Australia's National Science Agency

Weather sensitivity and battery identification

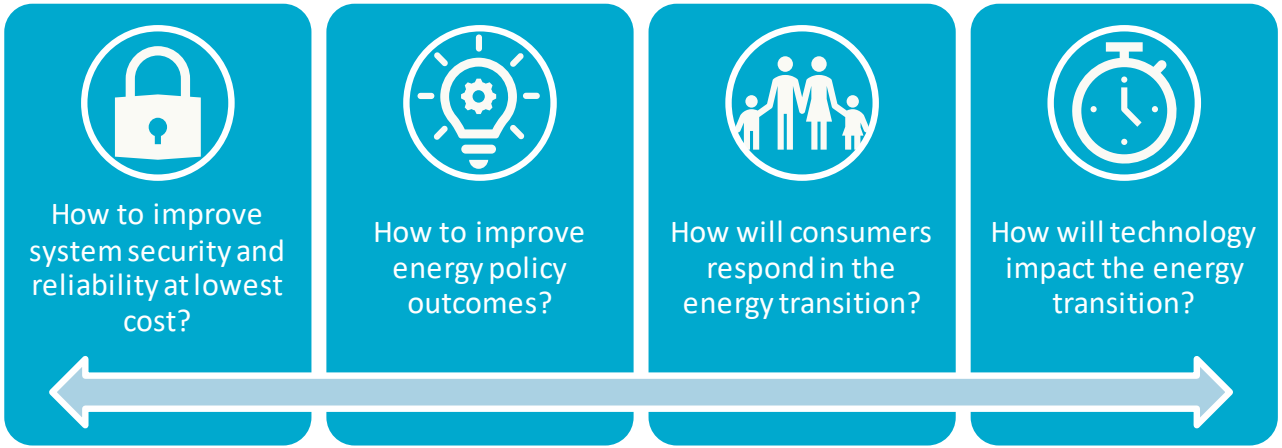
The National Energy Analytics Research (NEAR) Program's 2019/20 work with AEMO

Lachlan O'Neil | 28/10/2020

Research from: Lachlan O'Neil, Dilusha Weeraddana, Weihong Wang, Leala Darby, Peter Goldthorpe, Md. Mashud Rana, Khoa Nguyen, Andrew Spiers and Paul Graham

In collaboration with Greg Staib, Debborah Marsh, Luke Dowling, Ali Habibi Khalaj

Key research/policy themes for the NEAR Program



Australian Government
Department of Industry, Science, Energy and Resources

AEMO
AUSTRALIAN ENERGY MARKET OPERATOR

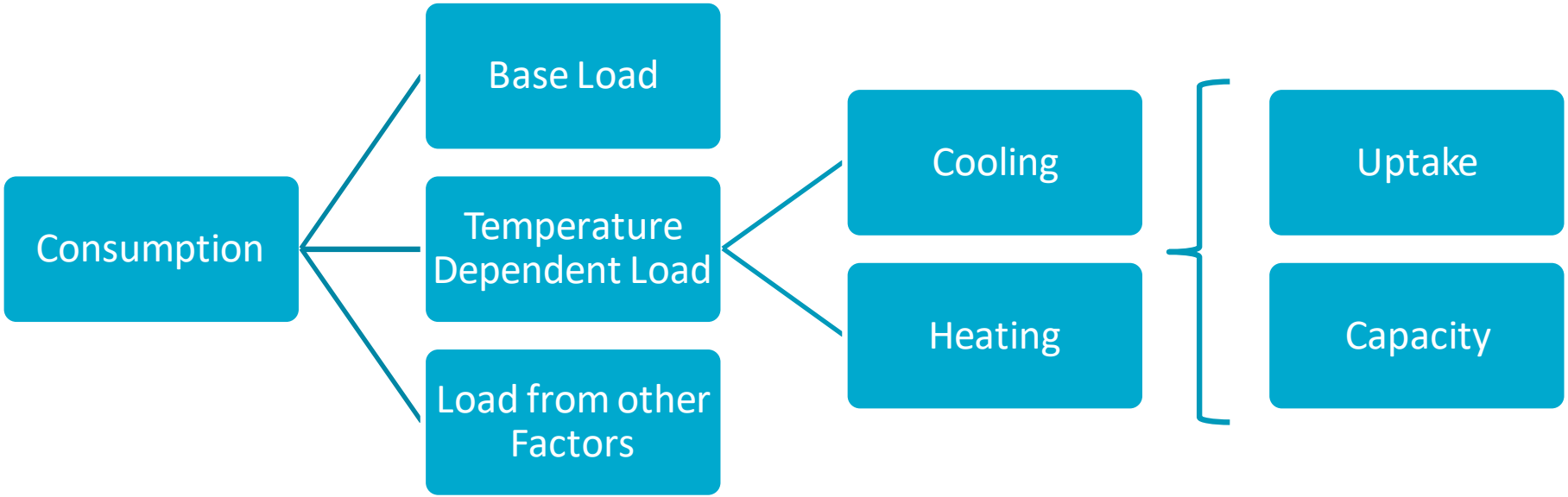
CSIRO

Weather sensitivity

Weather sensitivity

The challenge

Understanding components of consumption



Hypothesis 1 - a change in sensitivity to temperature over time, is due to changes in uptake and capacity

Hypotheses

- Hypothesis 1
  - A change in community sensitivity to temperature over time, is due to changes in uptake and capacity.
- Hypothesis 2
  - Other factors such as dwelling type or postcode will affect the heating and cooling capacity of a home.

