Taylor & Francis
Taylor & Francis Group

International Journal of Sustainable Energy

ISSN: 1478-6451 (Print) 1478-646X (Online) Journal homepage: https://www.tandfonline.com/loi/gsol20

Residential load and rooftop PV generation: an Australian distribution network dataset

Elizabeth L. Ratnam, Steven R. Weller, Christopher M. Kellett & Alan T. Murray

INTERNATIONAL JOURNAL OF SUSTAINABLE ENERGY, 2017
VOL. 36, NO. 8, 787–806
http://dx.doi.org/10.1080/14786451.2015.1100196

Residential load and rooftop PV generation: an Australian distribution network dataset

Elizabeth L. Ratnam[a], Steven R. Weller[a], Christopher M. Kellett[a] and Alan T. Murray[b]

[a]School of Electrical Engineering and Computer Science, University of Newcastle, University Drive, Callaghan NSW 2308 Australia; [b]Independent, Newcastle, Australia

ABSTRACT

Despite the rapid uptake of small-scale solar photovoltaic (PV) systems in recent years, public availability of generation and load data at the household level remains very limited. Moreover, such data are typically measured using bi-directional meters recording only PV generation in excess of residential load rather than recording generation and load separately. In this paper, we report a publicly available dataset consisting of load and rooftop PV generation for 300 de-identified residential customers in an Australian distribution network, with load centres covering metropolitan Sydney and surrounding regional areas. The dataset spans a 3-year period, with separately reported measurements of load and PV generation at 30-min intervals. Following a detailed description of the dataset, we identify several means by which anomalous records (e.g. due to inverter failure) are identified and excised. With the resulting 'clean' dataset, we identify key customer-specific and aggregated characteristics of rooftop PV generation and residential load.

ARTICLE HISTORY
Received 2 July 2015
Accepted 11 September 2015

KEYWORDS
Solar photovoltaics; residential load; dataset; time-of-use metering; feed-in tariffs

1. Introduction

In many countries, electricity distribution operators have installed metering equipment to separately record small-scale ('rooftop') photovoltaic (PV) generation and residential load. To incentivise the installation of rooftop PV the New South Wales (NSW) state government in Australia introduced the Solar Bonus Scheme (NSW Government 2010) on 1 January 2010, wherein customers were offered generous feed-in tariffs for grid-connected PV generation. As a consequence, the NSW-based utility Ausgrid installed additional metering infrastructure at premises of each eligible customer to enable the recording of power flows from PV inverters for the purpose of calculating feed-in tariff payments.

The financial policy of 'net metering' widely adopted in the USA also incentivises grid-connected rooftop PV generation (Black 2004; Campoccia et al. 2009; Darghouth, Barbose, and Wiser 2014; Ratnam, Weller, and Kellett 2015a,b). To offer the financial policy of net metering, an electricity distributor typically installs a bi-directional meter at the premises of each eligible customer. The bi-directional meter records any PV generation in excess of residential load for the purpose of calculating net metering credits. The bi-directional meter also records any load drawn from the electricity grid (i.e. load not met by the rooftop PV generator) for the purpose of calculating net metering bills (Black 2004; Campoccia et al. 2009; Ratnam, Weller, and Kellett 2015b). In contrast, implementing the feed-in tariff proposed under the NSW Solar Bonus Scheme requires two meters to be

CONTACT Elizabeth L. Ratnam elizabeth.ratnam@ieee.org

installed at residential premises. A dedicated meter recording solar generation direct from the PV inverter and a separate meter recording residential load are installed at premises for the purpose of subsidising and billing customers, respectively. A consequence of the NSW Solar Bonus Scheme is, therefore, more complete PV generation data available for collection.

Datasets with separate (more complete) records of load and PV generation are useful in the area of forecasting supply and demand in the electricity grid. To ensure generation supply meets electricity demand and that the power supplied to customers is of a high quality, network planners, and operators rely on forecasts of both system load and generation – including distributed PV generation (Holttinen et al. 2013). However, accurate predictions of fluctuations in day-ahead PV generation at a system level is an active research topic (Hart, Stoutenburg, and Jacobson 2012; Holttinen et al. 2013; Chow, Belongie, and Kleissl 2015). For example, fluctuations in aggregate residential PV generation as observed in an upstream feeder typically arise from moving cloud cover on timescales ranging from minutes to hours (Hart, Stoutenburg, and Jacobson 2012; Chow, Belongie, and Kleissl 2015). In contrast, fluctuations in aggregate load as observed in an upstream feeder are fairly slow and predictable for the day-ahead (Hart, Stoutenburg, and Jacobson 2012). Historical datasets with separately reported measurements of PV generation potentially assist in the development of forecasting algorithms that predict PV generation from prior PV measurements.

The recent dramatic increase in grid-connected solar PV has been driven by the decreasing cost of PV panels (Bazilian et al. 2013; Lang, Gloerfeld, and Girod 2015) together with generous feed-in tariffs and/or net metering polices (Ogimoto et al. 2013; Moosavian et al. 2013; von Appen et al. 2013). In Australia, Germany, and the USA, net metering policies and/or feed-in tariffs have been available to residents in recent years. As a consequence, in Australia, there was a 480% increase in solar PV installations in a single year from 2009 to 2010, of which 99% was grid-connected (Moosavian et al. 2013). In the USA, there is more than 16 GW of installed solar PV (Kroposki and Mather 2015), up from 0.8 GW in 2010 (Kroposki, Margolis, and Lynn 2011). PV plant installations in Germany exceed 1.2 million, and as of September 2012, peak PV capacity reached 31 GW with about 70% of this capacity being connected to the low voltage grid (von Appen et al. 2013).

The uptake of residential solar PV in NSW Australia was greater than expected under the generous Solar Bonus Scheme (Independent Pricing and Regulatory Tribunal 2012). In response, the NSW state government initiated a review of feed-in tariff prices to manage the cost of the Scheme together with encouraging further adoption of renewable energy in NSW (Ausgrid 2014e; Independent Pricing and Regulatory Tribunal 2012). As a consequence of this review, Ausgrid publicly released a small sample of PV data collected for feed-in tariff payments (Ausgrid 2014e). More specifically, in 2014 Ausgrid publicly released residential PV generation and load data for a subset of 300 (de-identified) customers spanning a 3-year period from 1 July 2010. Other utilities offering feed-in tariffs potentially store data similar to that released by Ausgrid (Nykamp et al. 2013; Independent Pricing and Regulatory Tribunal 2012). Furthermore, smart grid deployments in some countries have also provided opportunities for data collection relating to solar PV generation and/or residential load (Rhodes et al. 2014; Yang et al. 2014; Quilumba et al. 2015; Pereira et al. 2015). However, the public availability and analyses of these load and generation datasets in the open literature is very limited. The Ausgrid dataset, therefore, is a valuable resource for researchers and policy-makers alike.

In this paper, we present a detailed description of the Ausgrid dataset in Section 2 with a view to facilitating use of the dataset by interested researchers.[1] The Ausgrid dataset has already been used in several research publications investigating co-locating battery storage with solar PV (Ratnam, Weller, and Kellett 2013; Keerthisinghe, Verbic, and Chapman 2014a,b; Khalilpour and Vassallo 2015; Braun et al. Forthcoming; Worthmann et al. 2015; Ratnam, Weller, and Kellett 2015a,b), and this paper aims to serve as an archival reference for future research. In Section 3, we describe various anomalies that arise in the dataset. We remove customers with *any* anomalous data, leaving a 'clean dataset' consisting of reliable data for 54 customers for the entire 3 years covered by the dataset. It is worth noting that, for any given day, significantly more than the 54 customers in the clean dataset will have reliable data.[2] An analysis of the clean dataset is presented in Section 4.

2. Ausgrid dataset

The Ausgrid distribution network covers 22,275 km^2 and includes load centres in Sydney and regional NSW as depicted in Figure 1. The Ausgrid distribution network supplies in excess of 25,523 GWh of electricity annually to more than 1.64 million customers (2013/2014 financial year), comprised of over 1.4 million residential customers together with major industries including mining, manufacturing, and agriculture (Ausgrid 2014a). From the Ausgrid customer base, a subset of 300 residential customers were chosen as follows:

(1) Ausgrid identified residential customers with a separate meter that recorded PV generation directly from the PV inverter (i.e. customers receiving a feed-in tariff) over the period 1 July 2010–30 June 2013. Approximately 15,000 customers were included in this group (Ausgrid 2014b).
(2) From this group of around 15,000 customers, Ausgrid removed customers in the top or bottom 10% of annual household energy consumption (in kWh) or PV production (in kWh).

