Energy Policy 73 (2014) 607–619

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Using smart meter data to estimate demand response potential, with application to solar energy integration

Mark E.H. Dyson*, Samuel D. Borgeson, Michaelangelo D. Tabone, Duncan S. Callaway

Energy & Resources Group University of California, Berkeley 10 Barrows Hall Berkeley, CA 94720, USA

HIGHLIGHTS

- We use hourly electricity use data to estimate residential demand response (DR) potential.
- The residential cooling DR resource is large and well-matched to solar variability.
- Customer heterogeneity is large; programs should target high potential customers.

ARTICLE INFO

Article history:
Received 20 February 2014
Received in revised form
28 May 2014
Accepted 29 May 2014
Available online 20 June 2014

Keywords:
Demand response
Smart meter data
Renewable integration

ABSTRACT

This paper presents a new method for estimating the demand response potential of residential air conditioning (A/C), using hourly electricity consumption data ("smart meter" data) from 30,000 customer accounts in Northern California. We apply linear regression and unsupervised classification methods to hourly, whole-home consumption and outdoor air temperature data to determine the hours, if any, that each home's A/C is active, and the temperature dependence of consumption when it is active. When results from our sample are scaled up to the total population, we find a maximum of 270–360 MW (95% c.i.) of demand response potential over a 1-h duration with a 4 °F setpoint change, and up to 3.2–3.8 GW of short-term curtailment potential. The estimated resource correlates well with the evening decline of solar production on hot, summer afternoons, suggesting that demand response could potentially act as reserves for the grid during these periods in the near future with expected higher adoption rates of solar energy. Additionally, the top 5% of homes in the sample represent 40% of the total MW-hours of DR resource, suggesting that policies and programs to take advantage of this resource should target these high users to maximize cost-effectiveness.

1. Introduction

Variable renewable energy technologies including wind turbines and solar photovoltaics (PV) pose challenges for traditional modes of operating electricity systems, as they decrease the controllability of electricity supply and introduce new sources of forecast error (Makarov et al., 2009). The introduction of California's Renewable Portfolio Standard in 2012, requiring the state to get 33% of its electricity from renewable sources by 2020, will likely increase the share of these resources in the California electricity system, and may require substantial new capacity of flexible resources to ensure reliable electricity service (KEMA, 2010). Currently, most power system flexibility in California comes from power plants fueled by natural gas, and thus these are the resources forecast to be required in a more highly renewable future system (KEMA, 2010; Wei et al., 2013). However, several studies have recently proposed that increased grid flexibility can potentially come from demand response (DR), defined by the Federal Energy Regulatory Commission as changes in the pattern of end-use electricity consumption over time in response to grid conditions (FERC Staff, 2009; Kirby, 2007; Eto et al., 2009; Bode et al., 2013). These and other studies suggest that variability can be mitigated by controlling end-use load, especially air conditioning (A/C) equipment, potentially reducing the need for flexible gas generators to balance variation in renewable production. The specific grid need that DR may be able to meet is the evening solar down-ramp, where PV production declines just as load is peaking, and substantial flexibility is required to match the ramp rate of system load net of solar generation (Denholm and Margolis 2007; Mileva et al. 2013).

There is a large body of research relevant to the current focus on estimating the potential of air conditioning demand response in power systems. Research in economics lays the statistical groundwork for empirical identification of the impact of temperature on building

* Corresponding author.
E-mail addresses: mehd@berkeley.edu (M.E.H. Dyson), sborgeson@berkeley.edu (S.D. Borgeson), m.tabone@berkeley.edu (M.D. Tabone), dcal@berkeley.edu (D.S. Callaway).

http://dx.doi.org/10.1016/j.enpol.2014.05.053
0301-4215/

energy consumption (Henley and Peirson 1997; Aroonruengsawat and Maximilian 2009; Deschênes and Greenstone 2011). Work in the building science literature has applied similar statistical models to hourly data in order to identify the impacts of energy efficiency retrofits on temperature-dependent loads in buildings and/or disaggregate end-uses from whole-building consumption data (Schick et al., 1988; Kissock, Reddy, and Claridge 1998; Kissock, Haberl, and Claridge 2002; Raffio et al., 2007; Mathieu et al., 2011; Birt et al., 2012). Finally, there is a growing body of literature that aims to identify power system-scale impacts of demand response programs, primarily focused on peak load reduction and ancillary service provision(Kirby, 2007; Eto et al., 2009; Bode et al., 2013; Cappers and MacDonald 2013), with a few studies specifically focused on the implications for renewable energy integration (Callaway, 2009; Cappers et al., 2011; Mathieu, Dyson, and Callaway 2012).

