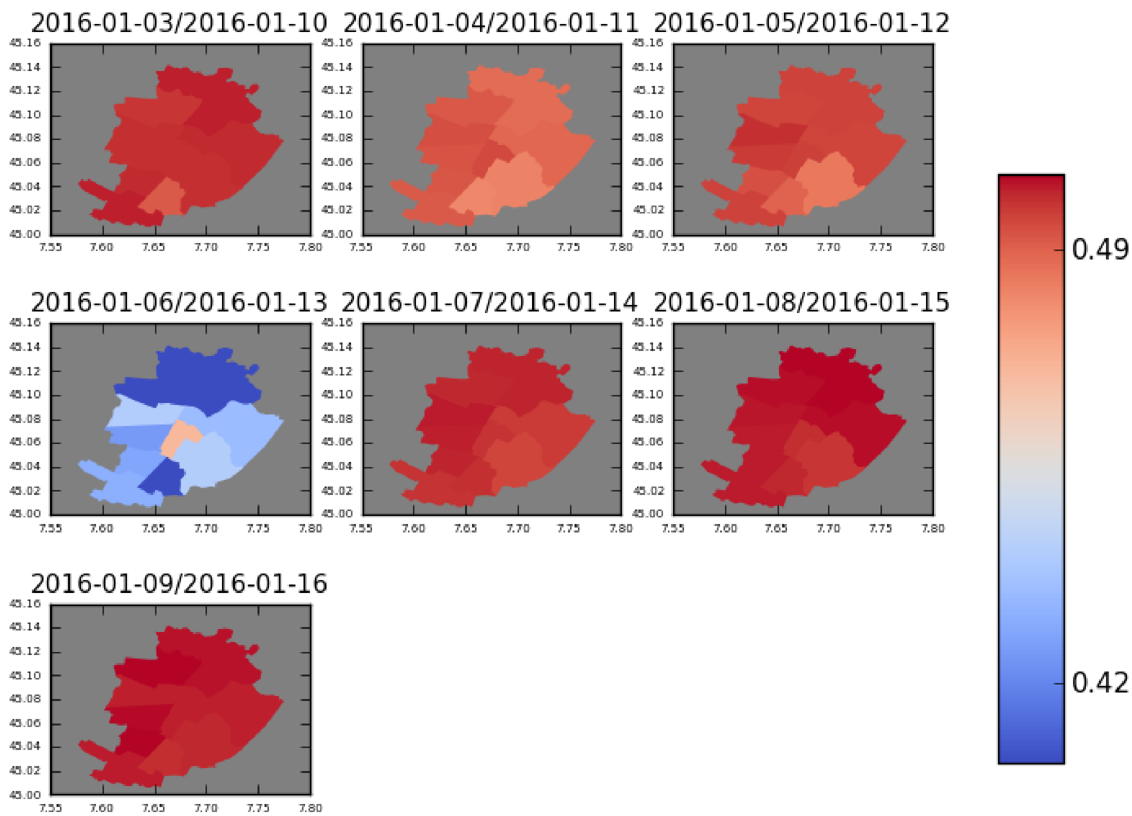
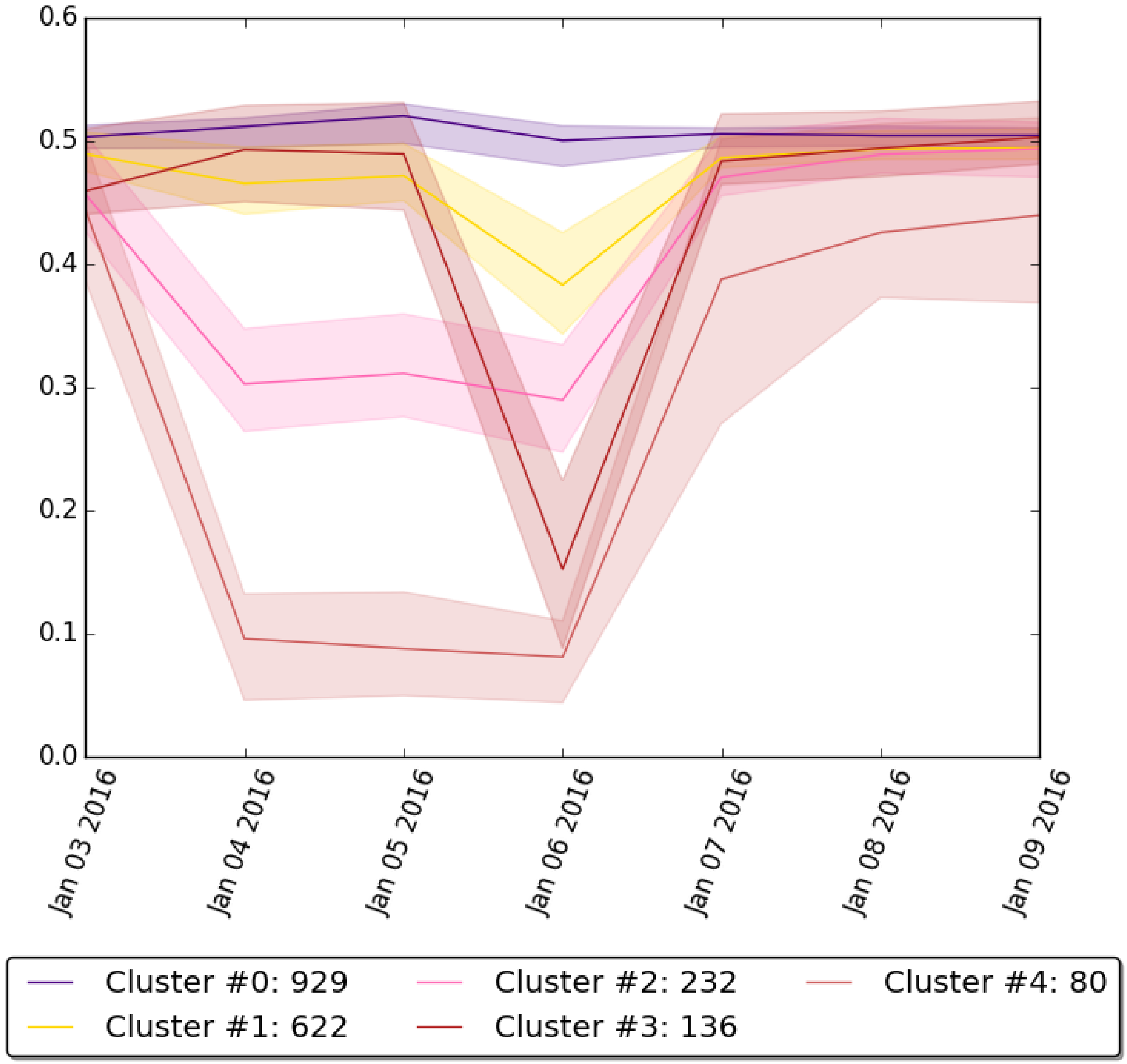


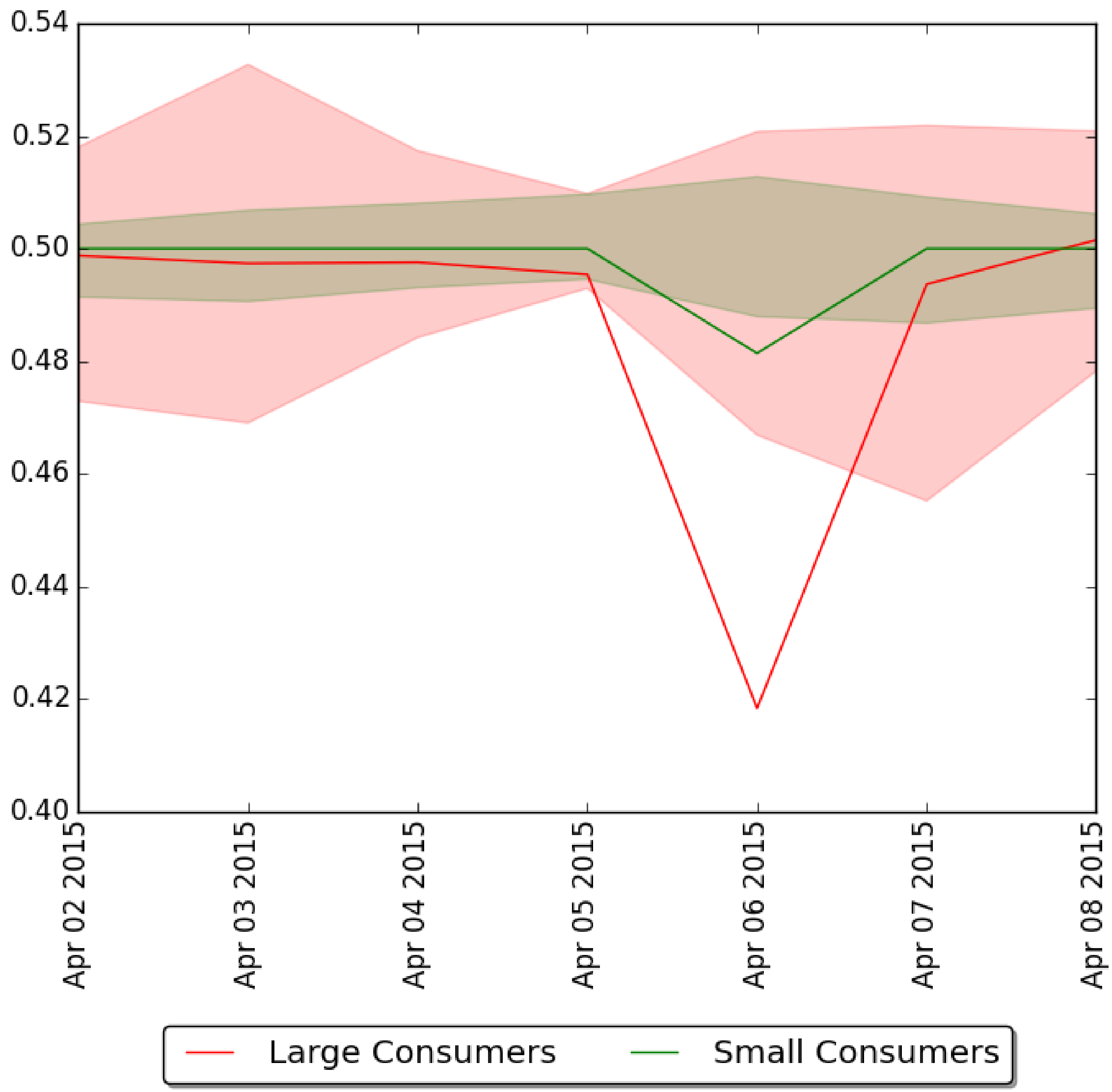
Figure 4-11: k-means Clustering for Epiphany Electricity Consumption as Fraction of Two Week Total for Large Consumers (Italy), Epiphany 2016 Week/(Epiphany Week + Next Week)



Immaculate Conception.