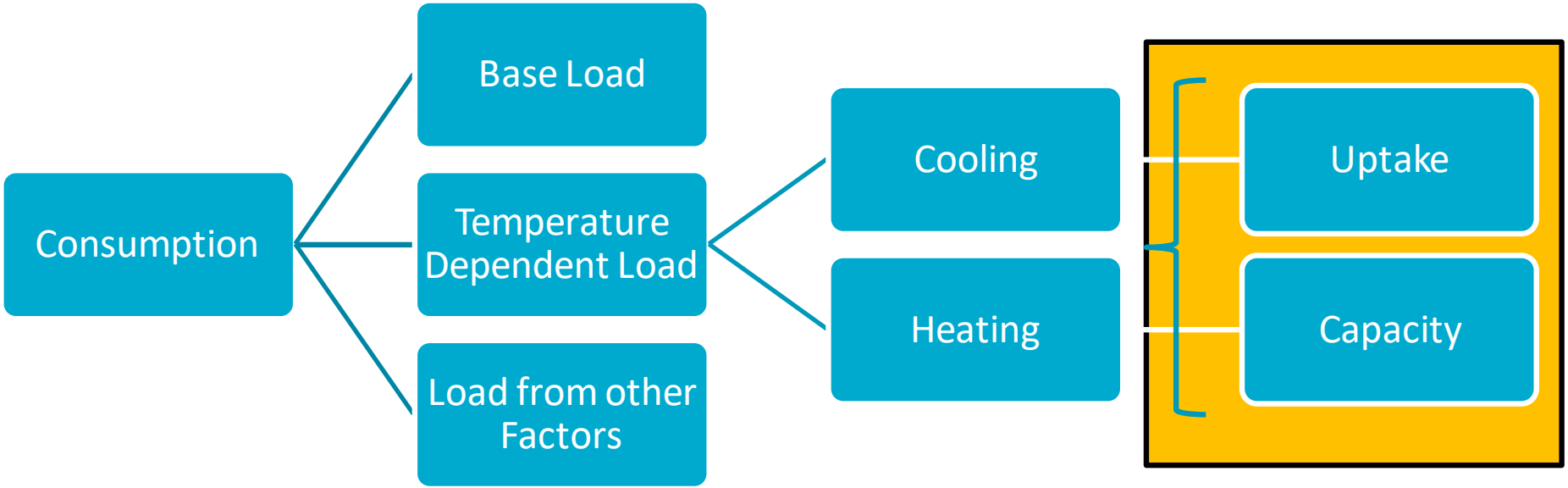
Heating & Cooling Capacity and Uptake are Critical to Forecasting

Heating and cooling capacity…

- Has a large effect on grid demand
- Is largely unknown
- Could be changing over time
- Could be calculated by observing consumption and temperature
- Will assist long-term forecasting if known*

*Though there are also ways in which this could assist operational forecasting, these models were developed specifically with the long-term forecasting benefits in mind.

Understanding components of consumption



Weather sensitivity

The methodology

Data limitations on preliminary results

- Only working within Victoria
- Only 5 postcodes
- Non-solar households only
- Sampled 1,000 households
- Only households with data from March 2014 through to February 2019

Note: More recent data is available this financial year (though we'll have to take into account COVID-19)