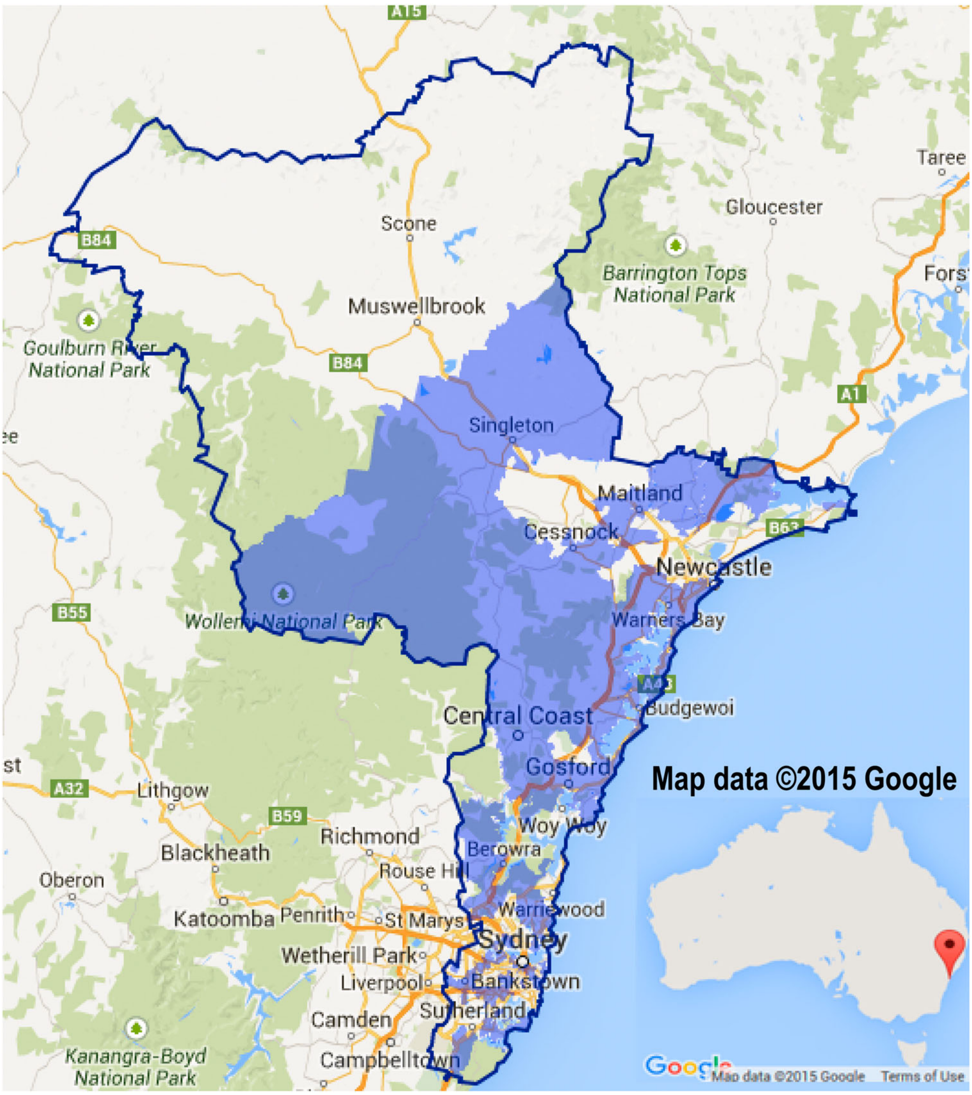

Figure 1. Outline of the Ausgrid distribution network that covers 22,275 km^2 and includes load centres in Sydney and regional NSW. Shaded regions within the Ausgrid network correspond to postcode areas included in the dataset of 300 customers (ABS 2006). Map data: ©2015 Google

(3) From the remaining residential customers, 300 were selected at random by Ausgrid.

To protect the privacy of these 300 residential customers, all personally identifiable information was removed from the Ausgrid dataset (Ausgrid 2014d). Consequently, the postcode of each customer provides the only context of geographical spread included in the Ausgrid dataset. In total, there are 100 unique postcodes in the Ausgrid dataset. The shaded regions in Figure 1 depict these 100 postcode areas within the Ausgrid boundary. Information regarding the relationship between socio-economic or demographic characteristics in the context of solar PV uptake in different areas of Australia is provided in ACIL Allen Consulting (2013).

By default, the time format in the Ausgrid dataset is Australian Eastern Standard Time (AEST), though for the summer period (approximately October–April) Australian Eastern Daylight Savings Time (AEDT) is used. For each of the 1096 days in the Ausgrid dataset (3 years from 1 July 2010), load and generation data were recorded at intervals of $\Delta = 0.5$ h duration. In what follows, we introduce terminology provided by Ausgrid (2014e) that describes the daily load and generation data recorded against each of the 48 intervals of 0.5 h duration. We also introduce additional terminology, where appropriate, with a view of facilitating use of the Ausgrid dataset by interested researchers.

2.1. *Residential load and generation data*

The Ausgrid dataset includes information on the installed capacity for each of the 300 residential-scale rooftop solar PV units (in kilowatt-peak kWp). The solar PV generation data for each customer in the Ausgrid dataset is obtained with a meter installed to measure solar generation for feed-in tariff payments.

The meter recording solar generation operates in gross metering mode, in which power flow measurements are recorded in a single direction (e.g. from the PV inverter) (Ratnam, Weller, and Kellett 2015a). In the Ausgrid dataset, residential PV generation (in kWh) is referred to as *Gross Generation* (GG), recorded at the conclusion of each half hour interval. More specifically, each customer in the Ausgrid dataset has 48 GG entries on each day, where each entry represents the energy produced by the PV panel over the preceding half hour, as seen by the meter. We denote by $\mathrm{gg}(j)$ the conversion of the daily half hour interval GG entries, to an average power (i.e. PV generation in kW) over the half hour period $((j-1)\Delta, j\Delta)$, where j denotes a time index.

A separate meter operating in gross metering mode is located at each residential premises to measure and record energy consumption. Residential energy consumption is billed according to a time-of-use or an inclining block rate, where each customer selects their respective preference (Ausgrid 2014c). In the Ausgrid dataset, residential energy consumption (in kWh) is referred to as *General Consumption* (GC), recorded at the conclusion of each half hour interval. More specifically, each customer has 48 GC entries on each day, where each entry represents the energy consumed by the customer over the preceding half hour. We denote by $\mathrm{gc}(j)$ the conversion of the daily half hour interval GC data to an average power (i.e. load in kW) over the half hour period $((j-1)\Delta, j\Delta)$, where j denotes a time index.

A third meter that measures and records *Controllable Load* (CL) associated with water heating is located in 137 of the 300 residential premises. That is, 137 of the 300 residents in the Ausgrid dataset allow the utility to control their all-electric-heated water systems for periods in the day (in a manner that ensures minimal impact to the network) given a financial incentive. The two financial incentives offered to customers are referred to as an off-peak 1, or an off-peak 2 tariff. The off-peak 1 tariff is offered to customers that allow the utility to switch 'off' their all-electric-heated water systems for 18 h per day. The more expensive off-peak 2 tariff is offered to customers that allow the utility to switch 'off' their all-electric-heated water systems for 8 h per day. While the precise tariff is not shown in the dataset, customer-specific switching times associated with off-peak 1 and off-peak 2 tariffs may be inferred in the CL dataset, where residential CL data (in kWh) is recorded after each half hour interval on each day. We denote by $\mathrm{cl}(j)$ the conversion of daily half hour interval CL data, to an average power (i.e. load in kW) over the period $((j-1)\Delta, j\Delta)$, where j denotes a time index.

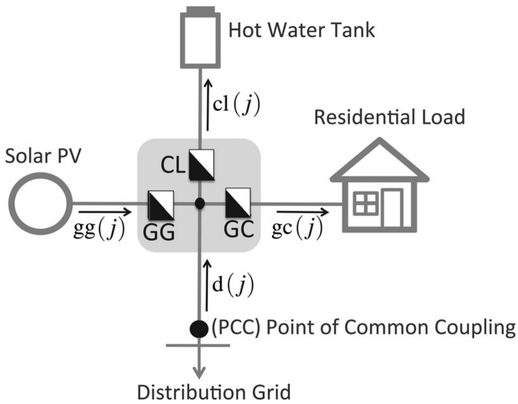

Figure 2. Residential system illustrating the direction of positive power flows and meters recording gross generation (GG), general consumption (GC), and controllable load (CL) consumption. Arrows associated with gg(j), gc(j), cl(j), and d(j) illustrate the direction of positive power flow, and j denotes a time index. Power flows against the direction of the arrow associated with d(j) are, therefore, negative. The meters in the shaded region are co-located at residential premises.

The meter recording CL is co-located with the meters recording GG and GC, depicted by the shaded region in Figure 2. The demand or average power flow (in kW) from (to) the grid to (from) the residential system over the period $((j-1)\Delta, j\Delta)$ is denoted by $\mathrm{d}(j) > 0$ ($\mathrm{d}(j) < 0$), where j is a time index. The power balance equation for the residential energy system depicted in Figure 2 is

$$\mathrm{d}(j) = \mathrm{gc}(j) - \mathrm{gg}(j) + \mathrm{cl}(j), \tag{1}$$

which must hold for all time indices j.

Note that customers that do not have a controllable load do not have a meter recording CL data. In this case, the power balance equation in Equation (1) reduces to $\mathrm{d}(j) = \mathrm{gc}(j) - \mathrm{gg}(j)$ for all j.

A summary of the key Ausgrid dataset parameters is included in Table 1. The consumption category in Table 1 is consistent with terminology provided by Ausgrid (2014e).

2.2. *Daylight saving time*

We present observations when moving from AEST to AEDT (or vice versa). This time change is commonly referred to as the start (or end) of 'daylight saving time'. On 3 October 2010 (the start of daylight saving time), we observe 0 entries from 2 am to 3 am in the respective GC and CL fields. Further, PV generation (in the GG field) is time-shifted back an hour. On 3 April 2011 (the end of

Table 1. Key Ausgrid dataset parameters.

Area of the Ausgrid network	22,275 km^2
Number of customers	300
Number of unique postcodes	100
Date: 1 July 2010 to the 30 June 2013	1096 days
Time format	AEST or AEDT
A single day: T	24 h from midnight
Number of daily intervals: s	48
Interval period: Δ	$\Delta = T/s = 0.5$ h
Generator capacity (in kWp)	Tested capacity of each PV unit
Consumption category: CL (in kWh)	Controllable load consumption
Consumption category: GC (in kWh)	General consumption
Consumption category: GG (in kWh)	Gross generation
Number of customers with CL data	137
Number of customers with GC data	300
Number of customers with GG data	300

daylight saving time), we do not observe an extra hour of data in any of the data-fields, GG, GC, CL, respectively. We therefore assume that turning back the clock over-writes the previous hour of data.

During October of each year, CL data entries vary due to the commencement of daylight saving time. On 4 October 2010, the hot water systems of some customers are switched on an hour later, others are not. On the last week of October, some of the CLs are switched on an hour earlier, consistent with the commencement in daylight saving time prior to 2007.[3] We recommend CL switching times be inferred from the CL data on each day, for each customer. When approximating these switching times, note that each CL is switched 'on' minutes or hours apart in a manner that ensures minimal impact to the network.