A Picnic Holiday: Easter

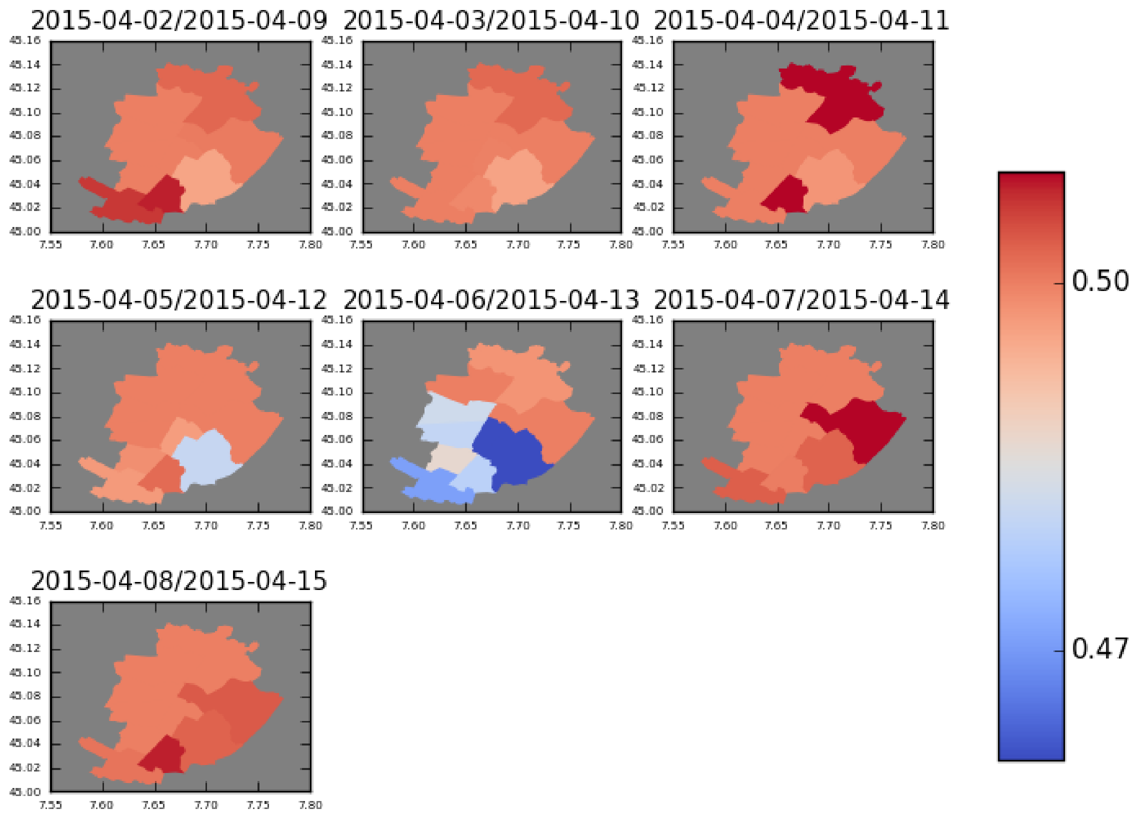
Figure 4-12: Median Ratio of Electricity Consumption by Contracted Power (Italy), Easter 2015 Week/(Easter 2015 Week + Next Week)



Easter provides some contrast with this typical winter pattern: both small and large consumer consumption dips beyond the norm, although the dip among large consumers is far larger.

As Figure 4-12 shows, the impact on consumption is concentrated almost exclusively on 6 April 2015, Easter Monday. Almost no change in consumption occurs on Easter itself (although there is a very slight dip in large consumers' consumption, and some small consumers do reduce consumption–while around 10% increase consumption). This suggests that the actual religious holiday observances do not much impact consumption patterns. Rather, it is Easter Monday–the public holiday–that drives down consumption. Indeed, Easter Monday is celebrated in Italy as *La Pasquetta*, with trips to the countryside for picnics a part of the tradition–which helpfully explains the drop in small consumer electric consumption on that day.

Figure 4-13: Map of Ratio of Electricity Consumption by Neighborhood for Small Consumers (Italy), Easter 2015 Week/(Easter 2015 Week + Next Week)



Curiously, though, the decline in small consumer consumption on Easter Monday is highly concentrated in the south of Turin. As Figure 4-13 shows, small consumer consumption in north Turin holds steady, whereas that in south Turin declines by 5-10%.

A Summer Holiday: The Feast of St. John

Lastly, we turn to a summer holiday, the Feast of St. John. The Feast of St. John is not actually a national holiday, but is rather celebrated only in a handful of northern Italian cities. Festivities vary by location, but in Turin they constitute some revelry the night before (the evening of 23 June) followed by an official local holiday on 24 June.

Figure 4-14: Median Ratio of Electricity Consumption by Contracted Power (Italy), Easter 2015 Week/(Easter 2015 Week + Next Week)

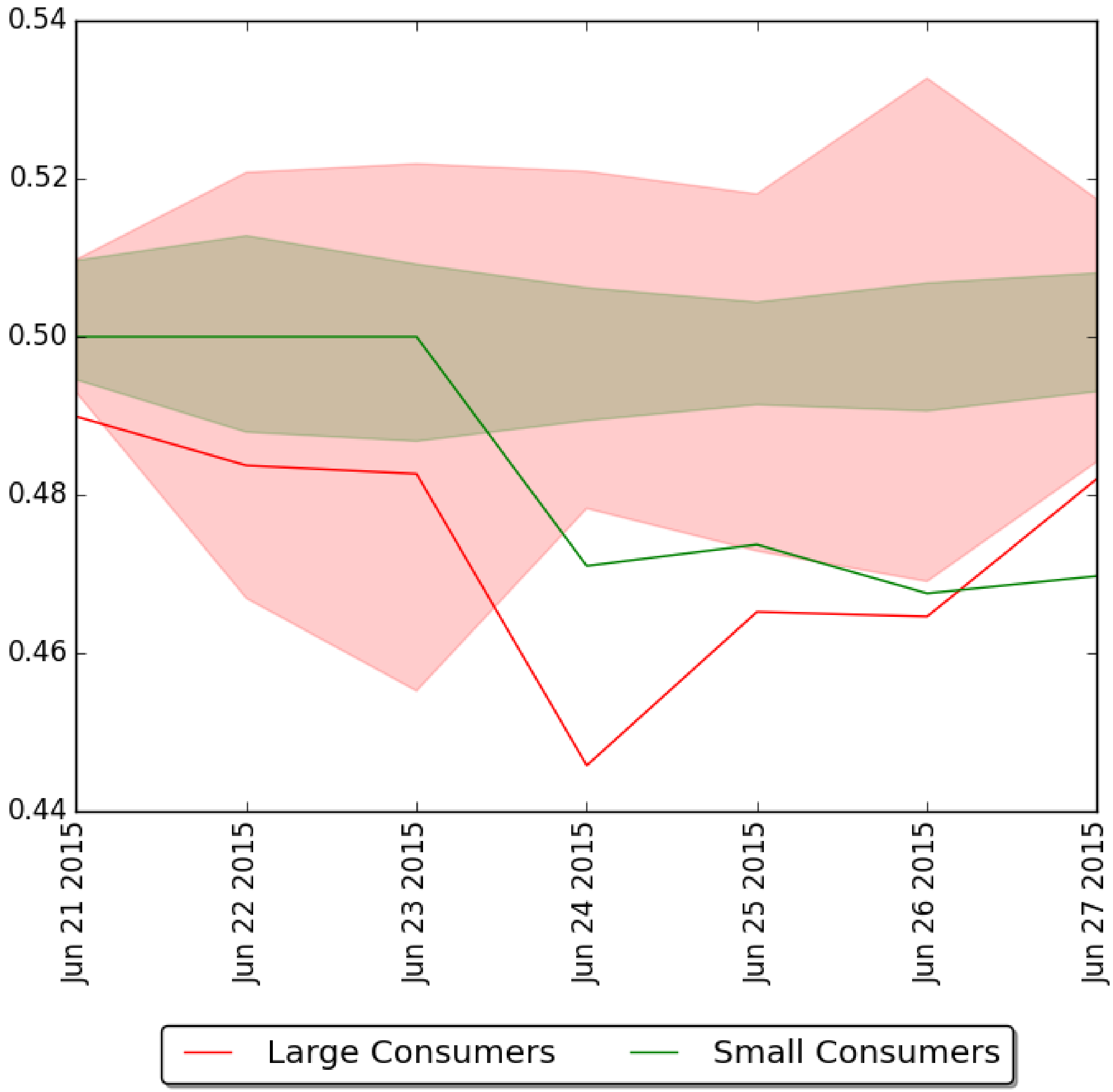


Figure 4-14 shows clearly that consumption drops significantly on 24 June. But consumption proceeds to remain low through the remainder of the workweek, even into Saturday 27 June. This holds true both for small and large consumers, and neither fully recovers until 2 July although large consumers recover more quickly. It appears that the Feast of St. John affords an opportunity for an extended vacation, both for producers and common residents. A similar pattern surrounds Republic Day in early June.