The five postcodes

Melbourne
Werribee
Ballarat
Bendigo
Cranbourne

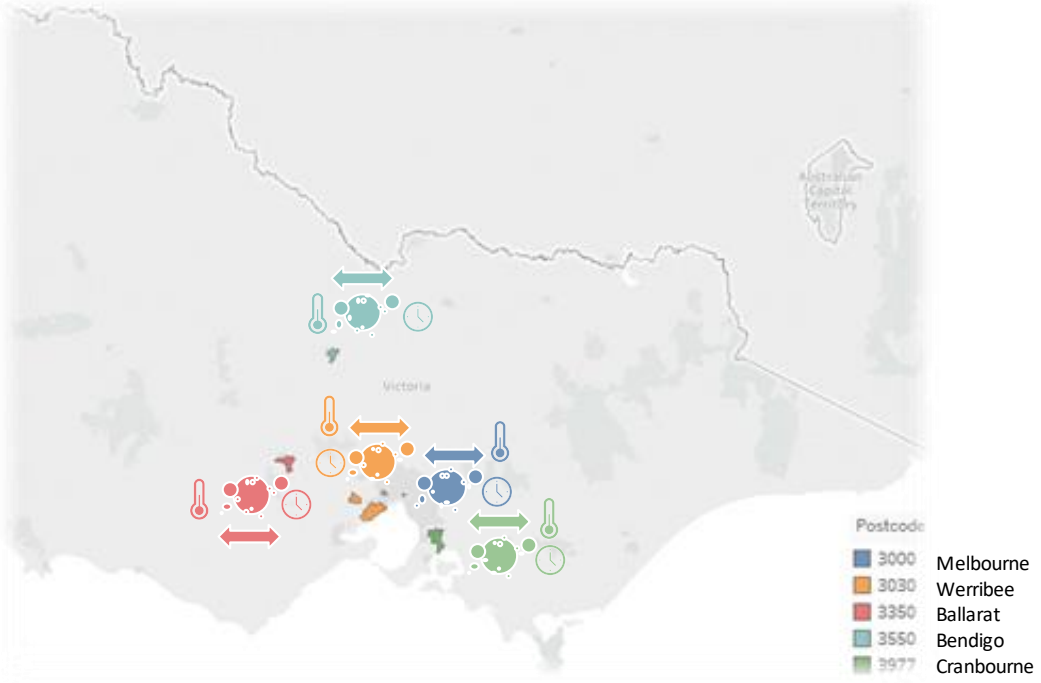
Sampling

Sample 1,000 households

Ensure they have consumption data for 2014-2019

Split data into seasons

Link with nearest weather station



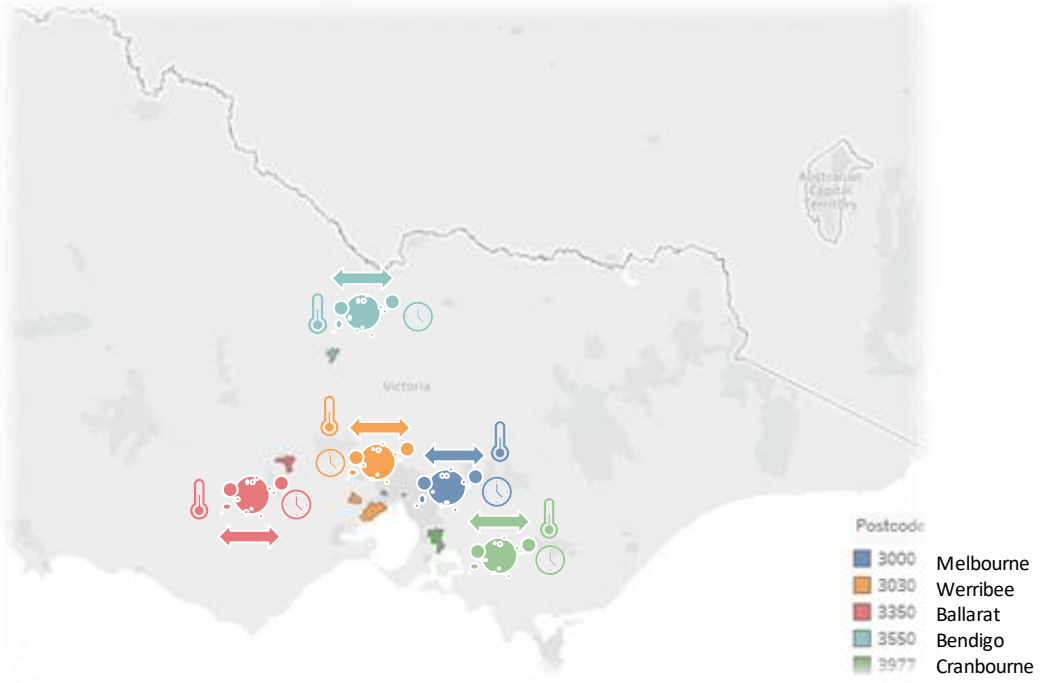
Methodology (1/3)

Sample 1,000 households

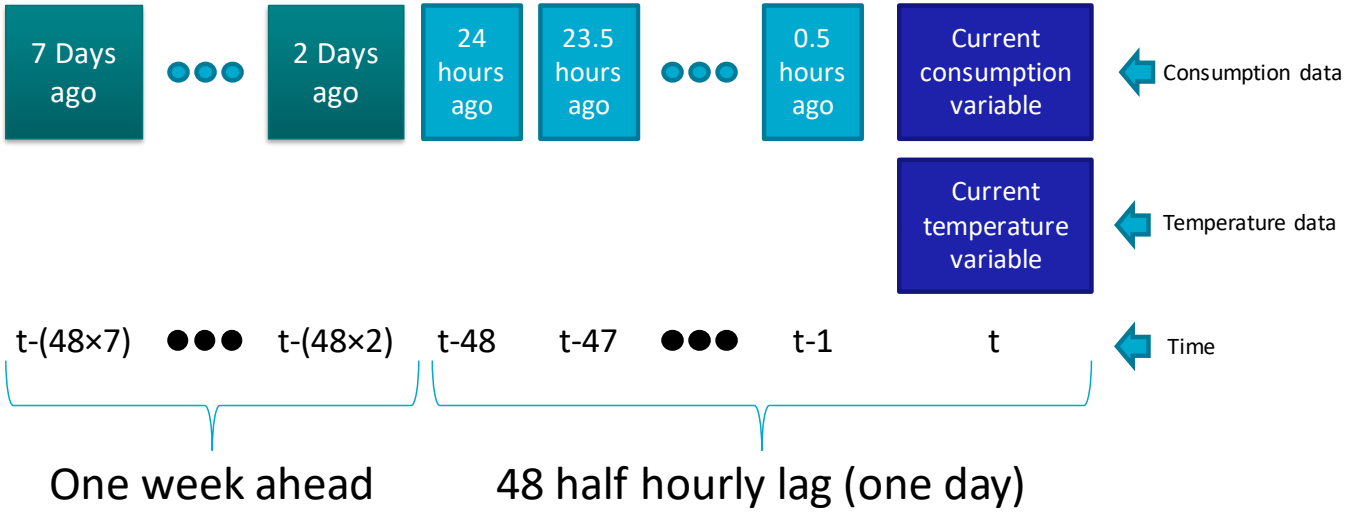
Ensure they have consumption data for 2014-2019

Split data into seasons

Link with nearest weather station



Methodology (2/3)



Methodology (3/3)

The final statistical distribution ends up looking like the below:

$$C_t \sim \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_{48} C_{t-48} + \beta_{49} C_{t-2\times 48} + \beta_{50} C_{t-3\times 48} + \cdots + \beta_{54} C_{t-7\times 48} + \beta_{55} T_t + \epsilon$$

C_τ = Consumption at half hourly period τ
T_τ = Temperature at half hourly period τ
β_n = Linear coefficient for the nth regressor
ϵ = The error term

Methodology (3/3)

The final statistical distribution ends up looking like the below:

$$C_t \sim \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_{48} C_{t-48} + \beta_{49} C_{t-2\times 48} + \beta_{50} C_{t-3\times 48} + \cdots + \beta_{54} C_{t-7\times 48} + \underline{\beta_{55}} T_t + \epsilon$$

The coefficient of temperature

C_τ = Consumption at half hourly period τ
T_τ = Temperature at half hourly period τ
β_n = Linear coefficient for the nth regressor
ϵ = The error term

Methodology (3/3)

The final statistical distribution ends up looking like the below:

$$C_t \sim \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_{48} C_{t-48} + \beta_{49} C_{t-2\times 48} + \beta_{50} C_{t-3\times 48} + \cdots + \beta_{54} C_{t-7\times 48} + \underline{\beta_{55}} T_t + \epsilon$$