3. Clean dataset: methodology

In principle, the dataset in the previous section consists of load and generation data for each of 300 customers at 30-min resolution for a 3-year period. In practice, however, several factors exist which lead to anomalous measurements in the dataset (e.g. when a PV inverter fails). In this section, we describe how these anomalous measurements are identified and subsequently excised, producing a 'clean dataset '. Moreover, our approach is very conservative as we place a higher value on the quality of the clean dataset rather than the quantity of data records. Further work to qualify anomalous data records removed in this section is certainly possible.

3.1. *Residential load: GC*

We remove customers with anomalous load recordings on any day in a 3-year period that potentially arise when customers go on holidays and disconnect (i.e. switch off) most appliances. More specifically, we remove customers with a maximum load less than 6 W on any day of the year (i.e. any day where $gc(j) < 0.006$ for all $j = 1, \ldots, s$). Customers removed often have data suitable for analysis on many days of each year, but not all. The customers removed based on this criterion are presented in Table 2.

In Figure 3, we present a small sample of days that contributed to four of the seven customers being removed from the Ausgrid dataset. That is, each subplot in Figure 3 corresponds to one of the four customers, where the daily load of each customer is presented on three separate occasions.

In Figure 3(a) and 3(c), we observe that Customer 9 and Customer 221, respectively, have days where no load is recorded. In contrast, Customer 191 and Customer 229 in Figure 3(b) and 3(d), respectively, have days with very low load recordings (i.e. 2 W). These daily low load (or no load) recordings potentially indicate that a customer is on holidays on the respective day.

The residential load of the remaining 293 customers is presented in Figure 4 in an aggregated form. We sum the load $gc(j)$ of each customer with reference to time index j, and we repeat this process for each time index j on each of the 1096 days, with the results presented in Figure 4(a). Each year in Figure 4(a) (on the x-axis) denotes 365 days (or 366 days during a leap year) from 1 July of the previous year. To present a load duration curve, we sort the aggregated data presented in Figure 4(a) across the complete set of 1096 days, which is depicted in Figure 4(b).

In Figure 4(a), we observe a peak load of 828 kW, which occurs during the 2010–2011 summer. This peak occurred during the first week in February 2011, one of the warmest days (i.e. high ambient temperatures) on record in Sydney (Australian Government Bureau of Meteorology 2011). During the 2012–2013 summer, the peak load in Figure 4(a) is 730 kW, which is again significantly greater than the winter peaks of each year. We also observe three occasions where the aggregated load is zero, consistent with the change from AEST to AEDT. In Figure 4(b), we observe that the aggregate load is below 400 kW at least 98% of the time.

Table 2. Anomalous load data: Customer ID.

Customer ID	9	121	150	191	221	229	260

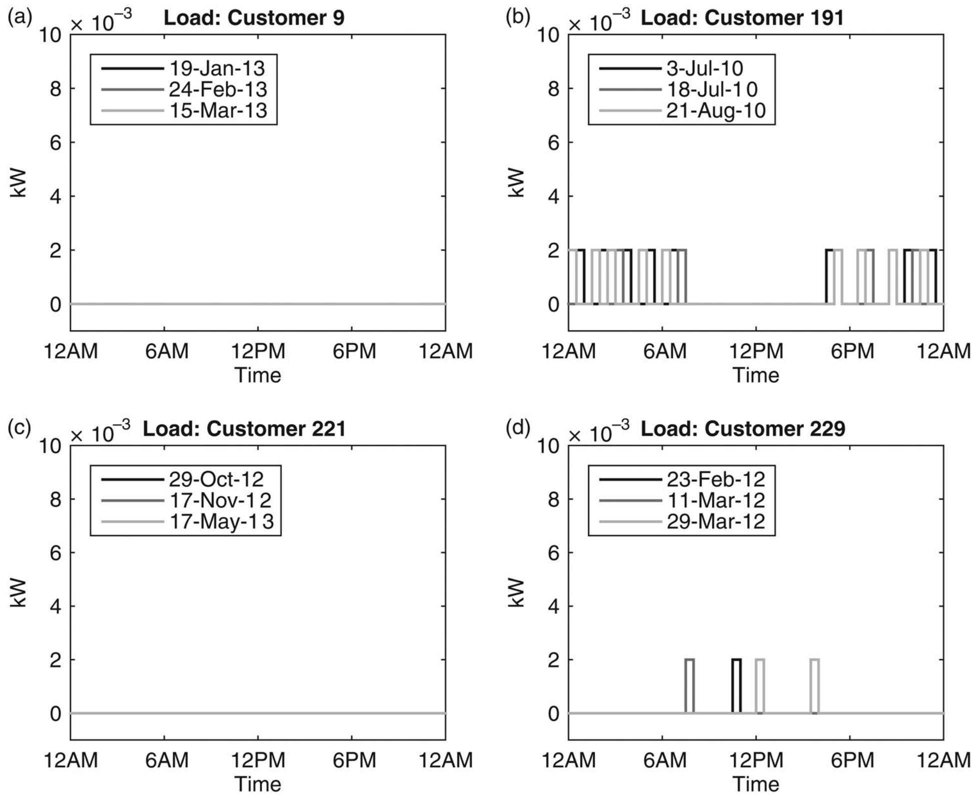

Figure 3. Example days and associated load consumption contributing to the removal of four customers from the Ausgrid dataset.

3.2. *Residential generation: GG*

In this section, we remove customers with anomalous generation recordings on any day in the 3-year period. Each customer removed often has data suitable for analysis on many days of each year, but not all. Customers removed have PV generation data that falls into one of the following three categories:

- Category 1 includes customers with a maximum generation less than 0.06 kW on any day of each year (i.e. any day where $gg(j) < 0.06$ for all $j = 1, \dots, s$). Data removed by this category potentially arise when PV inverters fail, or generation production is very low. A consequence of this category is the removal of 209 customers from the Ausgrid dataset.
- Category 2 includes customers with a daily generation less than 0.325 kWh and a maximum generation less than 0.101 kW on any of the 1096 days (i.e. any day where $gg(j) < 0.101$ for all $j = 1, \dots, s$, and $gg(1) + \cdots + gg(s) \leq 0.65$ kW). Data removed by this category potentially arise when daily PV generation profiles fall below a threshold whereby there is significant uncertainty regarding the quality of the data. Further work to confirm the quality of data removed by this category is certainly possible. A consequence of this category is the removal of 191 customers from the Ausgrid dataset, of which 36 customers are in addition to those customers identified in Category 1.
- Category 3 includes customers that generate more than 0.02 kWh during the early morning before 5 am on any day of each year (i.e. any day where $gg(1) + \cdots + gg(10) > 0.04$ kW). Data removed by this category potentially arise when measurement errors exist.[4] A consequence of this category is the removal of six customers from the Ausgrid dataset, of which one customer (ID 248) is in addition to those customers identified with Categories 1 and 2.

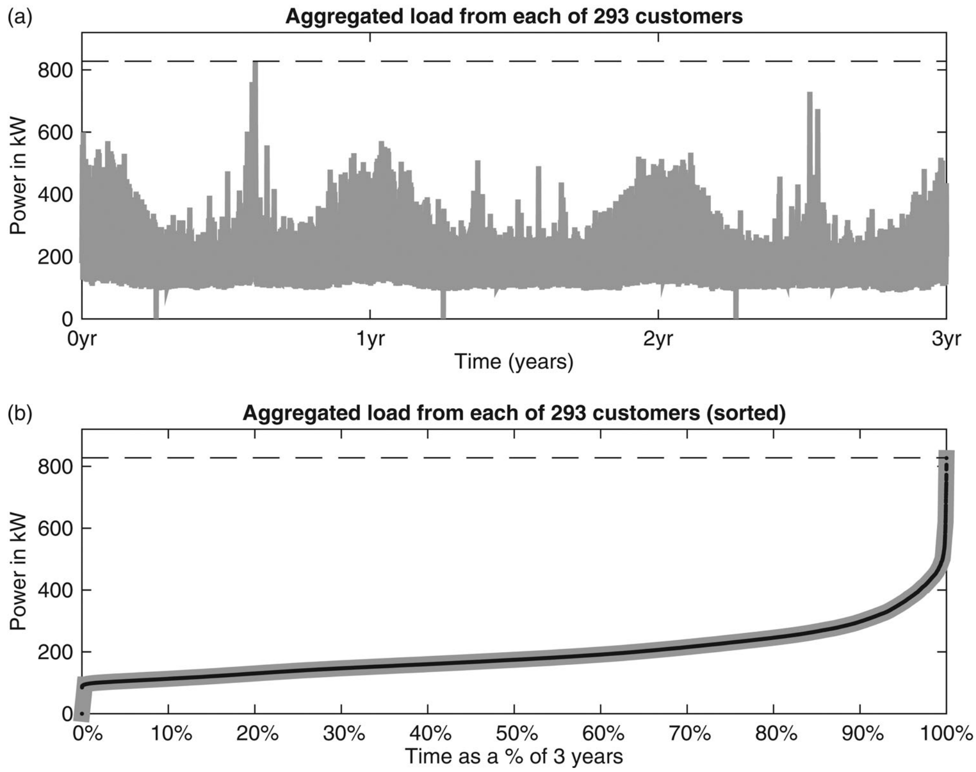

Figure 4. Aggregated load of 293 customers over 1096 days: (a) the aggregated load is presented chronologically from 1 July 2010 and (b) the aggregated load is sorted to obtain the load duration curve.

We present a small sample of days that contributed to 4 of the 246 customers being removed from the Ausgrid dataset. Each subplot in Figure 3 corresponds to one of the four customers, where the daily generation of each customer is presented on three separate occasions.

In Figure 5(a), we observe that Customer 1 has days where no generation is recorded, and these days were identified via a Category 1 removal. In Figure 5(c), we observe that Customer 215 has days where very little generation is recorded, and these days were identified via a Category 2 removal. In contrast, both Customer 145 and Customer 248 have days with a very small amount of generation recorded in the early morning before 5 am, and these days were identified via a Category 3 removal. A careful examination of Figure 5(b) and 5(d), respectively, is required to observe the very small amount of PV generation recorded in the early morning before 5 am for both Customer 145 and Customer 248.