This pattern of nearly weeklong dips in consumption repeats throughout the summer, even in the absence of a catalytic holiday. Indeed, some of the sharpest week-over-week dips in small consumer consumption occur during the summer, such as the week from 24 to 29 August (or even later, from 24 to 27 September). These appear to be simple summer vacationing, with no particularly strong geographic bias in consumption reductions across Turin. It is remarkable, however, the extent to which these non-holiday reductions in summer consumption are uniform across the city, suggesting high correlation in vacation timing (or errors in the data!).[2]

4.1.4 Conclusions

Generally, considering both the Country A and Italian data, we find that for winter holidays–including major, universally celebrated holidays–electricity consumption among residential or small consumers is unaltered, whereas consumption among large or industrial consumers declines dramatically. Although we lack summer data for Country A, in Italy we find that summer holidays are more likely to lead to reductions in small consumer energy consumption, and indeed may precipitate weeklong vacations. In both countries, holidays have different impacts in different parts of the city. In County B, holidays have more of an impact near the county seat and in more urbanized areas; in Turin, the broader patterns of consumption change on holidays are tempered in the downtown area.

[2] As will be discussed in Appendix A, this may relate to the abnormally high temperatures Turin experienced in summer 2015 [25].

4.2 Time Changes

We turn finally to ask what impact daylight savings time has on electricity consumption. Not only was daylight savings time initially adopted to reduce long-term energy consumption,[3] it has also been shown to have a significant short-term impact on sleep habits, which in turn may impact short-term electricity consumption patterns. This impact is particularly pronounced and long-lasting in the spring, with people losing sleep due to the initial jump forward in clocks during the transition to Daylight Savings Time, then adjusting slowly to the new time [16]. This adjustment is much slower in the spring than the fall [14].

The sleep deprivation that results in the spring matters. Janszky and Ljung show that the chance of a heart attack on the Monday after the spring time change is higher than in the surrounding weeks; after the fall time change (when more sleep rather than less is possible), it is lower [12]. Likewise, Varughese and Allen find that fatal car accidents occur in the U.S. at a higher rate on the Monday after the spring time change than on adjoining days–but also on the Sunday of the fall time change [30]. (Similar research in Sweden, however, has suggested no statistically significant change in the rate of car crashes [17].) Clearly, the beginning and end of Daylight Savings Time can have significant impacts on human behavior, generally assumed to be a result of altered sleep patterns. We ask if these changed sleep patterns are evident in electricity consumption data.

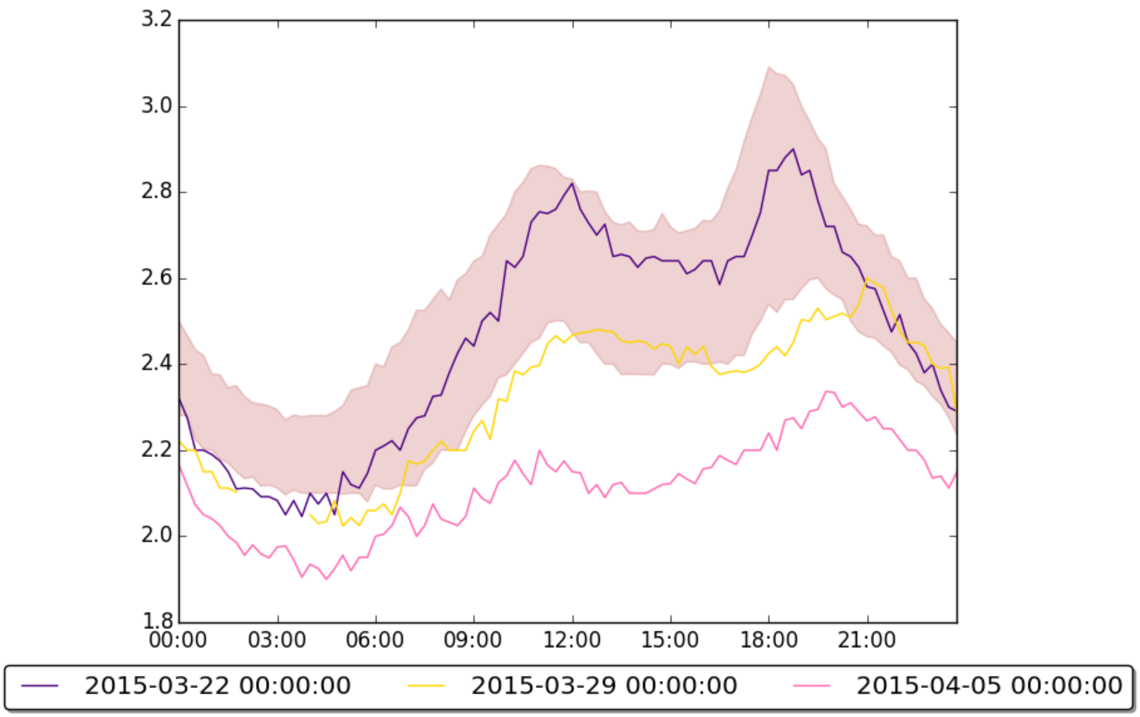
Unfortunately, the nature of the question requires looking at data at the individual day level, which is particularly noisy. Still, we attempt an analysis of the 2015 electricity consumption patterns surrounding both the March and October time changes. We do so only for Italy.

[3] In theory, this data should be able to evaluate this claim. However, the proximity of the spring implementation of DST to Easter impedes a meaningful direct comparison of aggregate consumption the week before and after DST implementation in the spring. In the fall, our results are inconclusive: DST appears to reduce consumption for small consumers and increase it for large consumers, but both changes are quite small and not necessarily meaningful (approximately 3%).

29 March: Clocks Move Forward 1 Hour (23 Hour Day)

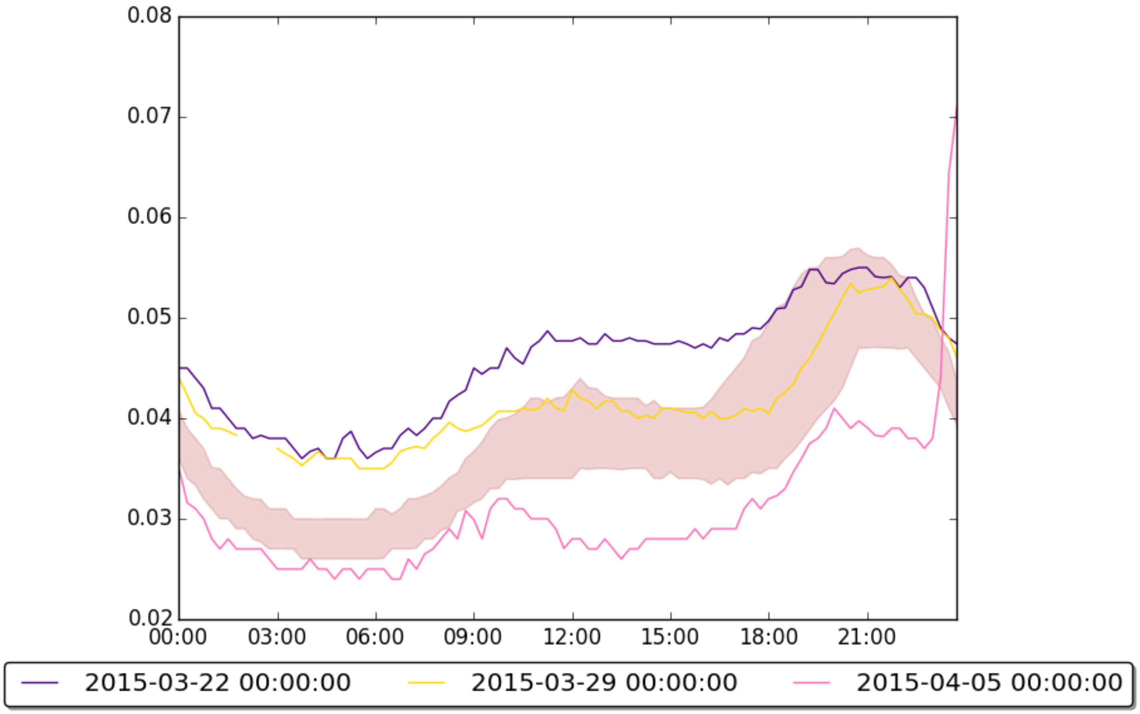
Daylight savings time for 2015 began on Sunday 29 March at 2 am, with clocks moving forward by 1 hour to read 3 am. As Figure 4-15 shows, large consumers' early morning consumption proceeded as if there had been no time change, with morning increases in consumption beginning about an hour later than they did on the previous Sunday. Consumption continued to rise at about an hour later than a week earlier until midday, when consumption plateaued at a lower level than the previous week. Consumption peaked in the evening later than it had a week earlier, but declined in the evening at roughly the same time as it had a week earlier. People do not stay late at work just because they arrived late. (Unfortunately, the next Sunday was Easter, limiting our ability to compare.)

