NZ GREEN Grid Household Electricity Demand Study

Research data overview (version 1.0)


Last run at: 2018-08-30 14:45:40 (Pacific/Auckland)

1 About

1.1 Report circulation:
- Public – this report is intended to accompany the data release.

1.2 License

This work is made available under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) License (https://creativecommons.org/licenses/by-sa/4.0/).

This means you are free to:
- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

Under the following terms:
- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
- No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Notices:
- You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.
- No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material. #YMVM

For the avoidance of doubt and explanation of terms please refer to the full license notice (https://creativecommons.org/licenses/by-sa/4.0/) and legal code (https://creativecommons.org/licenses/by-sa/4.0/legalcode).

1.3 Citation

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


This work is (c) 2018 the authors.

1.4 History

You may not be reading the most recent version of this report. Please check:
- the overall package documentation (https://cfsotago.github.io/GREENGridData/);
- this report’s edit history (https://github.com/CfSOtago/GREENGridData/commits/master/makeDocs/buildOverviewReport.Rmd)

1.5 Support

This work was supported by:
- The University of Otago (https://www.otago.ac.nz/);
- The University of Southampton (https://www.southampton.ac.uk/);
- The New Zealand Ministry of Business, Innovation and Employment (MBIE) (http://www.mbie.govt.nz) through the NZ GREEN Grid (https://www.otago.ac.nz/centre-sustainability/research/energy/otago050285.html) grant (Contract ID: UOCX1203);
2 Introduction

The NZ GREEN Grid household electricity demand study (https://cfsotago.github.io/GREENGridData/) recruited a sample of c 25 households in each of two regions of New Zealand (Stephenson et al. 2017). The first sample was recruited in early 2014 and the second in early 2015. Research data includes:

- 1 minute electricity power (W) data was collected for each dwelling circuit using GridSpy (https://gridspy.com/) monitors on each power circuit (and the incoming power). The power values represent mean(W) over the minute preceeding the observation timestamp;
- Dwelling & appliance surveys;
- Occupant time-use diaries (focused on energy use).

NB: Version 1 of the data package does not include the time-use diaries.

This report provides an overview of the GREEN Grid project (Stephenson et al. 2017) research data.

3 Data Package

Version 1.0 of the data package contains:

- powerData.zip: 1 minute power demand data for each circuit in each household. One file per household;
- ggHouseholdAttributesSafe.csv.zip: anonymised household attribute data;
- checkPlots.zip:
  - simple line charts of mean power per month per year for each circuit monitored for each household. These are a useful check;
  - tile plots (heat maps/carpet plots) of the number of observations per hour per day. Also a useful check…

4 Study recruitment

The project research sample comprises 44 households who were recruited via the local power lines companies in two areas: New Plymouth starting in May 2014 and Hawkes Bay starting in November 2014.

Recruitment was via a non-random sampling method and a number of households were intentionally selected for their ‘complex’ electricity consumption (and embedded generation) patterns and appliances (Giraldo Ocampo 2015, Stephenson et al. (2017), Jack et al. (2018), Suomalainen et al. (2017)).

The lines companies invited their own employees and those of other local companies to participate in the research and ~80 interested potential participants completed short or long forms of the Energy Cultures 2 household survey (Wooliscroft 2015). Households were then selected from this pool by the project team based on selection criteria relevant to the GREEN Grid project. These included:

- having the majority of their energy supply from electricity (i.e. not gas heating);
- household size;
- types of appliances owned.

After informed consent was obtained from each household, an electrician contracted by the two lines companies completed an appliance survey to record detailed information about the appliances in each house. This survey contained information about the number of appliances owned, brand, model number, efficiency and age. The electrician also installed the GridSpy units which recorded electricity power demand at a circuit level. The GridSpy units automatically upload the monitoring data to the GridSpy company’s secure database from where it was downloaded by the GREEN Grid research team.

As a result of this process the sample cannot be assumed to represent the population of customers (or employees) of any of the companies involved, nor the populations in each location (Stephenson et al. 2017).

Table 4.1 shows the number in each sample.

Table 4.1: Sample location

<table>
<thead>
<tr>
<th>Location</th>
<th>nHouseholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkes Bay</td>
<td>20</td>
</tr>
<tr>
<td>New Plymouth</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 4.2 shows the number for whom valid appliance and survey data is available in this data package. Note that even those which appear to lack appliance data may have sufficient survey data to deduce appliance ownership (see question numbers Q19_* and Q40_*).

Table 4.2: Sample information

<table>
<thead>
<tr>
<th>Location</th>
<th>hasShortSurvey</th>
<th>hasLongSurvey</th>
<th>hasApplianceSummary</th>
<th>nHouseholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawkes Bay</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Hawkes Bay</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>1</td>
</tr>
<tr>
<td>Hawkes Bay</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Hawkes Bay</td>
<td>Yes</td>
<td>NA</td>
<td>Yes</td>
<td>13</td>
</tr>
<tr>
<td>New Plymouth</td>
<td>NA</td>
<td>Yes</td>
<td>NA</td>
<td>12</td>
</tr>
<tr>
<td>New Plymouth</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>12</td>
</tr>
</tbody>
</table>
5 Data collection duration

Figure 5.1 shows the total number of households for whom GridSpy data exists on a given date by sample. The plot includes any data, including partial data and suggests that for analytic purposes the period from April 2015 to March 2016 (indicated) would offer the maximum number of households.