The *clean dataset* (defined later in Section 3.4) includes the PV generation and GC of the remaining 54 customers. That is, 54 customers did not present load or generation abnormalities as outlined above. The aggregate PV generation of these 54 customers is presented in Figure 6(a), where each year on the x-axis denotes 365 days (or 366 days during a leap year) from 1 July of the previous year. More specifically, in Figure 6(a), we sum the generation $gg(j)$ of each customer with reference to time index j, and we repeat this process for each time index j on each of the 1096 days. In Figure 6(b), we sort the aggregated generation in Figure 6(a) across the complete set of 1096 days.

We observe PV generation peaks at approximately 110 kW during each summer period in Figure 6(a). We observe in Figure 6(b), the PV units generate electricity 45% of the time (or do not generate power 55% of the time), consistent with the daily variability of solar irradiance.

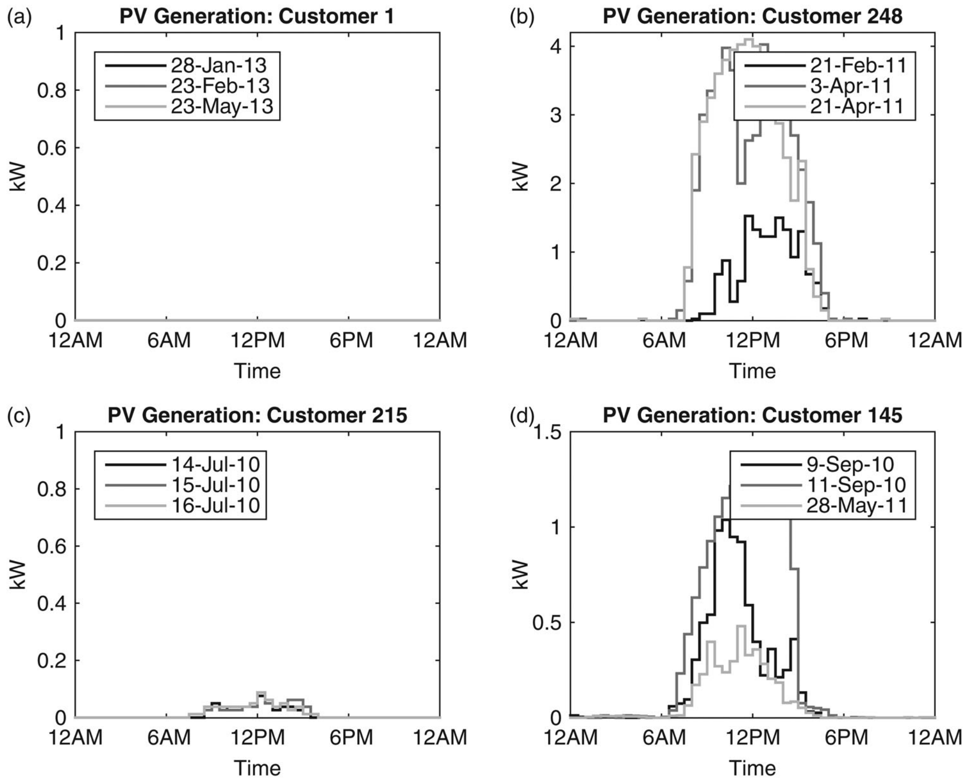

Figure 5. Example days and associated PV generation contributing to the elimination of four customers from the Ausgrid dataset.

3.3. *Controllable load*

In identifying the clean dataset, we ignored anomalous CL data. In this subsection, we look to assist the interested researcher seeking to identify anomalous CL data in the Ausgrid dataset.

The removal of utility controlled all-electric heated water systems often occurs when a customer upgrades to a gas-heated, solar-heated, or a heat pump water system. Customers that potentially remove utility controlled all-electric heated water systems are identified in the Ausgrid dataset via the CL field. That is, days where a customer allows a utility to control their all-electric heated water systems are identified via the presence of a daily CL field, and days where a customer removes permission for a utility to control their all-electric heated water systems are identified via the absence of a daily CL field. In the Ausgrid dataset, 11 customers with a daily CL field on a fraction of the 1096 days listed are included in Table 3.

Two additional customers not listed in Table 3 have potentially removed a utility controlled all-electric heated water system. That is, on each day no data are recorded against the CL field of Customer 37 and Customer 281 (with the slight exception of a single half hour interval entry for customer 37 on a single day). Hence, at least 13 customers have anomalous CL data in the Ausgrid dataset.

Further analysis to clean the CL dataset is certainly possible. For example, residential customers potentially reduce or increase the period of time a utility switches 'off' their all-electric heated water systems. This change in preference is reflected in the off-peak tariffs offered to residential customers.

Table 3. CL change: Customer ID.

ID	27	68	95	161	187	248	272	284	289	293	294

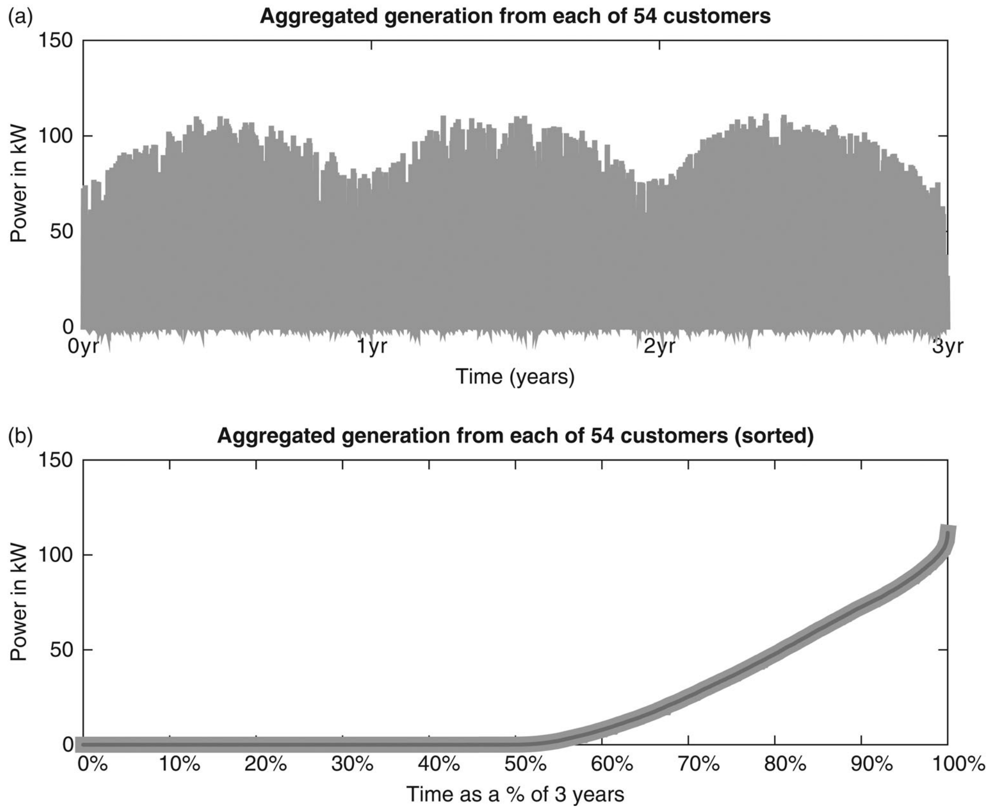

Figure 6. Aggregated generation of 54 customers over 1096 days: (a) the aggregate generation is presented chronologically from 1 July 2010 and (b) the aggregated generation is sorted to obtain the generation duration curve.

A change from an off-peak 1 to an off-peak 2 tariff (or vice versa) potentially leads to anomalous CL data within the Ausgrid dataset.

3.4. *Customer ID and location*

We define the *clean dataset* as the subset of 54 customers (from the total of 300), which are free of both load anomalies and PV generation anomalies. Note that 7 of the 300 customers had both load anomalies and PV generation anomalies. Customer 161 (cf. Table 3) is included in the clean dataset since we do not remove customers with anomalous CL. Customer 2 is included in the clean dataset since no anomalous load or generation recordings were identified; however, we note that from 12 October 2012 to 31 December 2012, data recordings were missing for this particular customer. We present the 28 postcode regions containing these 54 customers in Figure 7. The ID of each customer included in the clean dataset is presented in Table 4. In Figure 7, we observe that the majority of customers in the clean dataset are located in urban regions. Further, postcodes with smaller areas denote regions with higher population densities.

4. Clean dataset: analysis

In this section, we identify key characteristics of the clean dataset. We investigate aggregated demand, daily residential load and daily residential PV generation, and the orientation of 'rooftop' PV panels.

