



# Solent Acheiving Value from Efficiency (SAVE): 15-minute electricity consumption data archive report

Tom Rushby (t.w.rushby@soton.ac.uk\*<sup>\*</sup>@tom\_rushby)

Ben Anderson (b.anderson@soton.ac.uk<sup>†</sup>@dataknut)

Patrick James (p.a.james@soton.ac.uk<sup>‡</sup>@pab\_james)

AbuBakr Bahaj (a.s.bahaj@soton.ac.uk<sup>§</sup>)

Last run at: 2019-10-09 10:26:15

## Contents

<b>1 About</b>	<b>3</b>
1.1 Citation . . . . .	3
1.2 Circulation . . . . .	3
<b>2 Introduction</b>	<b>3</b>
<b>3 Data provided</b>	<b>4</b>
3.1 Data structure . . . . .	4
3.2 Example data . . . . .	4
3.3 Data summary . . . . .	5
3.4 Data files . . . . .	7
<b>4 Support</b>	<b>8</b>
<b>5 Known issues</b>	<b>8</b>
5.1 Out-of-sample IDs . . . . .	8
5.2 Communication outages and interpolation . . . . .	8
<b>6 References</b>	<b>10</b>

## Document history

**Table 1:** Document history

Version	Date	Author	Notes
Version 0_1	25th July 2019	T. Rushby	Initial preliminary run.
Version 0_2	9th October 2019	T. Rushby	Minor updates to text, data files section added.

## 1 About

This work was carried out by the Sustainable Energy Research Group<sup>1</sup> in the University of Southampton's Faculty of Engineering and Environment<sup>2</sup>.

The work was funded by the Low Carbon Network Fund (LCNF) through the Solent Achieving Value from Efficiency<sup>3</sup> project and is (c) 2019 the University of Southampton.

This report forms part of the documentation supporting the deposit with UK Data Archive of data collected under the SAVE project. Further details of the project, including documentation of the interventions tested, can be found on the project website<sup>4</sup>.

### 1.1 Citation

If you wish to refer to any of the material from this report please cite as:

- Rushby, T., Anderson, B., James, P.A.B. and Bahaj, A.S. (2019) Solent Achieving Value from Efficiency (SAVE): 15-minute electricity consumption data archive report, University of Southampton: Southampton, UK.

### 1.2 Circulation

This report is public, to accompany the data release associated with the Low Carbon Network funded (LCNF) Solent Achieving Value from Efficiency (SAVE) project.

## 2 Introduction

This document describes the electricity consumption data collected through the SAVE project (provided through the Navetas Loop system). It also describes the data deposited to the UK data archive - available for research purposes.

The Solent Achieving Value from Efficiency (SAVE) was funded by the Low Carbon Network (LCN) Fund and sought to establish to what extent energy efficiency measures and behaviour change programmes can be considered a cost effective tool for managing peak demand, and whether they provide an alternative to traditional low-voltage network reinforcement.

The SAVE sample recruited a stratified random address-based sample of households from:

- the County of Hampshire
- the City of Southampton
- the City of Portsmouth
- the Isle of Wight

---

<sup>1</sup><http://www.energy.soton.ac.uk>

<sup>2</sup><http://www.southampton.ac.uk/engineering/index.page>

<sup>3</sup><http://www.energy.soton.ac.uk/save-solent-achieving-value-from-efficiency/>

<sup>4</sup><http://www.save-project.co.uk/>

With some cautions this means that the sample is representative of households across these areas – but not of the UK in general.

The fieldwork initially recruited over 4,000 households in late 2016 but due to attrition this slowly declined. Efforts were taken to refresh and re-recruit new households to replace those that left.

Using a randomised control trial design, the SAVE project tested a number of interventions, targeted at engaging householders in reducing demand for electricity during the network peak hours between 4 and 8pm.

## 3 Data provided

### 3.1 Data structure

Consumption data is contained in a number of compressed (gzipped) comma separated value (.csv) files. Filenames use the following naming convention *NAX-YYYY-MM-DD.csv.gz*. The date in the filename refers to the end date of the one-week period of data covered by each file.