Detrends based on previous consumption

The coefficient of temperature

C_τ = Consumption at half hourly period τ

T_τ = Temperature at half hourly period τ

β_n = Linear coefficient for the nth regressor

ϵ = The error term

Methodology

The final statistical distribution ends up looking like the below:

$$C_t \sim \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_{48} C_{t-48} + \beta_{49} C_{t-2\times 48} + \beta_{50} C_{t-3\times 48} + \cdots + \beta_{54} C_{t-7\times 48} + \underline{\beta_{55}} T_t + \epsilon$$

Detrends based on previous consumption

The coefficient of temperature

C_τ = Consumption at half hourly period τ

T_τ = Temperature at half hourly period τ

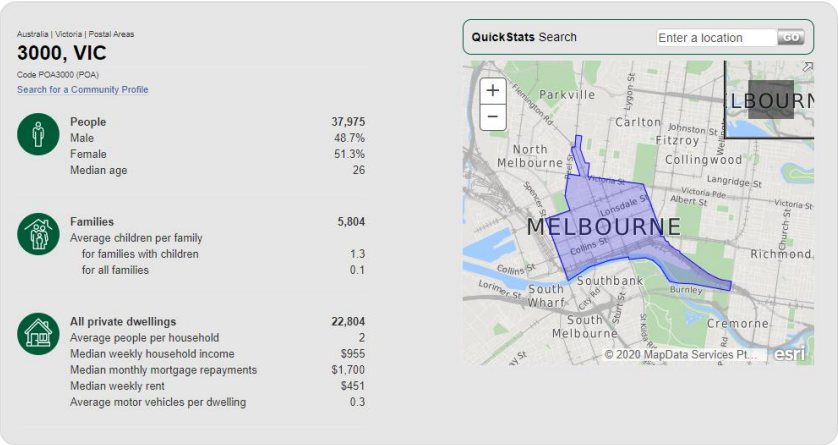
β_n = Linear coefficient for the nth regressor

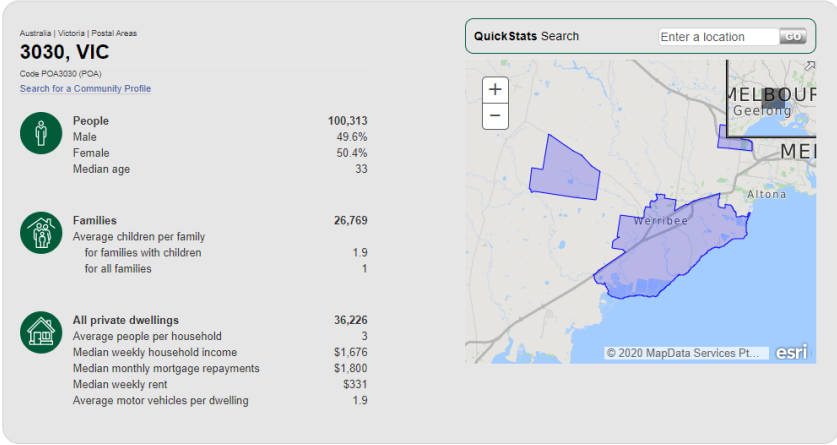
ϵ = The error term

Weather sensitivity

Deep dive

A case study





Dwelling Structure	3000 - Melbourne	3030 - Werribee
Separate House	0.2%	84%
Flat, apartment, etc.	98.1%	15.7%
Other	0.4%	0.2%

Stats and screenshots taken from ABS's 2016 Census QuickStats – Number do not add to 100% due to noise inserted by ABS for low counts
https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20QuickStats

Recall…

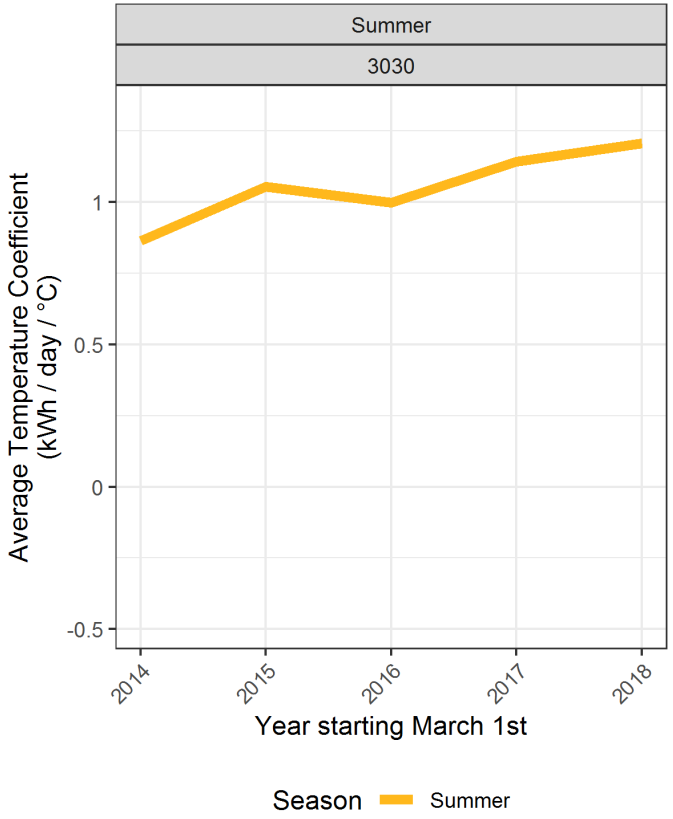
- 1,000 households sampled from each postcode
- Households must have data from March 2014 – February 2019
- Model is generated for every household, for every season, for every year
- This results in $1{,}000 \times 4 \times 5 = 20{,}000$ models
- We will be observing the temperature coefficient of these models

$$C_t \sim \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \cdots + \beta_{48} C_{t-48} + \beta_{49} C_{t-2\times 48} + \beta_{50} C_{t-3\times 48} + \cdots + \beta_{54} C_{t-7\times 48} + \underline{\beta_{55}} T_t + \epsilon$$