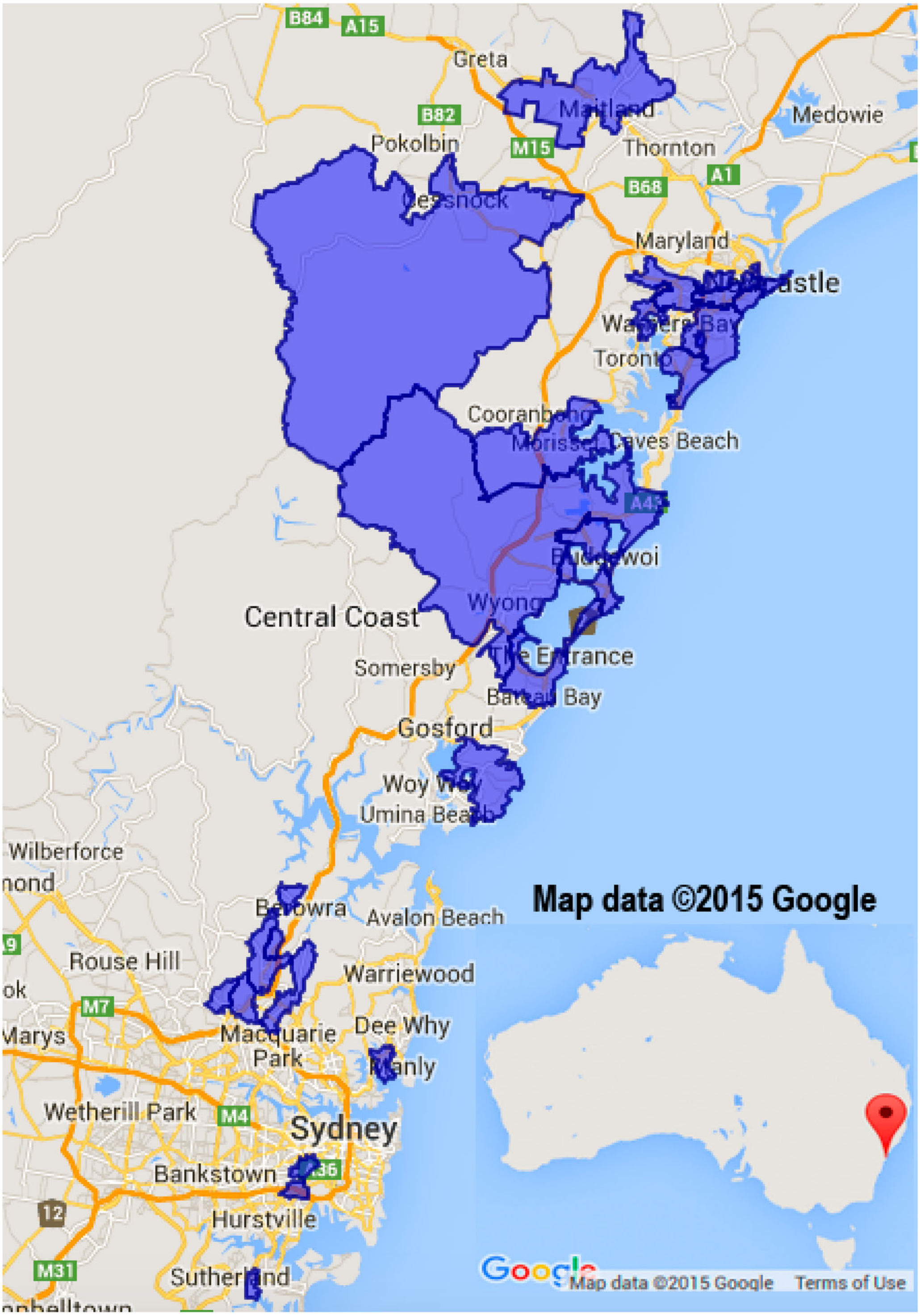

Figure 7. Each customer in the clean dataset belongs to a postcode region highlighted (ABS 2006). Map data: ©2015 Google

Table 4. Customer IDs in the clean dataset.

2	13	14	20	33	35	38	39	56
69	73	74	75	82	87	88	101	104
106	109	110	119	124	130	137	141	144
152	153	157	161	169	176	184	188	189
193	201	202	204	206	207	210	211	212
214	218	244	246	253	256	273	276	297

4.1. *Aggregated analysis*

The residential demand of 54 customers included in the clean dataset is presented in Figure 8 in an aggregated form. In Figure 8(a), we sum the demand $d(j)$ of each customer with reference to time index j, and we repeat this process for each time index j on each of the 1096 days. Each year on the x-axis commences on 1 July. In Figure 8(b), we sort the aggregated demand across the complete set of 1096 days.

In Figure 8(a), we observe the peak aggregated demand is 185 kW during the 2010–2011 summer. During the 2012–2013 summer, the peak aggregated demand is greater than the winter aggregated peak of each year. Aggregated negative demand that arises from surplus PV generation peaks at 85 kW. In Figure 8(b), we observe that the duration of the summer aggregated peak is less than 2% of the time, and aggregated negative demand (from surplus PV generation) occurs 23.4% of the time.

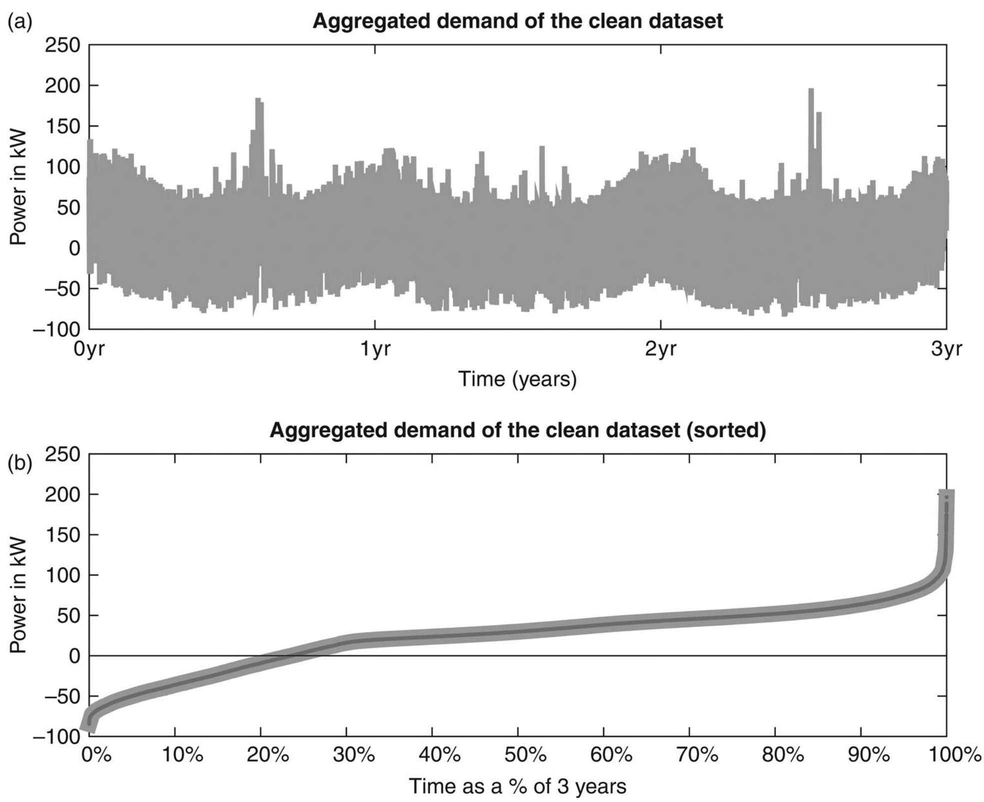

Figure 8. Aggregated demand of 54 customers over 1096 days: (a) the aggregate demand is presented chronologically from 1 of July 2010 and (b) the aggregated demand is sorted to obtain the demand duration curve.

To present the aggregated data in another form, we consider daily aggregate load and generation profiles, respectively, for each of the 54 customers in the clean dataset. In Figure 9, we present the daily mean and median aggregated generation profiles for summer and winter, respectively, with error bars that indicate one standard deviation about the mean. All references to standard deviation refer to one standard deviation ($\pm 1\sigma$) about a mean aggregate generation profile, or a mean aggregate load profile, where the context will make clear the intended meaning.

In Figure 9(a), we observe that the summer median aggregated generation is slightly greater than the summer mean aggregated generation during peak PV production (i.e. 85.3 kW compared to 76.7 kW). The summer PV generation occurs between 6.30 am and 7.30 pm. Also, the standard deviation in Figure 9(a) increases as PV production increases, highlighting that many days are impacted by cloud cover.

In Figure 9(b), we observe that the winter median aggregated generation profile is slightly greater than the winter mean aggregated generation profile during peak PV production (i.e. 72.7 kW compared to 62.3 kW). The winter PV generation occurs between 7 am and 5 pm. The standard deviation in Figure 9(b) increases as PV production increases. Thus, the results in Figure 9 are consistent with the availability of solar irradiance, which is greater in summer than winter and variable on each day.

In Figure 10, we present the daily mean and median aggregated load profile for summer and winter, respectively, with error bars that indicate one standard deviation about the mean. In Figure 10(a), we observe that the summer mean aggregated load is slightly greater than the summer median aggregated load during evening peak (i.e. 61.6 kW compared to 53.7 kW). Also, the summer residential load often peaks between 6 pm and 6.30 pm. The standard deviation in Figure 10(a) increases

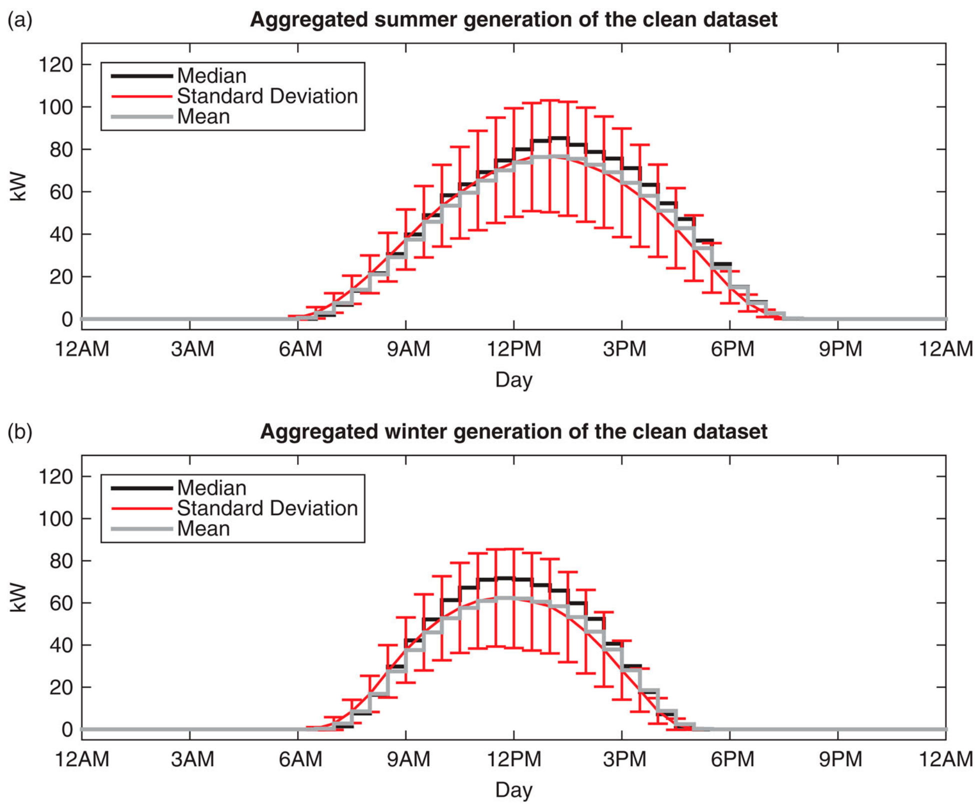

Figure 9. Aggregated generation: mean and median with error bars that indicate one standard deviation about the mean in (a) summer and (b) winter.