Figure 4-15: Median 29 March DST Consumption Pattern, Large Consumers



Looking instead at small consumers, it is hard to distinguish the consumption pattern of 29 March from that a week earlier. Consumption through daylight hours on 29 March is lower than that on 22 March, but it is unclear if people are waking up later than usual (or just not bothering to wake up at all?). People do appear to

Figure 4-16: Median 29 March DST Consumption Pattern, Small Consumers



wind down the day's electricity consumption at the same time, however.

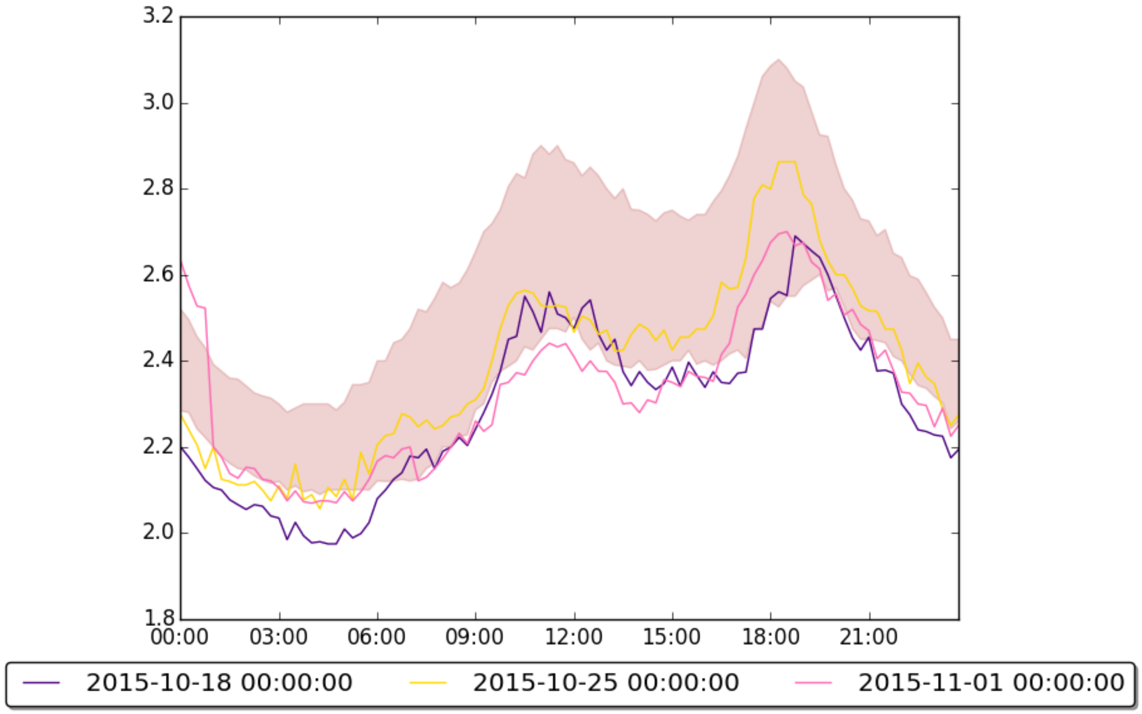
We can also consider electricity consumption the Monday after the time change. For both large and small consumers, median consumption is almost identical on the Monday after the time change as the Monday one week before. Only by considering the mean consumption curve do we notice that large enterprises, on average, did not consume as much electricity as early in the morning on the Monday after the time change–but only by a small margin.

25 October: Clocks Move Backward 1 Hour (25 Hour Day)

2015 daylight savings time ended at 2 am Sunday 25 October, with clocks moving backwards an hour to repeat the 1 am hour. This results in a particularly long day, and we might expect people to wake up earlier (on the clock)–having had a full night's sleep–and go to sleep a bit earlier–having exhausted themselves during the day.

Large consumers, it appears, simply work a longer day. Not only do they start consuming electricity earlier in the morning, but they also continue consumign more elec-

Figure 4-17: 25 October DST Consumption Pattern, Large Consumers

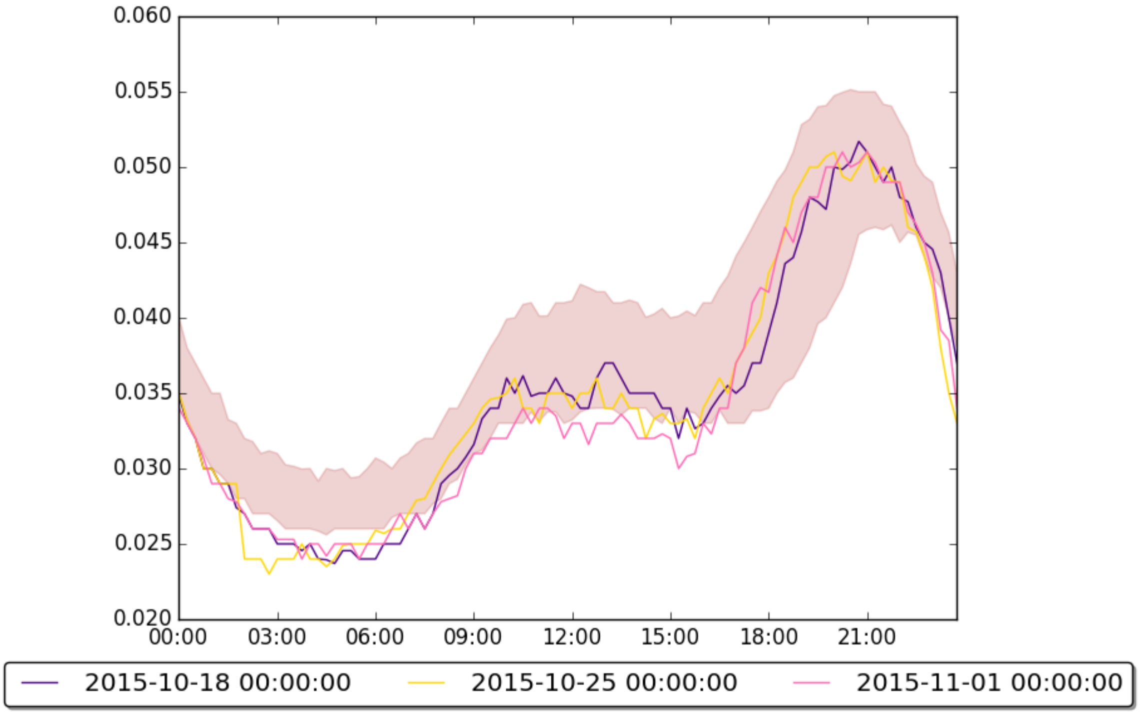


tricity late into the evening–the consumption curve for 25 October is almost strictly higher than that for either the preceding or succeeding Sundays (see Figure 4-17).

Small consumers, on the other hand, display negligible abnormal behavior (see Figure 4-18). Consumption on 25 October is almost identical to that on the preceding and succeeding Sundays, although there is potential evidence that small consumers might have woken up slightly earlier on 25 October. The definitely went to bed at almost exactly the same time–albeit perhaps ever so slightly earlier.

Consumption on Monday the 26th is largely normal. Median large consumer consumption is essentially identical to that a week earlier or later and mean consumption only suggests slightly earlier declines in consumption in the evening. Small consumers, on the other hand, begin and end their daily consumption at roughly the same time as they did a week earlier, but rather begin their evening peak in consumption (the global maximum of small consumer consumption) about an hour earlier. This, however, could be explained by turning the lights on earlier, as dusk would fall about an hour earlier after the time change–a theory reinforced by the even earlier onset of the

Figure 4-18: 25 October DST Consumption Pattern, Small Consumers



evening peak the week after the time change.

Conclusions

It is very hard to draw conclusions about the impact of daylight savings time on electricity consumption from the present data. The data we work with is relatively noisy compared to the one hour shift that–in the "best" case scenario–we would expect to observe. Smaller shifts become almost impossible to notice. Setting aside any concerns about the precision of our data and assuming that it suffers from no distortions due to noise, it appears that daylight savings time may have particularly pronounced impacts in the morning on the day of the time change, but these peter out by the Monday after the time change. Sleep times seem to go by the clock, although people appear to return home earlier (or perhaps turn on the lights earlier due to waning sunlight) on the Monday after the fall time change.

Chapter 5

Conclusion

This thesis builds on earlier research examining the geographic distribution of cell phone and internet use, as well as the various patterns of household electricity use, by examining the temporal and geographic distribution of electricity use across all major types of electricity consumers. Moreover, by analyzing electricity consumption from a suburban Country A county and a small Italian city, this thesis provides some preliminary insight into cross-national variation in electricity consumption.

We find that industrial and large consumers in Turin are more inclined to reduce consumption substantially on weekends than in County B. Industrial and large consumers in the two locales do share some characteristics, however: on holidays, they tend to cut electricity consumption quite dramatically, particularly in comparison to residential and small consumers, who often exhibit no changes in electricity consumption. In summer, however, particularly during unseasonably warm periods, small consumers in Turin do reduce consumption, and do so for extended periods of time. In Italy, in particular, we are able to show not only that consumption is highly correlated within clearly identifiable seasons, but also that consumption declines for holidays are highly correlated and concentrated among a particular subset of users. It seems entirely plausible that these results apply more broadly to Country A and Italy as a whole, but our data does not leave us in a position to assert as much with confidence. Future research into larger electricity consumption datasets could profitably explore this question, and indeed extend it to comparisons between additional countries and regions.

We also uncover various geographic patterns in the consumption of electricity. These, however, tend to be different in Turin and County B. In County B, we find that, in general, industrial and urban and suburban electricity consumption peaks in the morning and residential and rural electricity consumption peaks in the evening. In Turin, we detect a slight tendency for smaller consumers and those located downtown or in adjoining primarily residential areas to consume more in the evening, whereas those in the industrial suburbs are more likely to reach peak consumption in the morning. It thus becomes clear that residential consumption is typified by higher consumption in the evening and industrial consumption by higher consumption in the morning punctuated by a pronounced lunch break.