The variables contained in the files are as follows (column headers):

- `navetasID`
- `received_timestamp`
- `recorded_timestamp`
- `energy`

### 3.2 Example data

Table 2 contains example rows of Loop data. Each variable is described below.

**Table 2:** Example rows of Navetas Loop 15-minute electricity consumption data

<code>bmg_id</code>	<code>received_timestamp</code>	<code>recorded_timestamp</code>	<code>energy</code>
A2956616455	1483315201	1483315200	1669800
A1956626264	1483315202	1483303500	80141
A4956660123	1483315203	1482604200	251700
A1956639051	1483315205	1483315200	925440
A3956639150	1483315205	1483315200	2090252
a4956636535	1483315205	1483315200	2657218

#### 3.2.1 `navetasID` variable coding

The 11-character alphanumeric `navetasID` variable contains information related to the Loop installations.

The first character of the variable provides the installation type, with character given a value of A or B, indicating an ADSL or GSM data connection respectively, for example: A\*\*\*\*\* (ADSL).

The second character of the variable provides the trial group, with numeric value between 1 and 4, for example: A3\*\*\*\*\* (ADSL, trial group 3). Values outside of this range identify data that should not be included within the SAVE sample (i.e. test units).

At the end of September 2018, Navetas updated the coding of the `navetasID` to enable the identification of a number of groups associated with the trial implementation.

The change affected households in trial groups 3 and 4 only (TG3 and TG4) and the 1st character of `navetasID`. For TGs 3 and 4 the character changed to identify both the connection type (ADSL vs GSM), *and* the 'threshold group' to which each household was allocated during the third trial period. The values are shown below:

No sub-groups: ADSL = A, GSM = B (e.g. A2\*\*\*\*\* – TG2/ADSL)

Threshold group - Low: ADSL = C, GSM = D (e.g. C3\*\*\*\*\* – TG3/ADSL/Low threshold)

Threshold group - Med: ADSL = E, GSM = F (e.g. F4\*\*\*\*\* – TG4/GSM/Med threshold)

Threshold group - High: ADSL = G, GSM = H (e.g. G3\*\*\*\*\* – TG3/ADSL/High threshold)

No threshold (opted-out): ADSL = I, GSM = J (e.g. J4\*\*\*\*\* – TG4/GSM/opted-out)

No participating households in trial groups 3 and 4 should be left as existing as they should either be allocated to a threshold group, or have opted out. We therefore expect no households in these groups to retain A or B coding from the date of the change.

The 15-minute electricity consumption data can be linked to the household survey data by matching the unique household identifier. This is done using the **final** 9 characters of the `navetasID` variable and the `BMG_ID` variable from the household survey data.

### 3.2.2 `received_timestamp` variable

This variable contains the date and time that each observation was received by Navetas. The variable is recorded as a Unix timestamp (milliseconds, UTC).

### 3.2.3 `recorded_timestamp` variable

This variable contains the date and time that each observation was observed and relates to the end of the 15-minute interval, formatted as previous. Due to communication issues there may be a large lag (up to several months) between the recorded and received values.