Detrends based on previous consumption

The coefficient of temperature

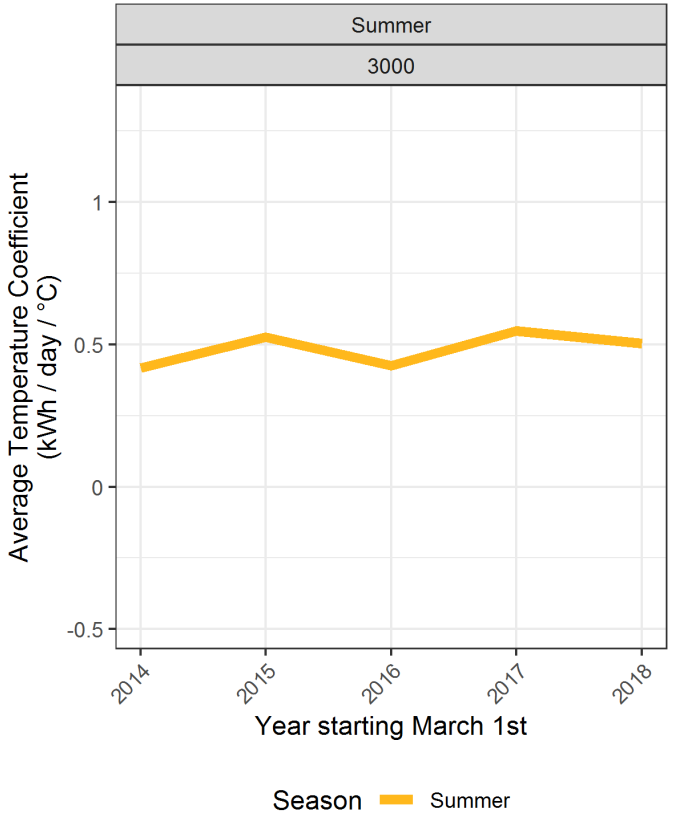
Postcode 3030 - Werribee



- On average, these households consume more electricity per degree Celsius in 2019, compared to 2014
- Our hypothesis is that they are using more space conditioning for cooling. This could be due to…
  - Installing bigger units
  - Installing more units in their house
  - Turning units on more frequently

- If heatwaves/urban heat islands become more frequent, future users of electricity may be increasingly likely to use their space cooling.
- Events like COVID-19 could change habits unlocking more latent cooling potential.

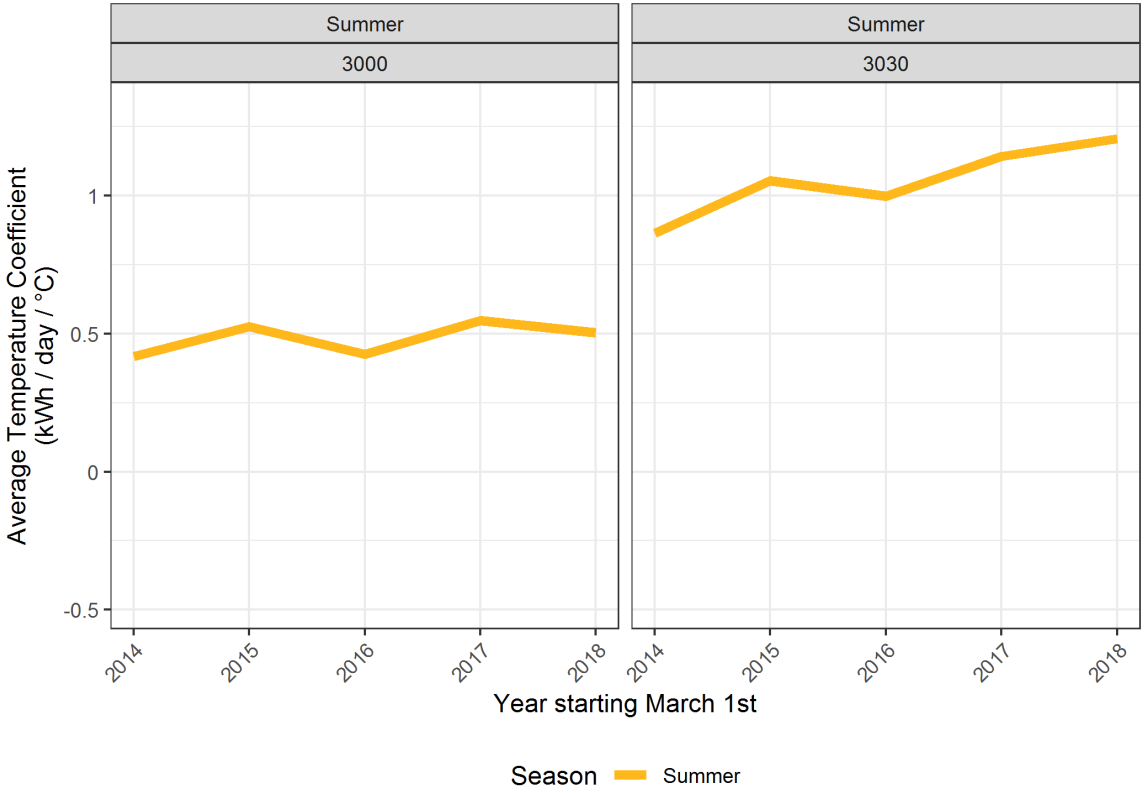
Postcode 3000 - Melbourne



- On average, these households consume around the same amount electricity per degree Celsius in 2019 as they did in 2014.
- Our hypothesis is that dwelling type (mostly apartments) might constrain the ability to install more space conditioning.
- Alternatively, these households might be far more efficient due to passive design in buildings.
- This sample is also likely to have physical dwelling sizes much more tightly distributed than separate houses.