This paper uses a newly available dataset coupled with empirical techniques to identify the DR resource present in residential air conditioners in Northern California, and examine this resource's potential for mitigating renewable variability. We use a dataset containing hourly consumption from 30,000 residential smart meters in the Pacific Gas & Electric (PG&E) service territory in central and northern California to estimate hourly DR potential in each building, for each hour of the studied year (2011).

The specific contributions of this paper are to extend the methodology of prior studies by employing an empirical approach that estimates (1) whether and during which hours of the year A/C is active in each building and (2) the expected power reduction in response to a broadcast set-point change. With these estimates for each building and load & solar generation data from the California Independent System Operator (CAISO), our final contribution is to (3) compare the aggregate DR resource to coincident patterns of renewable resource variability in California, and identify the level of DR resource heterogeneity among potential program participants in order to understand the importance of targeting specific customers for program participation.

Section 2 details our analysis approach and the data sources used. We present the results of our analysis in Section 3, starting with building-specific regression results and moving to the aggregated resource in the context of renewable integration. We introduce the implications of the results and potential sources of bias in Section 4, and conclude in Section 5 with policy implications and directions for future research.

2. Methods and data used

The basic approach we will use to characterize DR potential is to identify the relationship between outside air temperature and building electrical load in individual buildings. The central distinction between our approach and related work is that we first identify presence of air conditioner activity (by hour) and estimate cooling responsiveness only for the hours in which cooling activity is present. This allows us to drop hours when temperature may be high but for scheduling reasons the building is not cooling. To identify demand response potential we will make the central assumption that raising indoor temperature (e.g. raising the setpoint) by 1° will have the same impact as lowering the outdoor temperature by 1°. Thus, a model that estimates the impact of outside air temperature on electricity consumption can also be interpreted as estimating the impact of changing the building's thermostat setpoint.

2.1. Data sources

We use three data sources in this study: hourly interval electricity consumption data from residential buildings in Pacific Gas & Electric's (PG&E) service territory, weather data downloaded from Weather Underground, and California grid data obtained from CAISO. This section outlines these data sources and their use for this project.

2.1.1. Interval consumption data

The main dataset used in this study contains 1 h interval electricity consumption data representing up to four years of data from a random sample of 30,000 residential accounts in PG&E's service territory from 2008 through 2011. The data were provided confidentially by PG&E via the Wharton Customer Analytics Initiative[1].

The primary variables from the PG&E data used in the present study are the interval electricity consumption data and location data (5-digit zip code) associated with each smart meter. We clean the data in the following ways:

1. Interval consumption data before 2011 are dropped, to reduce the impact of unobserved year-to-year changes in occupancy and/or non-AC consumption pattern changes on our model estimates. Controlling for these changes may be possible by using a fixed effects approach, but that would introduce new challenges (e.g. identifying the correct time period to add the fixed effects) and, given the quantity of observations in 2011 alone, would not add meaningfully to the power of the statistical model.
2. Buildings with very high levels of consumption (>100 kW average demand) are removed from the data – these are likely mislabeled commercial facilities or very large residential (apartment) buildings, and we do not wish to mix results for this type of building with those of the remainder of the buildings; however, we acknowledge that this remainder of buildings is likely to include large, multi-family units and apartment buildings, given the 100 kW load cutoff. These buildings are included in the analysis; they are also viable targets for DR programs that utilize A/C equipment similar to that found in the single-family homes that make up the majority of the sample.
3. Buildings with missing data from 2011 are left as-is, however the models we use with these buildings are more likely to lack sufficient observations to successfully estimate model parameters, since the regression code automatically drops rows with missing data.

The spatial extent and density of observations are plotted in Fig. 1.

2.1.2. Hourly weather data

We use hourly weather data from Weather Underground, an internet service that archives weather data from thousands of stations in California and the rest of the country[2]. Using the zip code information from each account premise in the PG&E data, we downloaded weather data from up to seven weather stations within 10 km of the center of that zipcode, sub-sampled the data down to an hourly interval by averaging readings within each hour, and averaged the hourly readings across stations to come up with an estimate of zip code-level outdoor air temperatures at each hour. The buildings in each of 716 zip codes in the dataset are assigned that zip codes calculated hourly temperature. The regression drops hours with missing temperature values.

[1] http://www.wharton.upenn.edu/wcai/
[2] www.wunderground.com