Given our observations on the distinctiveness of industrial and residential consumption patterns, it is entirely consistent that using only consumption curves, we are able to classify meters into residential and industrial categories with relatively high accuracy. Moreover, given that we find significant overlap between the consumption patterns typical of other user categories and subcategories, it comes as little surprise that we are unsuccessful at classifying meters into more specific and more varied categories. We conclude from this exercise that electricity consumption patterns are not well suited to classifying land use using traditional categories. They may, however, present an alternative way of classifying land use that could prove useful from an electric grid management perspective.

The findings of this thesis are, in general, preliminary in nature. The data from County B is not representative of County B, let alone Country A. The data from Turin is representative of Turin, but probably not of all Italy. Still, the findings contribute a rudimentary sense of how consumption patterns likely vary between countries, and a stronger if not infallible sense of how they vary between consumer categories.

As larger datasets become available to explore electricity consumption, more exhaustive comparative work will become feasible. This thesis suggests that there is value in undertaking such research, and that the results will likely tell us much about how different parts of the world rest and work—and, in turn, the work-life balance that persists in various places, important questions for planners to understand when they approach a new community. This thesis also points to the value of geographic analysis in analyzing electricity consumption. Although some of our attempts at geographic analysis resemble a Rorschach test, others show clear differences between urban, suburban, and rural areas. These patterns do not appear to be identical in County B and Turin, suggesting that any theory that results will necessarily be nuanced and complicated. Nonetheless, we show that electricity consumption data can be used to show that different areas exhibit different behaviors, and further research could fruitfully explore the geographic correlates of these behavior differences. In turn, better understanding these geographic differences in electricity consumption patterns can help planners appropriately arrange land uses to balance electricity demand, as well as allocating electric grid resources for new areas according to typical geographic demand patterns.

Appendix A

Weather and Electricity Consumption

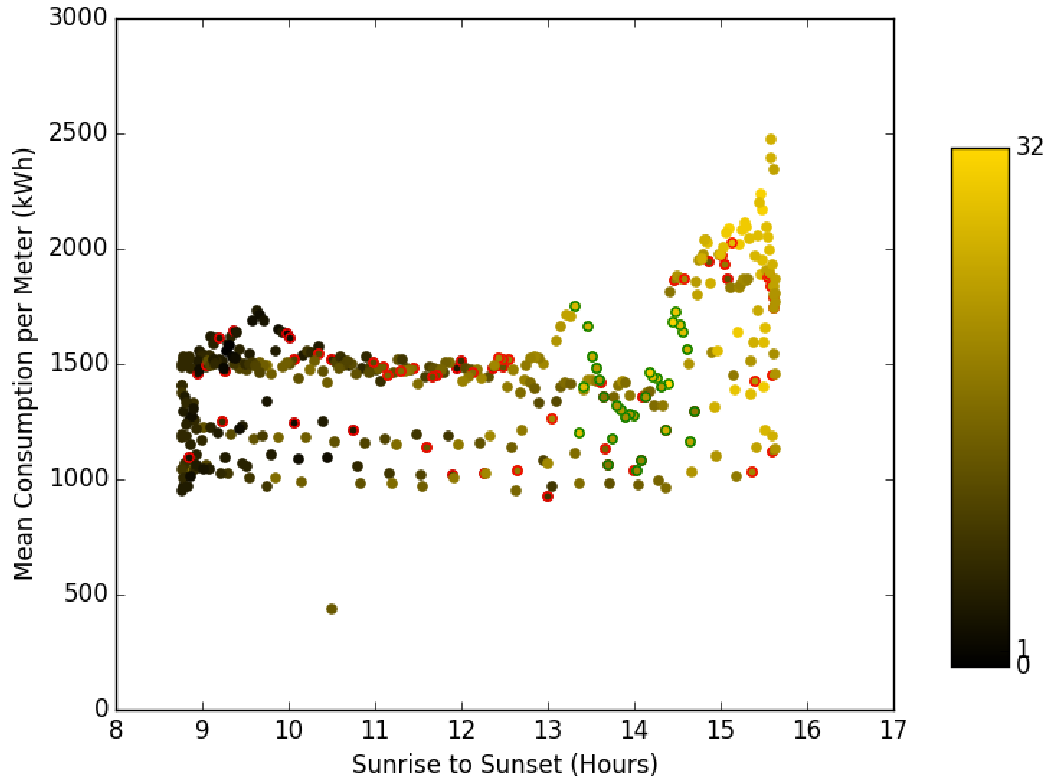
It is worth briefly exploring the relationship between weather, sunlight, and electricity consumption. Although this is not meant to be an exhaustive analysis of the subject, it points to some preliminary conclusions. Namely, consumers are relatively responsive to annual cycles in daily hours of sunlight, consuming more during darker periods of the year. Consumers do consume much more during summer periods with high temperatures, although vacation periods to some extent counteract this. Rain's impact on consumption appears to vary with temperature and user characteristics. Unfortunately, it is quite hard to tease apart different influences on electricity consumption: temperature is closely correlated with time of year, as is the number of hours from sunrise to sunset.

Italy

We begin by looking at Italy, because our data for Turin is more complete and thus our findings are more coherent. Large and small consumers respond quite differently to weather and variations in sunlight, with far greater responsiveness from small consumers.

Large consumers have relatively constant consumption patterns. Considered daily, their consumption during weekdays is relatively stable, with lower and stable weekend consumption. These patterns hold steady from the lowest number of hours of sunlight (really measured as the hours from sunrise to sunset) in the dead of winter (about

Figure A-1: Electricity Consumption by Hours of Sunlight, Large Consumers (Italy)

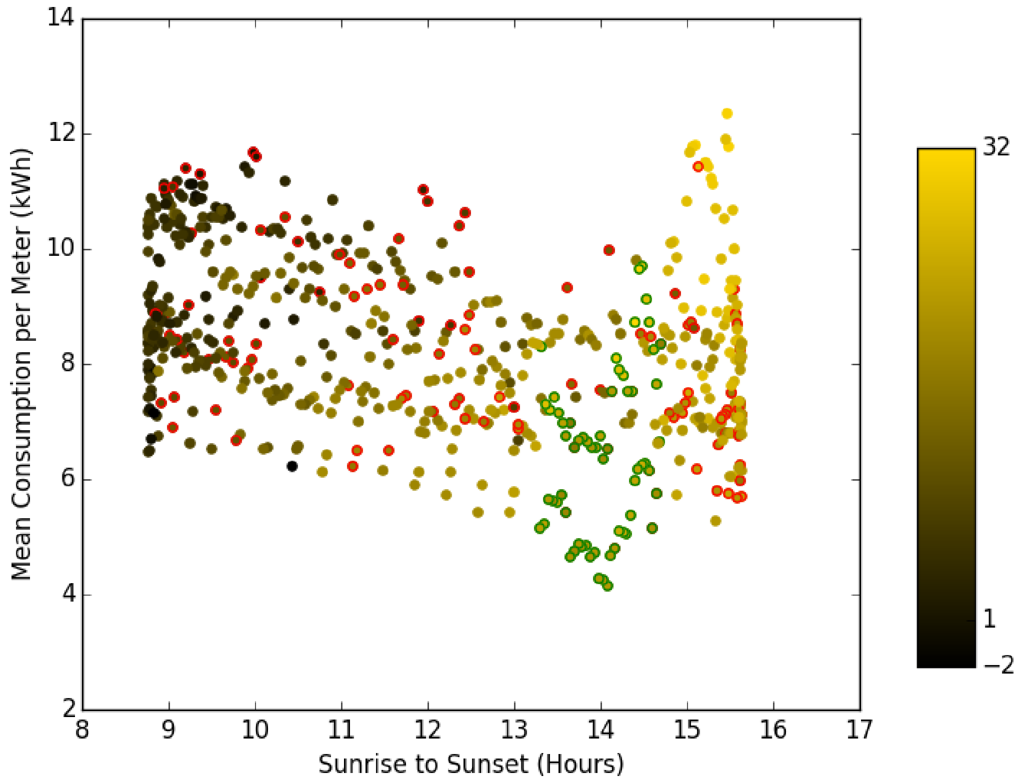


Note: Lighter colored dots are days with higher temperatures. Red perimeters denote rainy days. Green perimeters denote August days.

8.5 hours) to a little over 14 hours of sunlight. Close to the summer solstice, however, there is an abrupt increase in electricity consumption that correlates with the highest temperatures of the year. The abruptness of this increase might be explained by August's presence, with its negative impact on electricity consumption even for these large consumers. It is unclear, meanwhile, whether rain has much of an effect on large consumers' electricity consumption.

Small consumers, on the other hand, have two peaks of electricity consumption at the two solstices, both when sunlight is in particularly short supply and when it is particularly plentiful. The role of higher temperatures is not immediately clear for most seasons, but in summer, the hottest days clearly correlate with the highest electricity consumption. Interestingly, in summer, rain corresponds to lower electricity consumption (presumably reducing demand for air conditioning), whereas in

Figure A-2: Electricity Consumption by Hours of Sunlight, Small Consumers (Italy)



Note: Lighter colored dots are days with higher temperatures. Red perimeters denote rainy days. Green perimeters denote August days.

winter rain corresponds to relatively high consumption. August shows a clear dip in electricity consumption and indeed includes the days with the lowest consumption overall.