![Graph showing number of households per day for all clean GridSpy data](image)

Source: GREEN Grid Project data 2014-06-26 to 2018-08-01
Shaded area indicates 12 month period with largest number of households

Figure 5.1: Number of households sending GridSpy data by date

6 Key attributes

Table 6.1 shows key attributes for the recruited sample. Note that two GridSpy monitors were re-used and so require new hhIDs to be set from the date of re-use using the `linkID` variable. This is explained in more detail in the GridSpy processing (gridSpy1mProcessingReport_v1.0.html#reallocation) report. Linkage between the survey and GridSpy data should therefore always use `linkID` to avoid errors.

Table 6.1: Sample details

<table>
<thead>
<tr>
<th>hhID</th>
<th>linkID</th>
<th>Location</th>
<th>surveyStartDate</th>
<th>nAdults</th>
<th>nChildren0_12</th>
<th>nTeenagers13_18</th>
<th>notes</th>
<th>r_stopDate</th>
<th>hasApplianceSummary</th>
</tr>
</thead>
<tbody>
<tr>
<td>rf_06</td>
<td>rf_06</td>
<td>New Plymouth</td>
<td>2014-05-19 09:49:00</td>
<td>2</td>
<td>0</td>
<td>0 NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_07</td>
<td>rf_07</td>
<td>New Plymouth</td>
<td>2014-06-23 21:25:00</td>
<td>2</td>
<td>2</td>
<td>0 NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_08</td>
<td>rf_08</td>
<td>New Plymouth</td>
<td>2014-05-14 12:21:00</td>
<td>2</td>
<td>0</td>
<td>0 NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_09</td>
<td>rf_09</td>
<td>New Plymouth</td>
<td>2014-06-19 11:33:00</td>
<td>2</td>
<td>1</td>
<td>0 NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_10</td>
<td>rf_10</td>
<td>New Plymouth</td>
<td>2014-05-20 17:01:00</td>
<td>2</td>
<td>1</td>
<td>0 NA</td>
<td>NA</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>rf_11</td>
<td>rf_11</td>
<td>New Plymouth</td>
<td>2014-06-06 12:16:00</td>
<td>2</td>
<td>NA</td>
<td>NA NA</td>
<td>NA</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>rf_12</td>
<td>rf_12</td>
<td>New Plymouth</td>
<td>2014-06-16 07:34:00</td>
<td>1</td>
<td>0</td>
<td>0 NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rf_13</td>
<td>rf_13</td>
<td>New Plymouth</td>
<td>2014-05-14 12:07:00</td>
<td>2</td>
<td>1</td>
<td>1 NA</td>
<td>NA</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>rf_14</td>
<td>rf_14</td>
<td>New Plymouth</td>
<td>2014-06-10 11:51:00</td>
<td>1</td>
<td>1</td>
<td>0 NA</td>
<td>NA</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>rf_15</td>
<td>rf_15a</td>
<td>New Plymouth</td>
<td>2014-06-17 15:38:00</td>
<td>1</td>
<td>0</td>
<td>0 Connected</td>
<td>15/01/2015, Re-used</td>
<td>2015-01-15</td>
<td>NA</td>
</tr>
<tr>
<td>hhID</td>
<td>linkID</td>
<td>Location</td>
<td>surveyStartDate</td>
<td>nAdults</td>
<td>nChildren0_12</td>
<td>nTeenagers13_18</td>
<td>notes</td>
<td>r_stopDate</td>
<td>hasApplianceSummary</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------</td>
<td>---------------</td>
<td>----------------</td>
<td>------------------------------------------------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>rf_15</td>
<td>rf_15b</td>
<td>New Plymouth</td>
<td>2014-05-16 17:36:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>Re-user. Then disconnected 02/04/2016</td>
<td>2016-04-02</td>
<td>NA</td>
</tr>
<tr>
<td>rf_16</td>
<td>rf_16</td>
<td>New Plymouth</td>
<td>2014-06-10 15:29:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_17</td>
<td>rf_17a</td>
<td>New Plymouth</td>
<td>2014-05-14 20:04:00</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>Unusual &amp; specialist energy tech configuration. Disconnected 28/03/2016. Re-used.</td>
<td>2016-03-28</td>
<td>NA</td>
</tr>
<tr>
<td>rf_17b</td>
<td>rf_17b</td>
<td>New Plymouth</td>
<td>2014-05-22 09:16:00</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Re-user</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_18</td>
<td>rf_18</td>
<td>New Plymouth</td>
<td>2014-05-14 11:20:00</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_19</td>
<td>rf_19</td>
<td>New Plymouth</td>
<td>2014-05-22 13:37:00</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_20</td>
<td>rf_20</td>
<td>New Plymouth</td>
<td>2014-05-14 11:46:00</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rf_21</td>
<td>rf_21</td>
<td>New Plymouth</td>
<td>2014-05-20 16:30:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_22</td>
<td>rf_22</td>
<td>New Plymouth</td>
<td>2014-05-14 11:39:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_23</td>
<td>rf_23</td>
<td>New Plymouth</td>
<td>2014-05-15 15:51:00</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_24</td>
<td>rf_24</td>
<td>New Plymouth</td>
<td>2014-05-14 11:36:00</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_25</td>
<td>rf_25</td>
<td>New Plymouth</td>
<td>2014-06-18 13:57:00</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_26</td>
<td>rf_26</td>
<td>New Plymouth</td>
<td>2014-06-11 13:34:00</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_27</td>
<td>rf_27</td>
<td>New Plymouth</td>
<td>2014-07-03 15:37:00</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_28</td>
<td>rf_28</td>
<td>Hawkes Bay</td>
<td>2015-01-20 12:15:00</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_29</td>
<td>rf_29</td>
<td>Hawkes Bay</td>
<td>2015-02-10 11:39:00</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_30</td>
<td>rf_30</td>
<td>Hawkes Bay</td>
<td>2015-02-03 10:58:00</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_31</td>
<td>rf_31</td>
<td>Hawkes Bay</td>
<td>2015-02-09 08:05:00</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_32</td>
<td>rf_32</td>
<td>Hawkes Bay</td>
<td>2015-02-09 08:35:00</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_33</td>
<td>rf_33</td>
<td>Hawkes Bay</td>
<td>2015-02-09 16:05:00</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_34</td>
<td>rf_34</td>
<td>Hawkes Bay</td>
<td>2015-01-06 10:50:00</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
<tr>
<td>rf_35</td>
<td>rf_35</td>
<td>Hawkes Bay</td>
<td>2015-02-05 16:00:00</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
</tr>
</tbody>
</table>
7 Code examples
We have provided a number of code examples (https://github.com/CfSOtago/GREENGridData/tree/master/examples) for suggestions on how to load, further process and analyse the data.