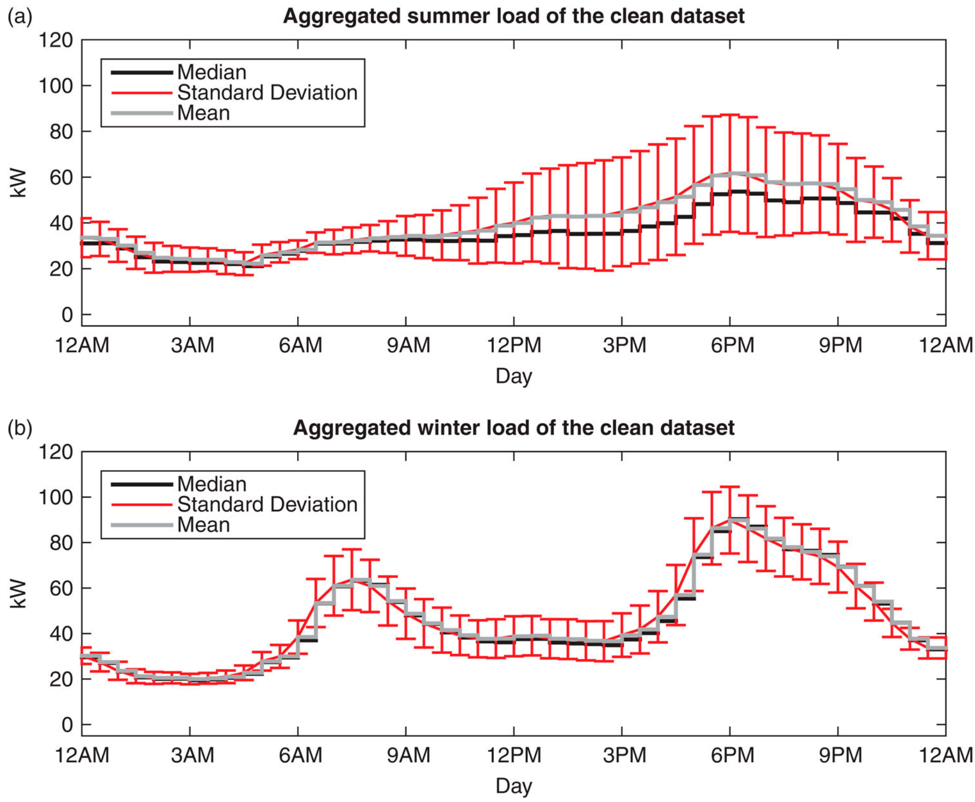

Figure 10. Aggregated load: mean and median with error bars that indicate one standard deviation about the mean in (a) summer and (b) winter.

from 9 am until the early evening when mean aggregate load peaks. In Figure 10(a), we observe that residential load from midnight till 9 am more closely corresponds to the time-of-day, since the standard deviation during this period is small.

In Figure 10(b), the winter mean aggregated load is slightly greater than the winter median aggregated load during the evening peak (i.e. 90.3 kW compared to 89.9 kW). We observe that residential loads often peak between 6 pm and 6.30 pm in winter. From midnight till 5 am, and from 10 am until midnight, the standard deviation in winter is clearly less than the standard deviation in summer, highlighting residential loads in winter are more closely tied to the time-of-day during periods outside the morning peak. Also, peak winter loads on most days are typically (but not always) larger than peak summer loads. Recall, in Figure 8(a) the peak aggregated demand often occurred in summer.

4.2. *Residential analysis*

In Section 4.1, we analysed aggregate load and generation profiles for the clean dataset. In this section, we investigate daily residential load and PV generation variability for a small number of customers in the clean dataset. With an intention to highlight the daily variability in customer-specific load and PV generation profiles, we select three arbitrary dates for the purpose of comparison. More specifically, residential load and generation profiles for the first four customers in the clean dataset are presented on three consecutive Mondays from 1 July 2010, respectively. These four customers are located in the Central Coast region of NSW towards the northern coastline, and are in close proximity to one another. Due to this close geographical proximity, it is to be expected that the PV generation profiles of each respective customer would be similarly affected by cloud cover.

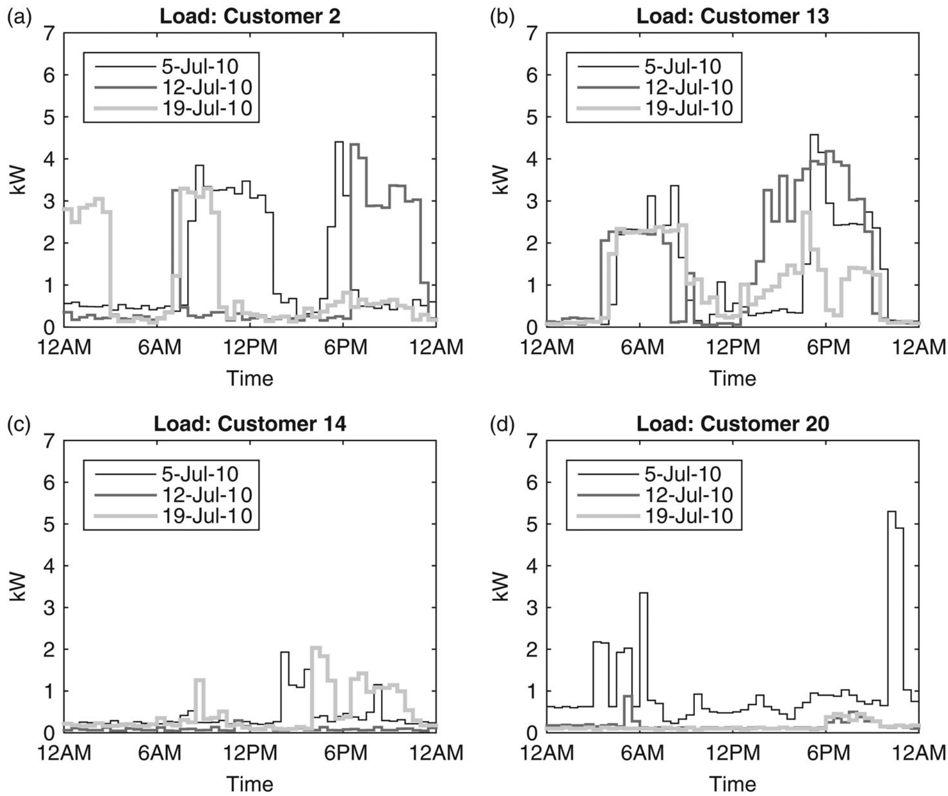

Figure 11. Load profiles for customers 2, 13, 14, and 20, on consecutive Mondays, respectively.

In Figure 11, we present load profiles on three consecutive Mondays for the first four customers in the clean dataset. We observe that the shape of the load profiles in Figure 11(a)–(c) is significantly different on each day, for each customer. In contrast, we observe that the shape of the load profile in Figure 11(d) on 12 July 2010 and 19 July 2010 is flatter and lower for Customer 20. As such, no visual confirmation regarding a strong relationship in energy consumption on consecutive weekdays is evident in Figure 11. Information regarding customer behaviour that affects energy use that leads to these variations in residential load profiles is provided in Slini, Giama, and Papadopoulos (2015) and Pothitou et al. (2014).

In Figure 12, we present generation profiles on three consecutive Mondays for the first four customers in the clean dataset. In Figure 12(a)–(d), we observe that daily peaks in PV generation vary across the customers, with peak generation on 19 July 2010 reaching 0.9 kW for Customer 2, 1.01 kW for Customer 13, 0.64 kW for Customer 14, and 0.89 kW for Customer 20. Likewise, the installed capacity of each rooftop PV unit varies, with the installed PV capacity in the Ausgrid dataset recorded as 1.62 kWp for Customer 2, 2.22 kWp for Customer 13, 1.48 kWp for Customer 14, and 1.57 kWp for Customer 20.

In Figure 12(b), we observe that the shape of the generation profiles is significantly different on each day. In contrast, we observe the shape of the generation profiles for Customer 2, Customer 14, and Customer 20 in Figure 12(a), 12(c), and 12(d) on 19 July 2010 are similar, yet vastly different to the generation profile of Customer 13 on 19 July 2010 (wherein PV production in the afternoon is much greater than PV production in the morning). We observe that fluctuations in PV production for some customers in close proximity to one another are potentially related. To assist the interested researcher looking to further describe these observations for each customer, information regarding solar irradiance, a key factor in PV production is provided in Elliston et al. (2015), Dehghan et al. (2014), and Šúri, Huld, and Dunlop (2005).