Performing linear regression largely confirms our observations from Figures A-1 and A-2. For large consumers, higher temperatures correspond to higher consumption. Rain leads to higher consumption, except during the summer months (May to August), when it appears to lead to lower consumption (although neither coefficient is not statistically significant even at 90%). On the other hand, summer consumption overall is much higher (400 kWh per day) than the rest of the year, an influence that is fully cancelled out by the decline of consumption in August. Furthermore, consumption declines by 300 kWh per day on Saturdays and by over 500 kWh per day on Sundays. Curiously, longer days are associated with lower consumption, the reverse of what appears in the graph without controlling for temperature, etc. Most remarkable is that this season- and weather-based model is remarkably good at predicting daily mean electricity consumption, with an r^2 of 0.76.

The linear regression model for small consumers yields a less complete description of the data, with an r^2 of only 0.40. The overall pattern is roughly the same: longer days lead to less consumption, as do August, Saturdays, and Sundays. However, August (tentatively, and within our error bars) has a more negative impact on consumption than Sundays, the reverse of the case for large consumers. Similarly, again, summer and rain lead to higher consumption, but summer rain leads to lower consumption–although again the coefficient on rain is not statistically significant. The only difference in sign on the coefficients is that higher temperatures lead to lower small consumer consumption but higher large consumer consumption, presumably for reasons related to heating and cooling systems.

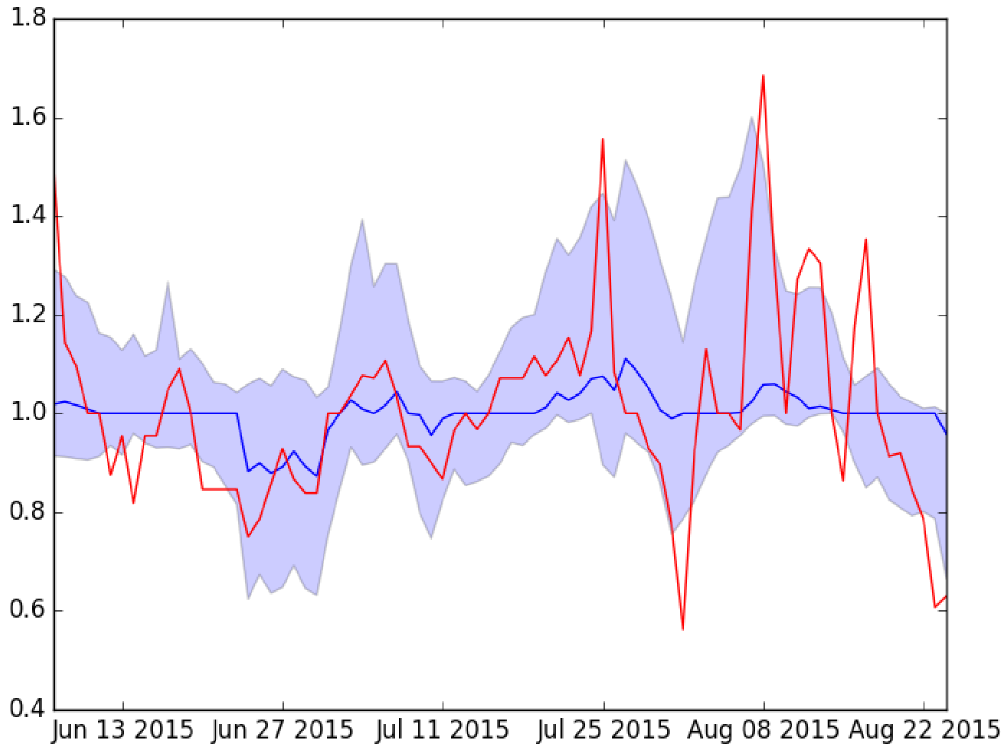
For reference, we include as well Figures A-3 and A-4, which show the median week-over-week ratio of consumption by day for both small and large consumers in Italy in comparison to the week-over-week ratio of mean temperatures for summer 2015. Clearly, the unseasonably hot temperatures Turin experienced in 2015–particularly in July–were closely correlated with changes in electricity consumption for large consumers. The pattern is much weaker for small consumers.

Overall, our main conclusion is that electricity consumption, at least at the aggregate level, is indeed highly determined by seasonal and climatic patterns.

Country A

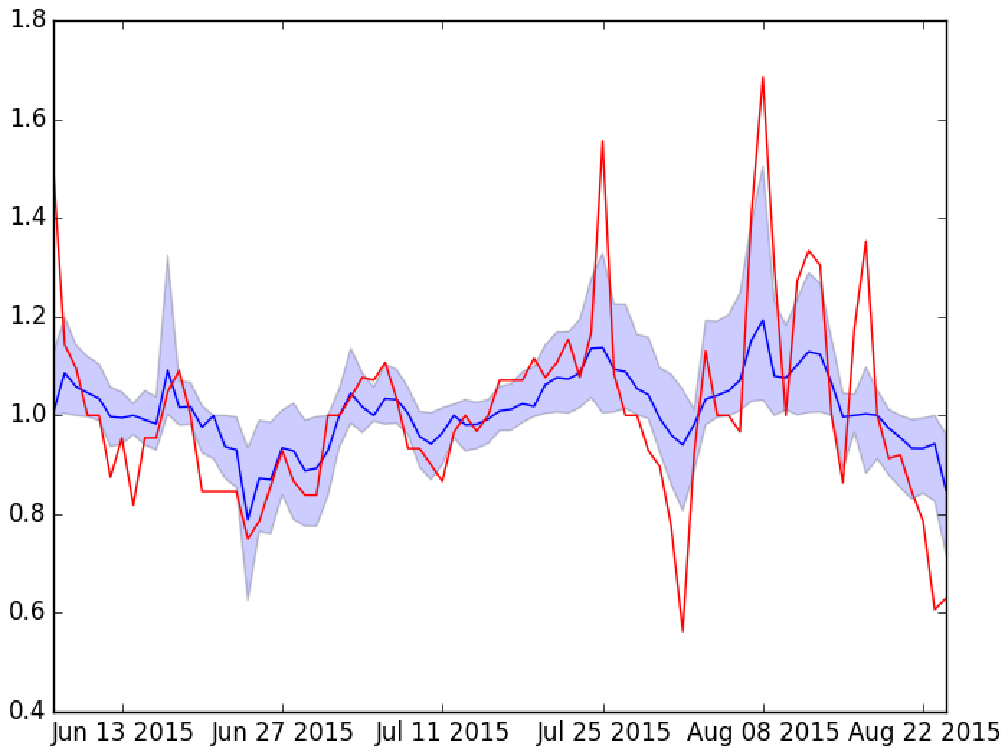
Our Country A data does not cater as well to this type of analysis because it includes only the winter months. Nonetheless, it is worth checking to see if any clear patterns emerge. Most notably, we see that shorter days correspond to higher electricity consumption. Likewise, a spate of chilly rainy days in March included relatively high electricity consumption. Matters are complicated by significant missing data from February, among other problems. Hence it is hard to draw meaningful conclusions.

Figure A-3: Median Week-over-Week Electricity Consumption Change Compared to Week-over-Week Temperature, Small Consumers (Summer 2015, Italy)



Note: Blue denotes the median over functioning meters of the ratios of electricity consumption on one day to that a week later. Blue denotes the interquartile range. Red denotes the week-over-week ratio of mean temperature..

Figure A-4: Median Week-over-Week Electricity Consumption Change Compared to Week-over-Week Temperature, Large Consumers (Summer 2015, Italy)



Note: Blue denotes the median over functioning meters of the ratios of electricity consumption on one day to that a week later. Blue denotes the interquartile range. Red denotes the week-over-week ratio of mean temperature..

Figure A-5: Electricity Consumption by Hours of Sunlight (Country A)

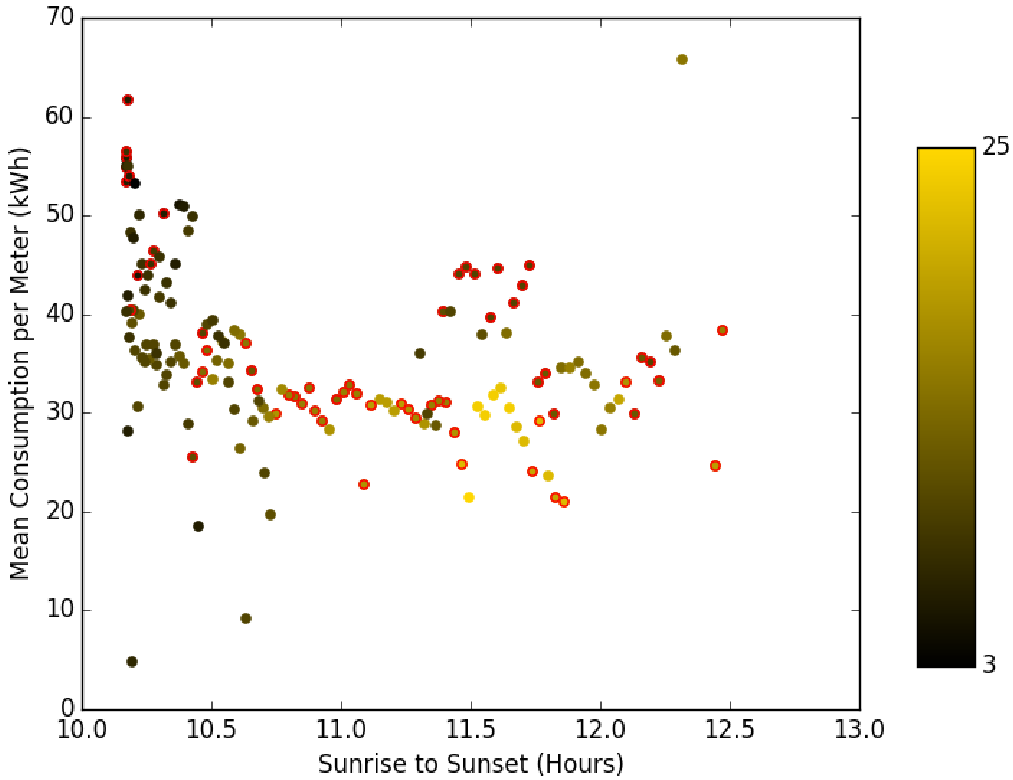


Table A.1: Linear Regression: Predicting Daily Mean Electric Consumption by Weather Conditions (Large Consumers, Italy)