8 Known issues
We maintain a known data issues list (https://github.com/CfSOtago/GREENGridData/labels/dataIssue) via our GitHub repository. If you think there is a data issue please check the repo list first and then add a new one if appropriate.

9 Runtime
Analysis completed in 5.34 seconds (0.09 minutes) using knitr (https://cran.r-project.org/package=knitr) in RStudio (http://www.rstudio.com) with R version 3.5.1 (2018-07-02) running on x86_64-apple-darwin15.6.0.

10 R environment
10.1 R packages used
- base R (R Core Team 2016)
- bookdown (Xie 2016a)
- GREENGridData (Anderson and Eyers 2018) which depends on:
  - data.table (Dowle et al. 2015)
  - dplyr (Wickham and Francois 2016)
  - hms (Müller 2018)
  - lubridate (Grolemund and Wickham 2011)
  - progress (Csárdi and FitzJohn 2016)
  - readr (Wickham, Hester, and Francois 2016)
  - readxl (Wickham and Bryan 2017)
  - reshape2 (Wickham 2007)
  - ggplot2 (Wickham 2009)
  - kableExtra (Zhu 2018)
  - knitr (Xie 2016b)
  - rmarkdown (Allaire et al. 2018)
10.2 Session info

```r
# R version 3.5.1 (2018-07-02)
# Platform: x86_64-apple-darwin15.6.0 (64-bit)
# Running under: macOS High Sierra 10.13.6
#
# Matrix products: default
# BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
# LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
#
# locale:
#
# attached base packages:
# [1] stats     graphics  grDevices utils     datasets  methods   base
#
# other attached packages:
# [1] stringr_1.3.1     bindrcpp_0.2.2    kableExtra_0.9.0
# [4] bookdown_0.7      rmarkdown_1.10    knitr_1.20.13
# [7] skimr_1.0.3       readxl_1.1.0      readr_1.1.1
# [10] lubridate_1.7.4   ggplot2_3.0.0     data.table_1.11.4
# [13] GREENGridData_1.0
#
# loaded via a namespace (and not attached):
# [1] Rcpp_0.12.18      highr_0.7         cellranger_1.1.0
# [4] pillar_1.3.0      compiler_3.5.1    plyr_1.8.4
# [7] bindr_0.1.1       prettyunits_1.0.2 tools_3.5.1
# [10] progress_1.2.0    digest_0.6.15     viridisLite_0.3.0
# [13] evaluate_0.11     tibble_1.1.2       dplyr_0.7.6
# [16] pkgconfig_2.0.2   rlang_0.2.2       rstudioapi_0.7
# [19] yaml_2.2.0        xfun_0.3          xml2_1.2.0
# [22] httr_1.3.1        withr_2.1.2       dplyr_0.7.6
# [25] hms_0.4.2         rprojroot_1.3-2   grid_3.5.1
# [28] tidyselect_0.2.4  glue_1.3.0        R6_2.2.2
# [31] reshape2_1.4.3    magrittr_1.5      scales_1.0.0
# [34] backports_1.1.2   assertthat_0.2.0  colorspace_1.3-2
# [37] rvest_0.3.2       lazyeval_0.2.1   crayon_1.3.4
#```

References