- Even if heatwaves/urban heat islands become more frequent, future users of electricity in Melbourne may not change their space cooling.

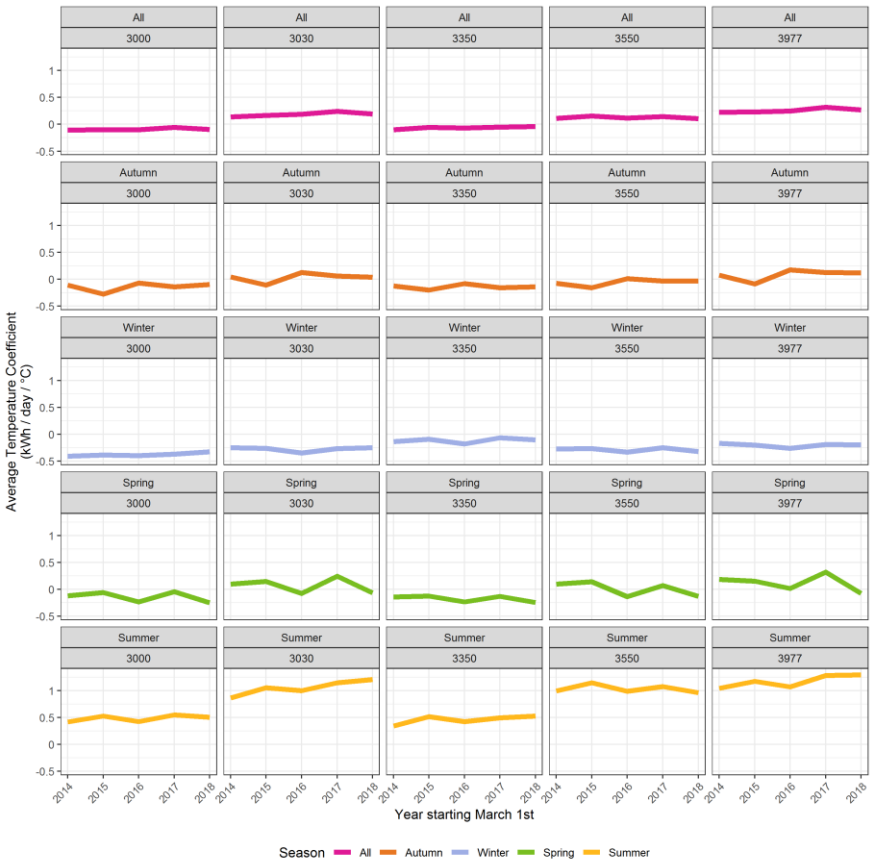
A case study



- Comparing the two suggests that perhaps an understanding of
  - Where dwellings are built
  - The type of dwellings being built

  Are important metrics for the industry to monitor and understand as population grows into both metro and outer suburbs.

Many more comparisons



Weather sensitivity

Further research

Further research

- Expand individual household modelling
  - Include households newer than 2014 in the sampled case studies.
- Deeper diver into segmentation work
  - Investigate data quality for segmenting by dwelling type, age, and postcode.
  - Generate models by segments, rather than individual households.
  - Try and confirm hypotheses around dwelling type and location.
- Expand work to more regions
  - Generate a more comprehensive view of Victoria.
  - Expand work to other states
- Look at specific events
  - Max/min temperature days
  - Max/min demand days

More further research

- Looking for suggestions!

Identifying batteries

Identifying batteries

The challenge

The DER Register

- In late 2018, the Australian Energy Market Commission (AEMC) amended the National Electricity Rules enabling AEMO to establish a DER Register
- From 1 March 2020, a regulatory requirement mandates NSPs provide additional information about DER equipment installed on their networks including:
  - Battery Storage Capacity (kWh)
  - Battery Power Capacity (kW)
  - Installation date

Data provided to AEMO

Pre- March 2020

- Historical data provided to AEMO
- Data collection was not mandatory
- Accuracy and completeness variable

Post- March 2020

- Up-to-date data provided to AEMO
- Data collection mandatory

Data provided to AEMO

Pre- March 2020

- Historical data provided AEMO
- Data collection was not ndatory
- Accuracy and complete s variable

Post- March 2020

- Up-to-date data provided to AEMO
- Data collection mandatory

Detecting households with batteries

- Use AEMO's MSATS database (half hourly energy recordings)
- Train models based on known battery households
- Combine machine learning with domain expertise
- The result – Identifying existing batteries, to support the forecasting of future battery numbers

Impact

- Correctly predicting the existence of batteries can be used to better inform the starting point of future battery forecasts.
- An additional source of info to (for example) SunWiz for cross-reference an improved certainty.
- Expanded sets of known battery households for analysis.