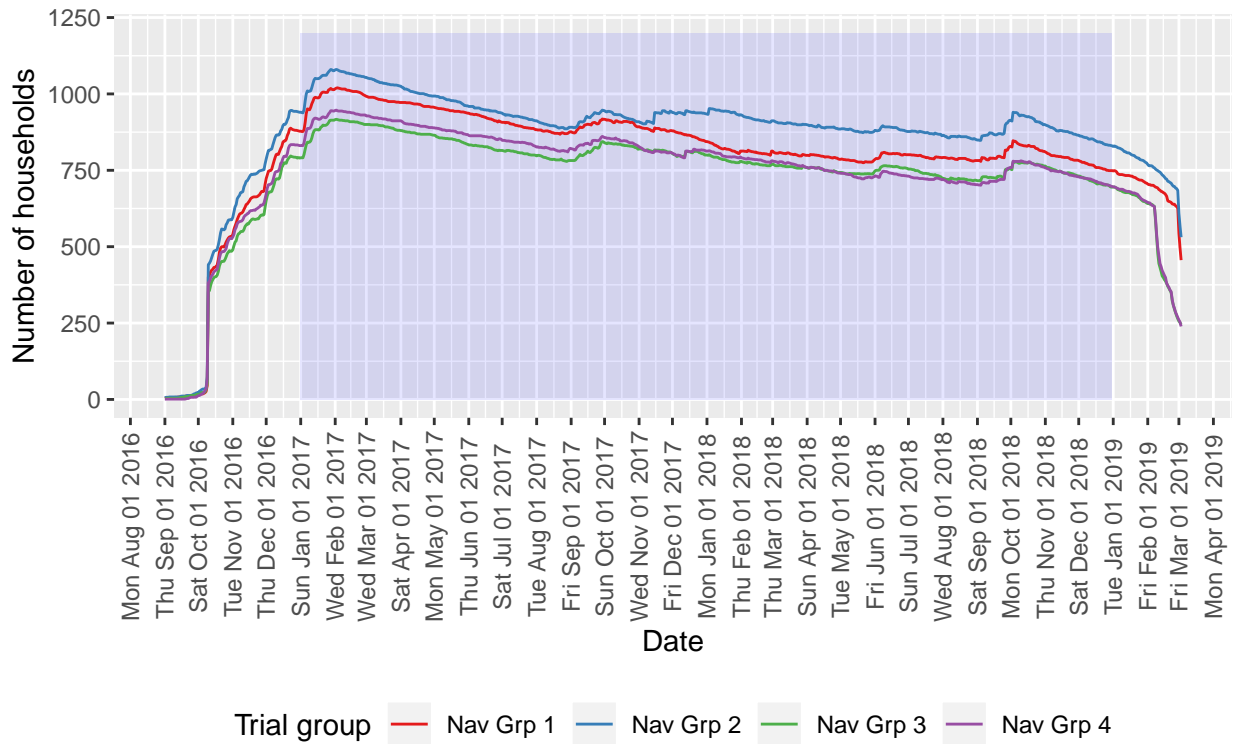
### 3.2.4 `energy` variable

For each recorded timestamp the energy consumed variable provides a value in Watt-hours (Wh) for the cumulative electricity consumption at that time.

## 3.3 Data summary

This report provides a summary of the 15-minute electricity consumption data for the period shown above. The summaries are generated using code developed by the project team to pre-process the Navetas data before use.

Figure 1 shows the number of Loop systems in each of the four trial groups contributing data on any given day during the data period. The shaded box indicates the period covered by the data provided to the UK Data Archive (1st January 2017 to 31st December 2018).



SAVE sample households  
 15-minute Wh data: 2016-09-01 to 2019-02-28  
 Shaded area indicates data supplied in UKDA data deposit

**Figure 1:** SAVE 15-minute consumption: households with data

Monthly summary statistics for the data are provided in Table 3 below.

**Table 3:** 15-minute consumption data: summary statistics by month

Year	Month	Mean kWh	Std Dev	Min kWh	Max kWh	Observations	Households
2016	Sep	0.081	0.099	0	1.090	84457	66
2016	Oct	0.105	0.161	0	4.196	4031638	2169
2016	Nov	0.121	0.186	0	4.238	7201562	2837
2016	Dec	0.128	0.193	0	5.377	9575254	3532
2017	Jan	0.132	0.205	0	5.958	11212847	4018
2017	Feb	0.120	0.188	0	5.809	10544499	3970
2017	Mar	0.108	0.167	0	4.388	11394443	3875
2017	Apr	0.098	0.151	0	3.994	10766955	3789
2017	May	0.092	0.142	0	4.416	10869924	3704
2017	Jun	0.086	0.141	0	134.835	10238032	3600
2017	Jul	0.086	0.131	0	65.241	10318739	3520
2017	Aug	0.086	0.169	0	209.989	10045388	3471
2017	Sep	0.091	0.151	0	141.020	9944643	3684
2017	Oct	0.097	0.184	0	276.029	10454422	3596
2017	Nov	0.112	0.281	0	402.022	9889325	3541
2017	Dec	0.124	0.263	0	309.620	10153420	3509
2018	Jan	0.120	0.523	0	1004.283	9969649	3429
2018	Feb	0.122	0.231	0	305.194	8811204	3368
2018	Mar	0.115	0.368	0	860.664	9641020	3316
2018	Apr	0.098	0.170	0	152.194	9186934	3246
2018	May	0.087	0.144	0	76.797	9305954	3207
2018	Jun	0.083	0.154	0	240.666	9159341	3250
2018	Jul	0.081	0.127	0	106.909	9316905	3197
2018	Aug	0.082	0.147	0	159.210	9149879	3143
2018	Sep	0.085	0.168	0	251.961	8958944	3349
2018	Oct	0.092	0.173	0	171.608	9792189	3375
2018	Nov	0.106	0.192	0	208.414	9085986	3245
2018	Dec	0.119	0.193	0	55.304	9026539	3116
2019	Jan	0.124	0.203	0	5.992	8564484	2974
2019	Feb	0.120	0.194	0	5.999	6143090	2761
2019	Mar	0.120	0.179	0	4.272	438877	1668