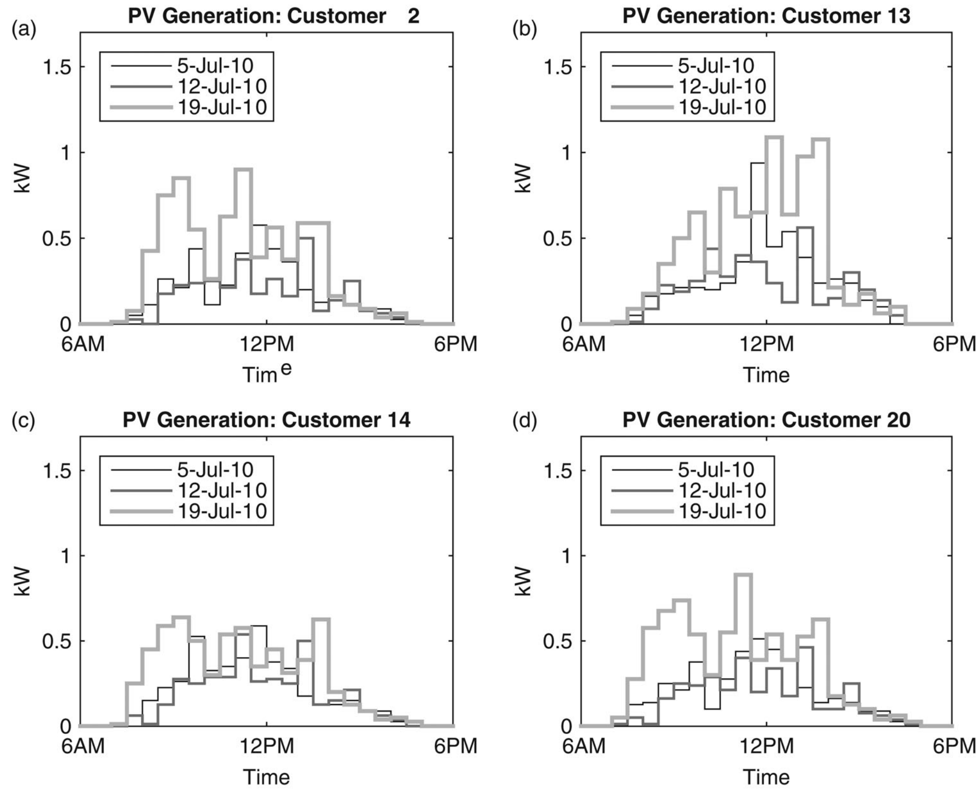

Figure 12. Generation profiles for customers 2, 13, 14, and 20, on three consecutive Mondays, respectively.

4.3. *Orientation of the PV panels*

To increase the amount of solar energy received on PV panels, residential customers generally orientate their PV panels due north in the southern hemisphere (Kacira et al. 2004; Šúri et al. 2007; Chatters 2015). However, PV incentives designed to reduce the evening peak demand (i.e. time-of-use net metering) potentially encourage customers to orientate their PV panels towards the west (Mondol, Yohanis, and Norton 2007). Another motivation to orientate PV panels to the west or east is the prevalence of shade from multistory dwellings (or large trees) that cover north-facing PV panels. Therefore, the optimum PV panel orientation potentially shifts from due north (in the southern hemisphere) to another orientation that is more specific to an individual residential dwelling.

The PV panel orientation of each customer in the Ausgrid dataset is unknown. Each customer in the clean dataset potentially orientates their rooftop PV panels close to due north. To assist with future research that verifies the PV panel orientation of each customer in the Ausgrid dataset, we examine generation profiles ($gg(j)$ for all $j = 1, \ldots, s$) for four of the customers in the clean dataset. We present PV generation profiles on 5 February 2011 (a day in summer), and on 9 July 2012 (a day in winter) for Customer 2 in Figure 13(a), for Customer 69 in Figure 13(b), for Customer 106 in Figure 13(c), and for Customer 214 in Figure 13(d).

In Figure 13, we observe the generation profiles of Customer 2, Customer 69, and Customer 214 peak around midday on 5 February 2011 and on 9 July 2012. These customers potentially orientated their respective PV panels due north (Ward, Moore, and Lindsay 2012). In contrast, the generation profile of Customer 106 peaked around 3 pm on 5 February 2011, consistent with a west-facing PV panel orientation (Ward, Moore, and Lindsay 2012). Further work to verify and infer the PV panel orientation of each customer in the clean dataset is certainly possible.

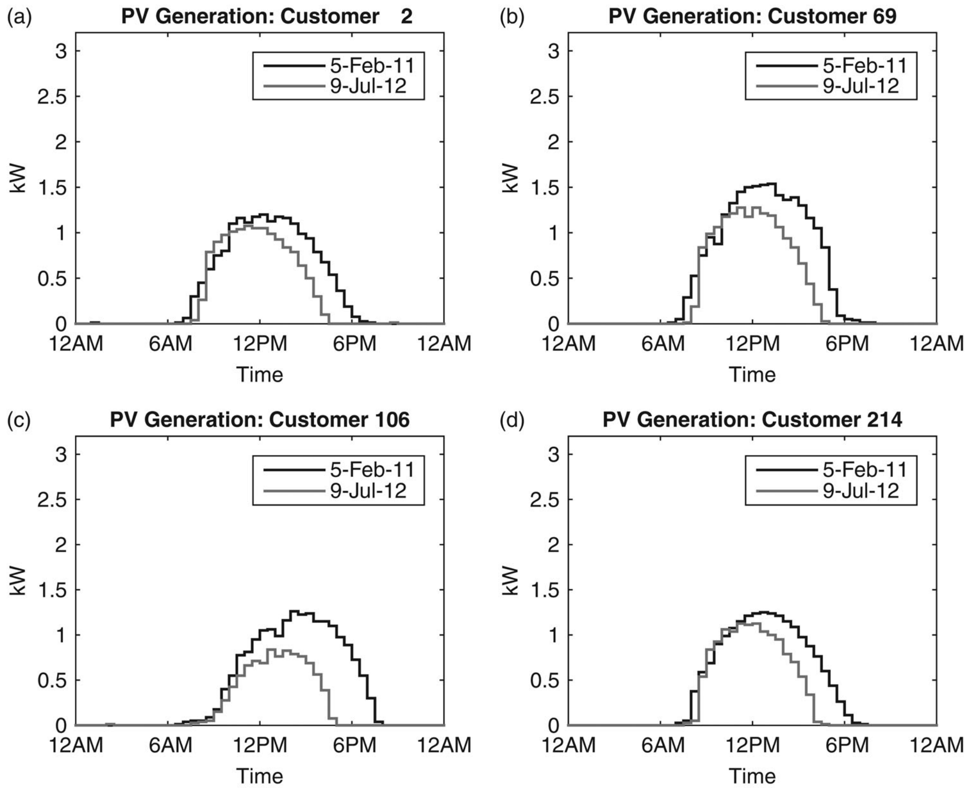

Figure 13. PV generation for (a) Customers 2, (b) Customers 69, (c) Customers 106, and (d) Customers 214, on 5 February 2011 and 9 July 2012, respectively.

5. Conclusions

In this paper, we have reported a publicly available dataset of measured load and PV generation from 300 residential customers located in an Australian distribution network. This dataset is a valuable resource to researchers and policy-makers alike since (1) residential load is measured separately to residential PV generation, (2) time-of-use meters record residential load and PV generation energy measurements after each half hour interval on each day, and (3) the dataset spans a 3-year period. To facilitate use of the dataset, we have presented an approach to remove customer-specific anomalous load and anomalous generation profiles (e.g. when a PV inverter fails), leaving a so-called clean dataset. Analysis of the clean dataset is presented on daily, seasonal, and annual timescales. We envision that the clean dataset will assist future research in the area of categorisation and forecasting of PV generation variability and intermittency.

Notes

1. Ausgrid information published in this paper is publicly available and no permission for publication was required.
2. For the time period 1 July 2012–30 June 2013, 187 customers make it into the 'clean dataset'.
3. In 2007 the AEDT commencement date was moved from the last Sunday of October to the first Sunday of October.
4. Anomalous data relating to clock recording were rectified by Ausgrid in March 2015. These anomalous data were identified via a category 3 elimination.

Acknowledgments

The assistance of Dr Robert Simpson in facilitating access to the Ausgrid dataset is gratefully acknowledged.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

ACIL Allen Consulting. 2013. *Drivers of Domestic PV Uptake Characteristics of Households with Solar Photovoltaic Systems*. The Australian Renewable Energy Agency. Tech. Rep. Accessed February 16, 2015. http://arena.gov.au/resources/.

von Appen, J., M. Braun, T. Stetz, K. Diwold, and D. Geibel. 2013. "Time in the Sun: the Challenge of High PV Penetration in the German Electric Grid." *IEEE Power and Energy Magazine* 11 (2): 55–64.

Ausgrid. 2014a. *Ausgrid Annual Report 2013/14*. Accessed May 1, 2015. http://www.ausgrid.com.au.

Ausgrid. 2014b. *Data to Share*. Accessed February 16, 2015. http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/Data-to-share.aspx.

Ausgrid. 2014c. *Network Price List 2013–2014*. Accessed February 3, 2015. http://www.ausgrid.com.au/Com-mon/Industry/Regulation/Network-prices/Price-lists-and-policy/Historical-network-prices.aspx.

Ausgrid. 2014d. *Privacy Policy*. Accessed February 16, 2015. http://www.ausgrid.com.au/Common/About-us/Privacy-Policy.aspx.

Ausgrid. 2014e. *Solar Home Electricity Data*. Accessed January 28, 2015. http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/Data-to-share/Solar-household-data.aspx.

ABS (Australian Bureau of Statistics). 2006. "Census of Population and Housing: Census Geographic Areas Digital Boundaries, Australia, 2006." Data cube: Postal Areas (POA) 2006 Digital Boundaries in MapInfo Interchange Format, cat. no. 2923.0.30.001. Accessed February 1, 2015. http://www.abs.gov.au/ausstats/abs@.nsf/detailspage/2923.0.30.0012006?opendocument.

Australian Government Bureau of Meteorology. 2011. *Sydney in February 2011: 2nd Warmest on Record*. Accessed February 11, 2015. http://www.bom.gov.au/clim-ate/current/month/nsw/archive/201102.sydney.shtml.

Bazilian, M., I. Onyeji, M. Liebreich, I. MacGill, J. Chase, J. Shah, D. Gielen, D. Arent, D. Landfear, and S. Zhengrong. 2013. "Re-Considering the Economics of Photovoltaic Power." *Renewable Energy* 53, 329–338.

Black, A. J. 2004. "Financial Payback on California Residential Solar Electric Systems." *Solar Energy* 77 (4): 381–388.