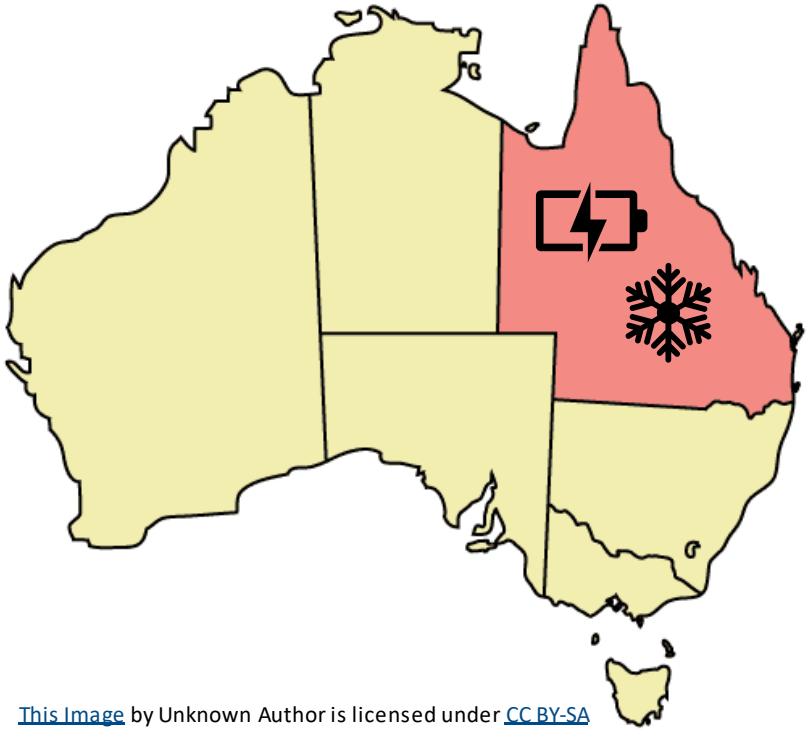
Identifying batteries

The approach

Data limitations on preliminary results

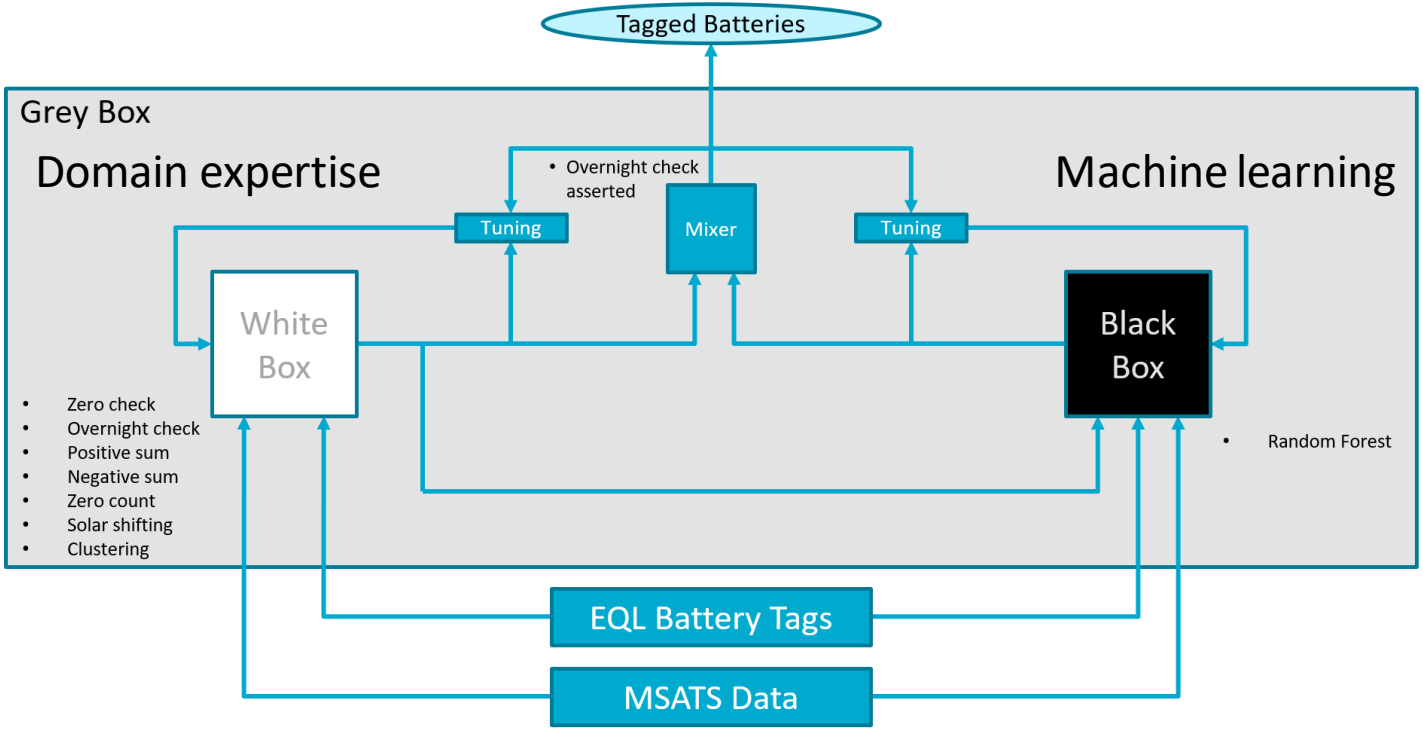
- Only working within Queensland
- Only set of ~1,500 known battery households
- Only data for 4 months for validation (Winter)

Note: More data is available this financial year –expect results to improve for minimal new overhead