Note the very large values for the maximum recorded 15-minute consumption (Max kWh) from June 2017 onward. This was due to an issue with non-communicating systems coming back online providing large cumulative consumption observations. These values should be addressed prior to the use of the data (see 'Known issues' below).

### 3.4 Data files

The household electricity consumption dataset made available consists of a set of files provided in compressed (gzipped) comma-separated value (.csv) format. Each file contains the consumption data for all SAVE participants for a single week (see 3.2 above).

For convenience, the files have been bundled together into four compressed (zipped) folders, each containing six months of data with the coverage indicated in the filenames, for example SAVE\_Consumption\_Data\_2017\_1 contains data for January through June 2017, SAVE\_Consumption\_Data\_2017\_2 contains data for July through December 2017.

The files are listed below:

```
## [1] "SAVE_Consumption_Data_2017_1_v0-1.zip"  
## [2] "SAVE_Consumption_Data_2017_2_v0-1.zip"  
## [3] "SAVE_Consumption_Data_2018_1_v0-1.zip"  
## [4] "SAVE_Consumption_Data_2018_2_v0-1.zip"
```

## 4 Support

A repository has been set up through the University of Southampton GitLab service to support the SAVE project data deposit at <https://git.soton.ac.uk/SERG/save-public>. Newly identified issues with the data will be logged by the project team and updated through the repository. New issues can be raised at <https://git.soton.ac.uk/SERG/save-public/issues> (registration required) and we will endeavour to respond although note that SAVE project funding has now ended and team members have moved on to other projects.

## 5 Known issues

### 5.1 Out-of-sample IDs

Some Loop monitoring systems were provided to the project team for testing. These do not form part of the SAVE sample but appear within the dataset. The systems can be identified using the `navetasID` variable and should be removed prior to any analysis of the data.

### 5.2 Communication outages and interpolation

The Loop system provided some protection against data loss through communication issues. Clamps stored consumption measurements during periods of non-communication with the hub and/or Navetas system (e.g. where a hub was disconnected from the ADSL router). The clamps had adequate storage to allow up to a month of consumption data to be retained, after which time the device would begin overwriting data. If the memory was full, the clamp preserved data at 30-minute resolution with missing observations being imputed using linear interpolation.

In some cases, hubs were offline for a long period of time, resulting in a single large cumulative consumption value being recorded when the system resumed communication. In some cases this large consumption value was used to impute missing observations by backwards interpolation:

- Filling in the missing observations (and giving them the same received `dateTime` as the single catch-up);
- Evenly distributing the large value over these missing readings.

To flag them: they will be sequences of observations (incrementing `obsDateTime`) with:

- Identical received timestamps, and
- Identical change in consumption across a series of adjacent observations.

Occasionally, the interpolation of large values was not performed by the Navetas system, resulting in erroneously large consumption values for a 15-minute period. They can be identified by the large lag-times between consecutive observation dates and removed as required.



## Runtime

Report generated using knitr<sup>5</sup> in RStudio<sup>6</sup> with R version 3.6.1 (2019-07-05) running on x86\_64-apple-darwin15.6.0.

Analysis completed in 31.958 seconds ( 0.53 minutes).

---

<sup>5</sup><https://cran.r-project.org/package=knitr>

<sup>6</sup><http://www.rstudio.com>

## 6 References