OLS Regression Results

Dep. Variable:	meandailyconsumption	R-squared:	0.761
Model:	OLS	Adj. R-squared:	0.757
Method:	Least Squares	F-statistic:	154.4
Date:	Thu, 02 Jun 2016	Prob (F-statistic):	2.04e-115
Time:	00:07:30	Log-Likelihood:	-2531.4
No. Observations:	396	AIC:	5081.
Df Residuals:	387	BIC:	5117.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf.	Int.]
const	1670.0127	69.423	24.055	0.000	1533.518	1806.507
day_hours	-26.8188	7.665	-3.499	0.001	-41.890	-11.748
meantempm	8.9215	2.010	4.438	0.000	4.969	12.874
rain	41.2667	25.794	1.600	0.110	-9.448	91.981
August	-393.1288	32.092	-12.250	0.000	-456.225	-330.033
Saturday	-300.5111	21.269	-14.129	0.000	-342.328	-258.694
Sunday	-510.4395	21.267	-24.001	0.000	-552.254	-468.625
summer	404.8438	32.219	12.566	0.000	341.499	468.189
summer rain	-60.7945	43.289	-1.404	0.161	-145.906	24.317

Omnibus:	120.614	Durbin-Watson:	0.882
Prob(Omnibus):	0.000	Jarque-Bera (JB):	729.442
Skew:	-1.145	Prob(JB):	4.01e-159
Kurtosis:	9.242	Cond. No.	203.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Table A.2: Linear Regression: Predicting Daily Mean Electric Consumption by Weather Conditions (Small Consumers, Italy)

```
                            OLS Regression Results
==================================================================================
Dep. Variable:     meandailyconsumption   R-squared:                       0.408
Model:                              OLS   Adj. R-squared:                  0.400
Method:                   Least Squares   F-statistic:                     51.64
Date:                  Wed, 01 Jun 2016   Prob (F-statistic):           1.75e-63
Time:                          00:40:31   Log-Likelihood:                -975.99
No. Observations:                   609   AIC:                             1970.
Df Residuals:                       600   BIC:                             2010.
Df Model:                             8
Covariance Type:              nonrobust
=================================================================================
                  coef    std err          t      P>|t|      [95.0% Conf. Int.]
---------------------------------------------------------------------------------
const          11.0928      0.483     22.967      0.000        10.144    12.041
day_hours      -0.1439      0.054     -2.682      0.008        -0.249    -0.039
meantempm      -0.0529      0.013     -3.985      0.000        -0.079    -0.027
rain            0.1549      0.161      0.963      0.336        -0.161     0.471
August         -1.8060      0.195     -9.274      0.000        -2.188    -1.424
Saturday       -1.1098      0.142     -7.796      0.000        -1.389    -0.830
Sunday         -1.4348      0.143    -10.068      0.000        -1.715    -1.155
summer          0.8924      0.223      4.005      0.000         0.455     1.330
summer rain    -0.8017      0.260     -3.089      0.002        -1.312    -0.292
================================================================================
Omnibus:                       18.137   Durbin-Watson:                   0.211
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.931
Skew:                           0.419   Prob(JB):                     7.75e-05
Kurtosis:                       3.208   Cond. No.                         220.
================================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

Appendix B

Night Lighting and Electricity Consumption

The large number of countries with unreliable official economic statistics has led to a cottage industry in studies using night lighting as a proxy for economic activity. The logic holds that high energy consumption correlates well with economic activity, and moreover high energy consumption correlates well with night lighting. We briefly explore whether our data is able to support or challenge this second supposed correlation.

We use 2013 night lighting data from NASA.[1] We then compare this with total electricity consumption in areas that correspond to the pixels of these night lighting images. Unfortunately, we encounter three problems with the data:

- Night lighting data saturates easily in large cities. Thus for the area of Turin where our data comes from, we only have three levels of night lighting (61, 62, 63) on a scale of integers from 0 to 63. This narrow band of values does not cater to fine-grained analysis.

- The imprecision of our Italian data's geographic information becomes a source of error as we assign night lighting pixels to the grid of geocoded locations.

- Our unrepresentative sample in Country A would make it impossible for us to

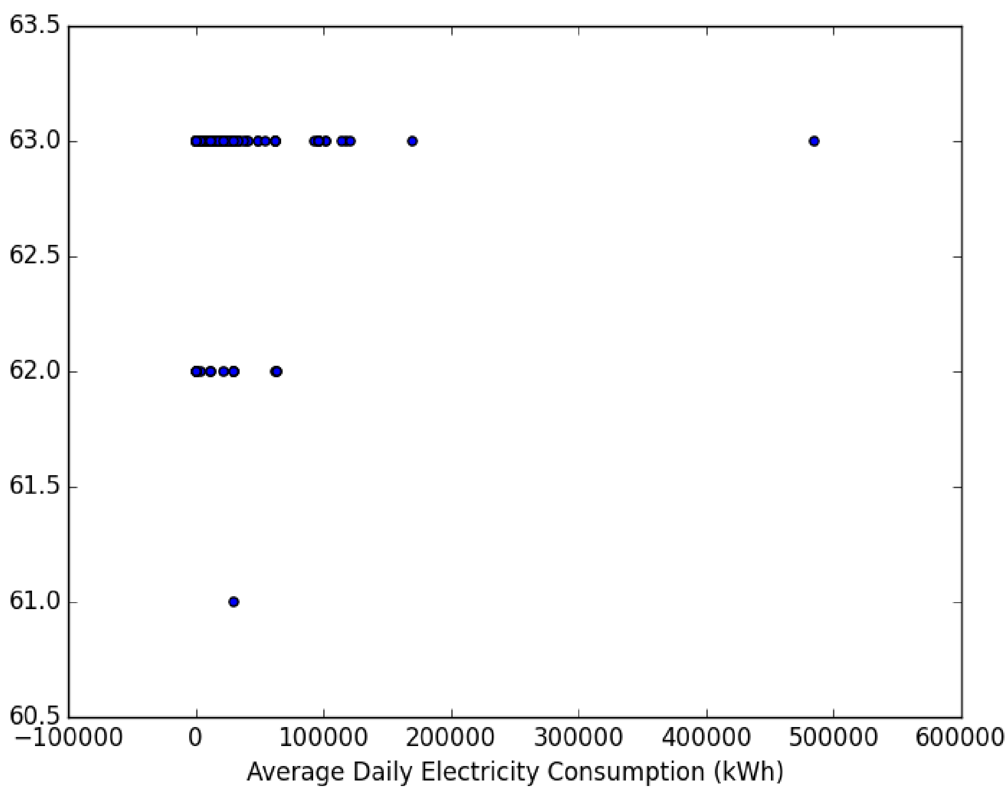
[1] http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

claim confidently that electricity consumption and night lighting are not closely related.

We proceed understanding these serious limitations.

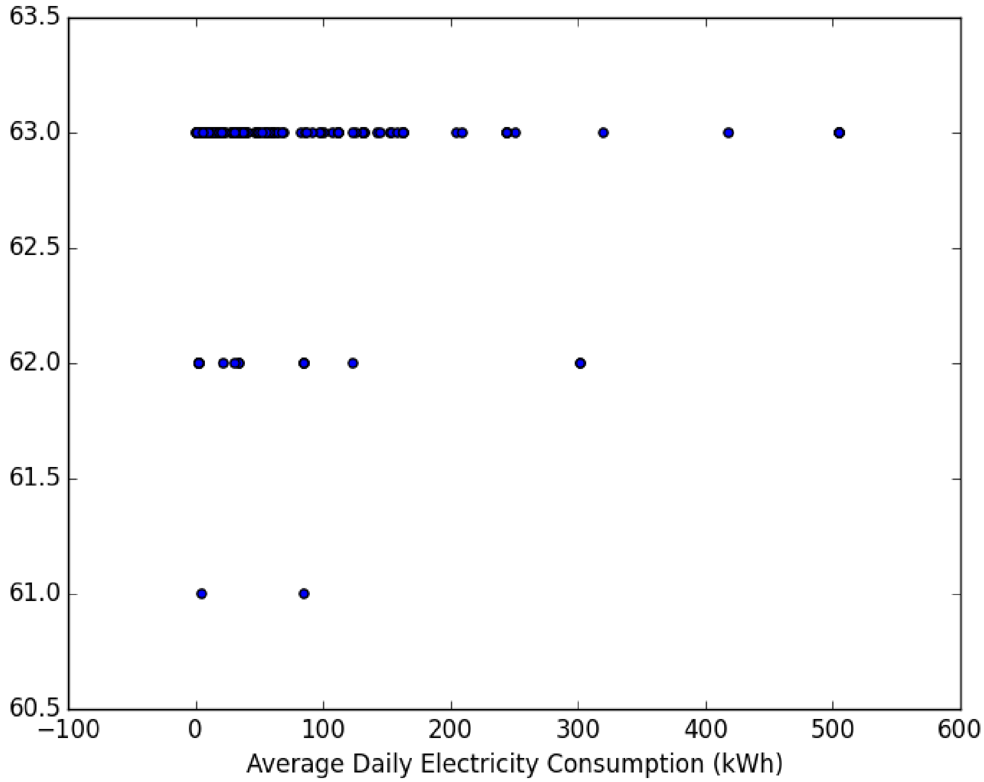
Italy

Figure B-1: Electricity Consumption versus Night Lighting, Large Consumers (Italy)



For both large and small consumers in Italy, it is clear that neighborhoods with particularly high consumption tend to be maximally lit. However, neighborhoods that have low or typical consumption may be very well lit or less well lit. Combined, this yields a correlation coefficient indistinguishable from zero, suggesting that neighborhoods with bright lights might not actually be home to particularly high electricity consumption. On the other hand, it appears safe to conclude that areas without bright lights do not consume large amounts of electricity. In sum, caution probably ought to be used in using night lighting as a proxy for electricity consumption (and thus for economic activity).

Figure B-2: Electricity Consumption versus Night Lighting, Small Consumers (Italy)

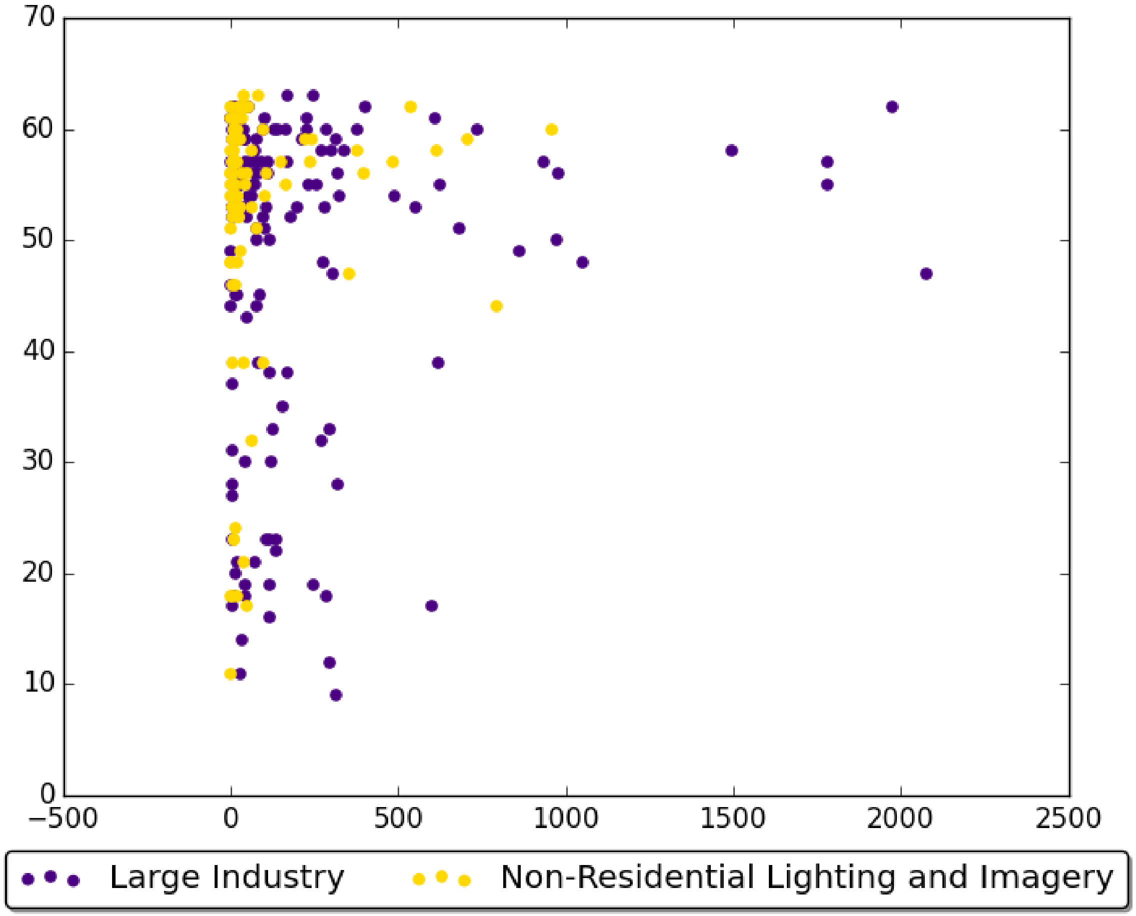


Country A