Braun, P., L. Grüne, C. M. Kellett, S. R. Weller, and K. Worthmann. Forthcoming. "A Distributed Optimization Algorithm with Centralized Performance Properties." *Proceedings of the 54th IEEE Conference on Decision and Control (CDC 2015)*, Osaka, Japan.

Campoccia, A., L. Dusonchet, E. Telaretti, and G. Zizzo. 2009. "Comparative Analysis of Different Supporting Measures for the Production of Electrical Energy by Solar PV and Wind Systems: Four Representative European Cases." *Solar Energy* 83 (3): 287–297.

Chatters, M. "DGDS215 – Optimising Photovoltaic Panel Orientation for Customer Benefits – A Grid PS+EDGE Modelling Platform Case Study." Smart Grid Smart City. Tech. Rep. Accessed February 16, 2015. https://ich.smartgridsmartcity.com.au/Home/SGSC-Report/Technical-Compendia/Distributed-Generation-and-Distributed-Storage.

Chow, C. W., S. Belongie, and J. Kleissl. 2015. "Cloud Motion and Stability Estimation for Intra-Hour Solar Forecasting." *Solar Energy* 115, 645–655.

Darghouth, N. R., G. Barbose, and R. H. Wiser. 2014. "Customer-Economics of Residential Photovoltaic Systems (Part 1): The Impact of High Renewable Energy Penetrations on Electricity Bill Savings with Net Metering." *Energy Policy* 67, 290–300.

Dehghan, A., A. A. Prasad, S. C. Sherwood, and M. Kay. 2014. "Evaluation and Improvement of TAPM in Estimating Solar Irradiance in Eastern Australia." *Solar Energy* 107, 668–680.

Elliston, B., I. MacGill, A. Prasad, and M. Kay. 2015. "Spatio-Temporal Characterisation of Extended Low Direct Normal Irradiance Events Over Australia Using Satellite Derived Solar Radiation Data." *Renewable Energy* 74, 633–639.

Hart, E. K., E. D. Stoutenburg, and M. Z. Jacobson. 2012. "The Potential of Intermittent Renewables to Meet Electric Power Demand: Current Methods and Emerging Analytical Techniques." *Proceedings of the IEEE* 100 (2): 322–334.

Holttinen, H., A. Tuohy, M. Milligan, E. Lannoye, V. Silva, S. Muller, and L. Soder. 2013. "The Flexibility Workout: Managing Variable Resources and Assessing the Need for Power System Modification." *IEEE Power and Energy Magazine* 11 (6): 53–62.

Independent Pricing and Regulatory Tribunal. 2012. *Final Report– Solar feed-in tariffs– March 2012*. Accessed February 16, 2015. http://www.ipart.nsw.gov.au/.

Kacira, M., M. Simsek, Y. Babur, and S. Demirkol. 2004. "Determining Optimum Tilt Angles and Orientations of Photovoltaic Panels in Sanliurfa, Turkey." *Renewable Energy* 29 (8): 1265–1275.

Kroposki, B., R. Margolis, and K. Lynn. 2011. "Power to the People." *IEEE Power and Energy Magazine* 9 (3): 16–22.

Kroposki, B., and B. Mather. 2015. "Rise of Distributed Power: Integrating Solar Energy into the Grid." *IEEE Power and Energy Magazine* 13 (2): 14–18.

Keerthisinghe, C., G. Verbic, and A. C. Chapman. 2014a. "Addressing the Stochastic Nature of Energy Management in Smart Homes." *Power Systems Computation Conference 2014 (PSCC)*, Wroclaw, Poland, 1–7.

Keerthisinghe, C., G. Verbic, and A. C. Chapman. 2014b. "Evaluation of a Multi-Stage Stochastic Optimisation Framework for Energy Management of Residential PV-Storage Systems." *Australasian Universities 2014 Power Engineering Conference (AUPEC)*, Perth, Australia, 1–6.

Khalilpour, R., and A. Vassallo. 2015. "Leaving the Grid: An Ambition or a Real Choice?" *Energy Policy* 82, 207–221.

Lang, T., E. Gloerfeld, and B. Girod. 2015. "Don't Just Follow the Sun – A Global Assessment of Economic Performance for Residential Building Photovoltaics." *Renewable & Sustainable Energy Reviews* 42, 932–951.

Mondol, J. D., Y. G. Yohanis, and B. Norton. 2007. "The Impact of Array Inclination and Orientation on the Performance of a Grid-Connected Photovoltaic System." *Renewable Energy* 32 (1): 118–140.

Moosavian, S. M., N. A. Rahim, J. Selvaraj, and K. H. Solangi. 2013. "Energy Policy to Promote Photovoltaic Generation." *Renewable & Sustainable Energy Reviews* 25, 44–58.

NSW Government. 2010. *Solar Bonus Scheme*. Accessed February 16, 2015. http://www.resourcesandenergy.nsw.gov.au/energy-consumers/solar/solar-bonus-scheme.

Nykamp, S., M. G. C. Bosman, A. Molderink, J. L. Hurink, and G. J. M. Smit. 2013. "Value of Storage in Distribution Grids–Competition or Cooperation of Stakeholders?" *IEEE Transactions on Smart Grid* 4 (3): 1361–1370.

Ogimoto, K., I. Kaizuka, Y. Ueda, and T. Oozeki. 2013. "A Good Fit: Japan's Solar Power Program and Prospects for the New Power System." *IEEE Power and Energy Magazine* 11 (2): 65–74.

Pereira, R., J. Figueiredo, R. Melicio, V. M. F. Mendes, J. Martins, and J. C. Quadrado. 2015. "Consumer Energy Management System with Integration of Smart Meters." *Energy Reports* 1, 22–29.

Pothitou, M., A. J. Kolios, L. Varga, and S. Gu. 2014. "A Framework for Targeting Household Energy Savings through Habitual Behavioural Change." *International Journal of Sustainable Energy* 1–15. July 16, 2014.

Quilumba, F. L., W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados. 2015. "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities." *IEEE Transactions on Smart Grid* 6 (2): 911–918.

Ratnam, E. L., S. R. Weller, and C. M. Kellett. 2013. "An Optimization-Based Approach for Assessing the Benefits of Residential Battery Storage in Conjunction with Solar PV." *Proceedings of the IX IEEE International Symposium on Bulk Power System Dynamics and Control (IREP'13)*, Rethymnon, Greece, August 25–30, 1–8.

Ratnam, E. L., S. R. Weller, and C. M. Kellett. 2015a. "An Optimization-Based Approach to Scheduling Residential Battery Storage with Solar PV: Assessing Customer Benefit." *Renewable Energy* 75, 123–134.

Ratnam, E. L., S. R. Weller, and C. M. Kellett. 2015b. "Scheduling Residential Battery Storage with Solar PV: Assessing the Benefits of Net Metering." *Applied Energy* 155, 881–891.

Rhodes, J. D., C. R. Upshaw, C. B. Harris, C. M. Meehan, D. A. Walling, P. A. Navrátil, A. L. Beck, et al. 2014. "Experimental and Data Collection Methods for a Large-Scale Smart Grid Deployment: Methods and First Results." *Energy* 65, 462–471.

Slini, T., E. Giama, and A. M. Papadopoulos. 2015. "The Impact of Economic Recession on Domestic Energy Consumption." *International Journal of Sustainable Energy* 34 (3–4): 259–270.

Šúri, M., T. A. Huld, and E. D. Dunlop. 2005. "PV-GIS: A Web-Based Solar Radiation Database for the Calculation of PV Potential in Europe." *International Journal of Sustainable Energy* 24 (2): 55–67.

Šúri, M., T. A. Huld, E. D. Dunlop, and H. A. Ossenbrink. 2007. "Potential of Solar Electricity Generation in the European Union Member States and Candidate Countries." *Solar Energy* 81 (10): 1295–1305.

Ward, J. K., T. Moore, and S. Lindsay. 2012. "The Virtual Power Station – Achieving Dispatchable Generation from Small Scale Solar." *Proceedings of the 50th Annual Australian Solar Council Conference on Solar 2012 (AuSES)*, Melbourne, Australia, December 6–7, 1–10.

Worthmann, K., C. M. Kellett, P. Braun, L. Grüne, and S. R. Weller. 2015. "Distributed and Decentralized Control of Residential Energy Systems Incorporating Battery Storage." *IEEE Transactions on Smart Grid* 6 (4): 1914–1923.

Yang, Y., H. Li, A. Aichhorn, J. Zheng, and M. Greenleaf. 2014. "Sizing Strategy of Distributed Battery Storage System with High Penetration of Photovoltaic for Voltage Regulation and Peak Load Shaving." *IEEE Transactions on Smart Grid* 5 (2): 982–991.

Appendix. Selected customer ID data

The Customer ID's and corresponding dates presented in Figures 3, 5, 11–13 of this paper are included in Table A1.

Table A1. Index of each Customer ID and corresponding dates presented in Figures 3, 5, 11–13.

Section	Figure	Customer ID	Date		
3.1	3	9	19 January 13	24 February 13	15 March 13
		191	3 July 10	18 July 10	21 August 10
		221	29 October 12	17 November 12	17 May 13
		229	23 February 12	11 March 12	29 March 12
3.2	5	1	28 January 13	23 February 13	23 May 13
		145	9 September 10	11 September 10	28 May 11
		215	14 July 10	15 July 10	16 July 10
		248	21 February 11	3 April 11	21 April 11
4.2	11	2	5 July 10	12 July 10	19 July 10
		13	5 July 10	12 July 10	19 July 10
		14	5 July 10	12 July 10	19 July 10
		20	5 July 10	12 July 10	19 July 10
4.2	12	2	5 July 10	12 July 10	19 July 10
		13	5 July 10	12 July 10	19 July 10
		14	5 July 10	12 July 10	19 July 10
		20	5 July 10	12 July 10	19 July 10
4.3	13	2	5 February 11	9 July 12	
		69	5 February 11	9 July 12	
		106	5 February 11	9 July 12	
		214	5 February 11	9 July 12	