Although our data from Country A is not as complete, we are able to categorize meters by user type. We take particular interest in Large Industry and Non-Residential Lighting and Imagery Consumers. Since our dataset is biased by including only high voltage customers, our sample of Large Industry customers is more likely to be complete. And the safest correlate of night lighting is likely street lighting, which is conveniently included in our Non-Residential Lighting and Imagery category.

Again, we do not find particularly pretty results. The same pattern in which areas with particularly high consumption are relatively well lit holds, although not to the same extent as in Turin. Curiously, Non-Residential Lighting and Imagery does not

Figure B-3: Electricity Consumption versus Night Lighting, by User Type (Country A)



correlate particularly well with night lighting. But nor does large industry. Plenty of areas with low electricity consumption by these two categories have bright night lighting. And plenty of areas with relatively high (although not extraordinarily high) electricity consumption have dim night lights.

It is hard to draw any definitive conclusions from this brief exploration. However, we emphasize that we have not found strong evidence that bright areas of a city are necessarily consuming a great deal of electricity, which should give pause to users of night lighting data as a proxy for economic data.

Bibliography

[1] Smart Electricity Meters to Total 780 Million in 2020, Driven by China's Rollout. *T.D. World Magazine.*

[2] Tom Baum and Svend Lundtorp, editors. *Seasonality in Tourism.* Routledge, December 2001.

[3] F. Calabrese, J. Reades, and C. Ratti. Eigenplaces: Segmenting Space through Digital Signatures. *IEEE Pervasive Computing*, 9(1):78–84, January 2010.

[4] A. P. Carleer and E. Wolff. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, 27(6):1035–1051, March 2006.

[5] Liqun Chen. *Redefining the typology of land use in the age of big data.* Thesis, Massachusetts Institute of Technology, 2014.

[6] Leyna Sadamori Christian Beckel. Revealing Household Characteristics from Smart Meter Data. *Energy*, 78:397–410, 2014.

[7] D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, April 1979.

[8] B. Dixon and N. Candade. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both? *International Journal of Remote Sensing*, 29(4):1185–1206, February 2008.

[9] Laxmi Narayana Eeti, Krishna Mohan Buddhiraju, and Avik Bhattacharya. A SINGLE CLASSIFIER USING PRINCIPAL COMPONENTS VS MULTI-CLASSIFIER SYSTEM : IN LANDUSE-LANDCOVER CLASSIFICATION OF WORLDVIEW-2 SENSOR DATA. December 2014.

[10] C. Efthymiou and G. Kalogridis. Smart Grid Privacy via Anonymization of Smart Metering Data. In *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 238–243, October 2010.

[11] John L. Harris and Lon-Mu Liu. Dynamic structural analysis and forecasting of residential electricity consumption. *International Journal of Forecasting*, 9(4):437–455, December 1993.

[12] Imre Janszky and Rickard Ljung. Shifts to and from Daylight Saving Time and Incidence of Myocardial Infarction. *New England Journal of Medicine*, 359(18):1966–1968, October 2008.

[13] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda. Privacy for Smart Meters: Towards Undetectable Appliance Load Signatures. In *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 232–237, October 2010.

[14] Thomas Kantermann, Myriam Juda, Martha Merrow, and Till Roenneberg. The Human Circadian Clock's Seasonal Adjustment Is Disrupted by Daylight Saving Time. *Current Biology*, 17(22):1996–2000, November 2007.

[15] Amir Kavousian, Ram Rajagopal, and Martin Fischer. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, 55:184–194, June 2013.

[16] Tuuli A. Lahti, Sami LeppÃďmÃďki, Jouko LÃűnnqvist, and Timo Partonen. Transition to daylight saving time reduces sleep duration plus sleep efficiency of the deprived sleep. *Neuroscience Letters*, 406(3):174–177, October 2006.

[17] Mats Lambe and Peter Cummings. The shift to and from daylight savings time and motor vehicle crashes. *Accident Analysis & Prevention*, 32(4):609–611, July 2000.

[18] Dengsheng Lu and Qihao Weng. Urban Classification Using Full Spectral Information of Landsat ETM+ Imagery in Marion County, Indiana. *Photogrammetric Engineering & Remote Sensing*, 71(11):1275–1284, November 2005.

[19] Ulrike Von Luxburg. A Tutorial on Spectral Clustering. 2007.

[20] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, September 2003.

[21] Fintan McLoughlin, Aidan Duffy, and Michael Conlon. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, 141:190–199, March 2015.

[22] AndrÃľs Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. Private Memoirs of a Smart Meter. In *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 61–66, New York, NY, USA, 2010. ACM.

[23] City of Turin. AperTO - Gli OpenData della Citta di Torino.

[24] Gang Pan, Guande Qi, Zhaohui Wu, Daqing Zhang, and Shijian Li. Land-Use Classification Using Taxi GPS Traces. *IEEE Transactions on Intelligent Transportation Systems*, 14(1):113–123, March 2013.

[25] Agenzia Regionale per la Protezione Ambientale. Il Clima in Piemonte nel 2015, January 2016.

[26] Elias Leake Quinn. Privacy and the New Energy Infrastructure. SSRN Scholarly Paper ID 1370731, Social Science Research Network, Rochester, NY, February 2009.

[27] Jonathan Reades, Francesco Calabrese, and Carlo Ratti. Eigenplaces: analysing cities using the space âĂŞ time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36(5):824–836, 2009.

[28] Sarah L. Richardson and John Crompton. Vacation patterns of French and English Canadians. *Annals of Tourism Research*, 15(3):430–435, January 1988.

[29] Jameson Toole, Michael Ulm, Dietmar Bauer, and Marta Gonzalez. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 2012.

[30] Jason Varughese and Richard P Allen. Fatal accidents following changes in daylight savings time: the American experience. *Sleep Medicine*, 2(1):31–36, January 2001.