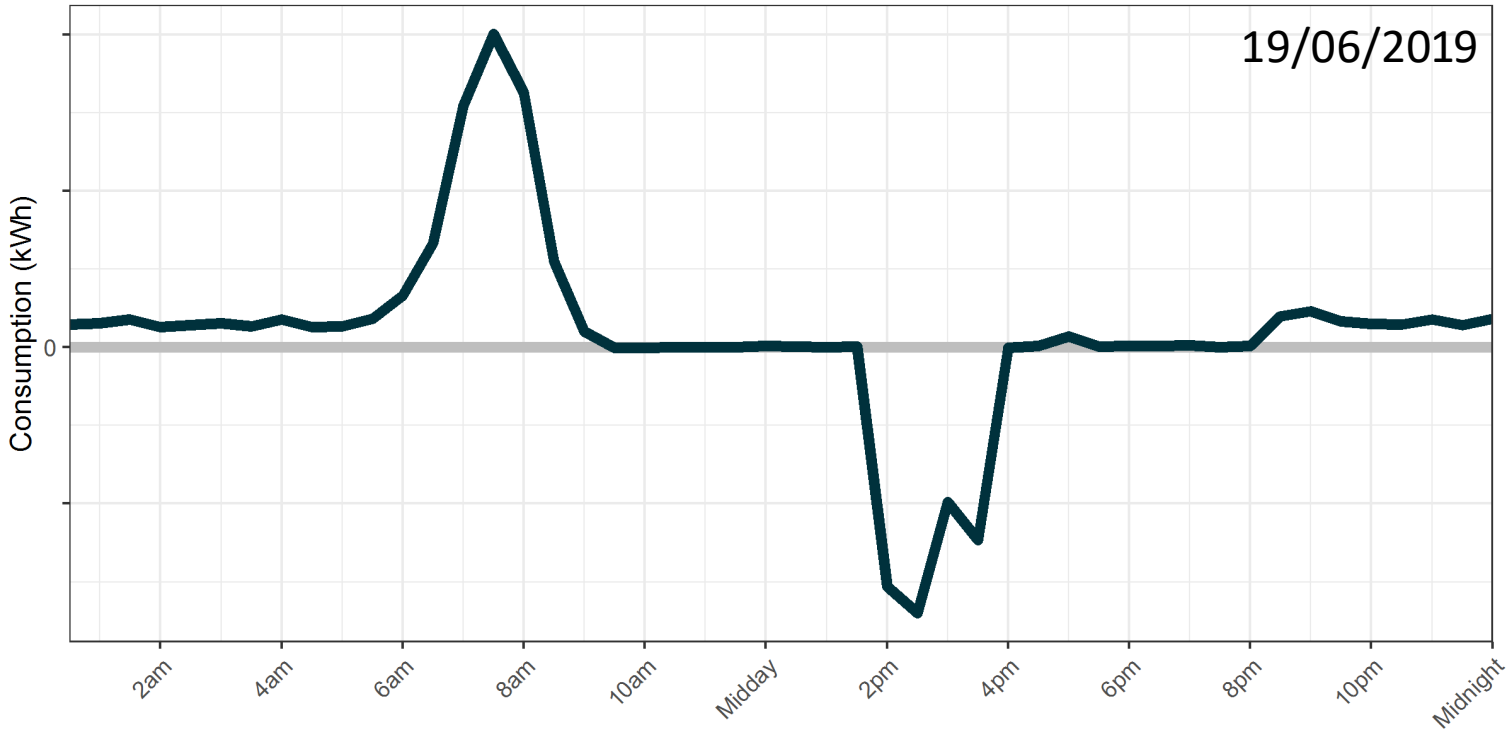


This Image by Unknown Author is licensed under CC BY-SA

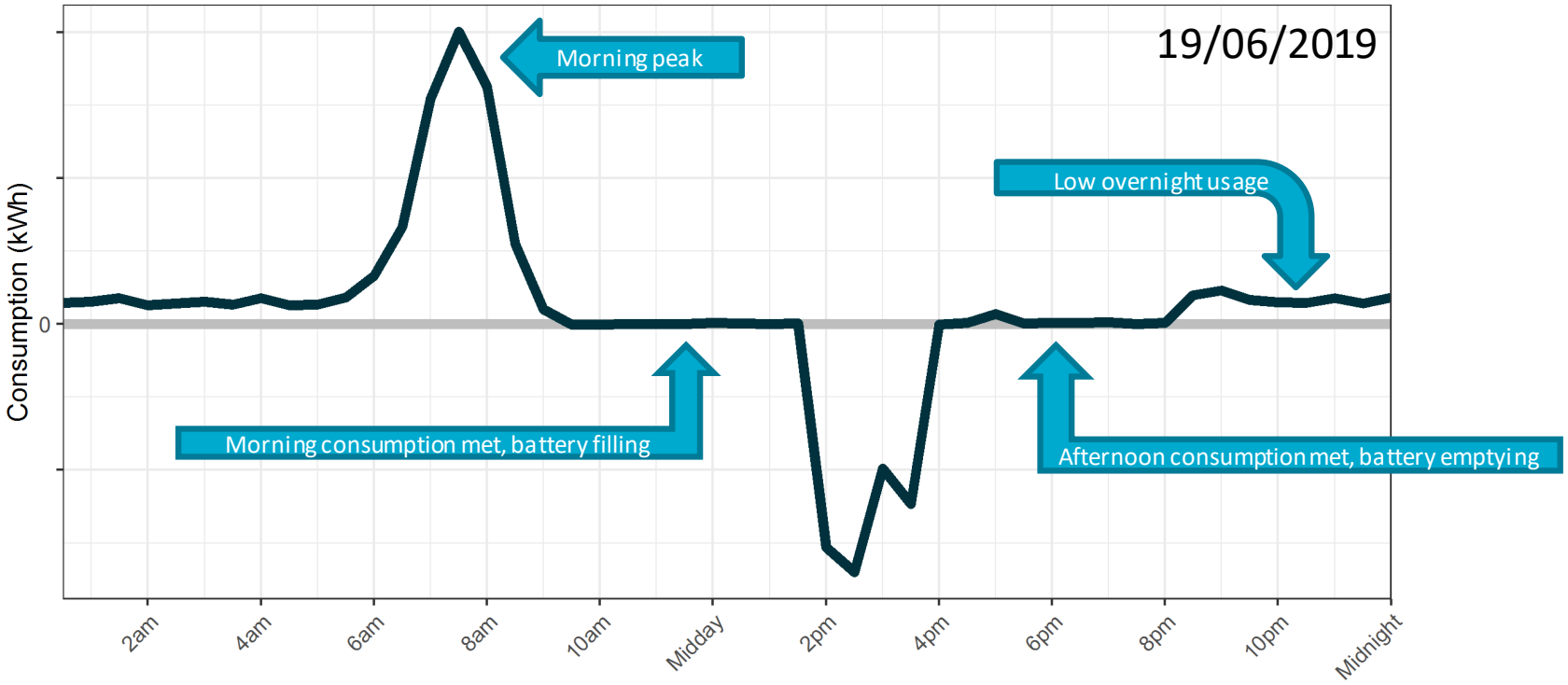
Methodology



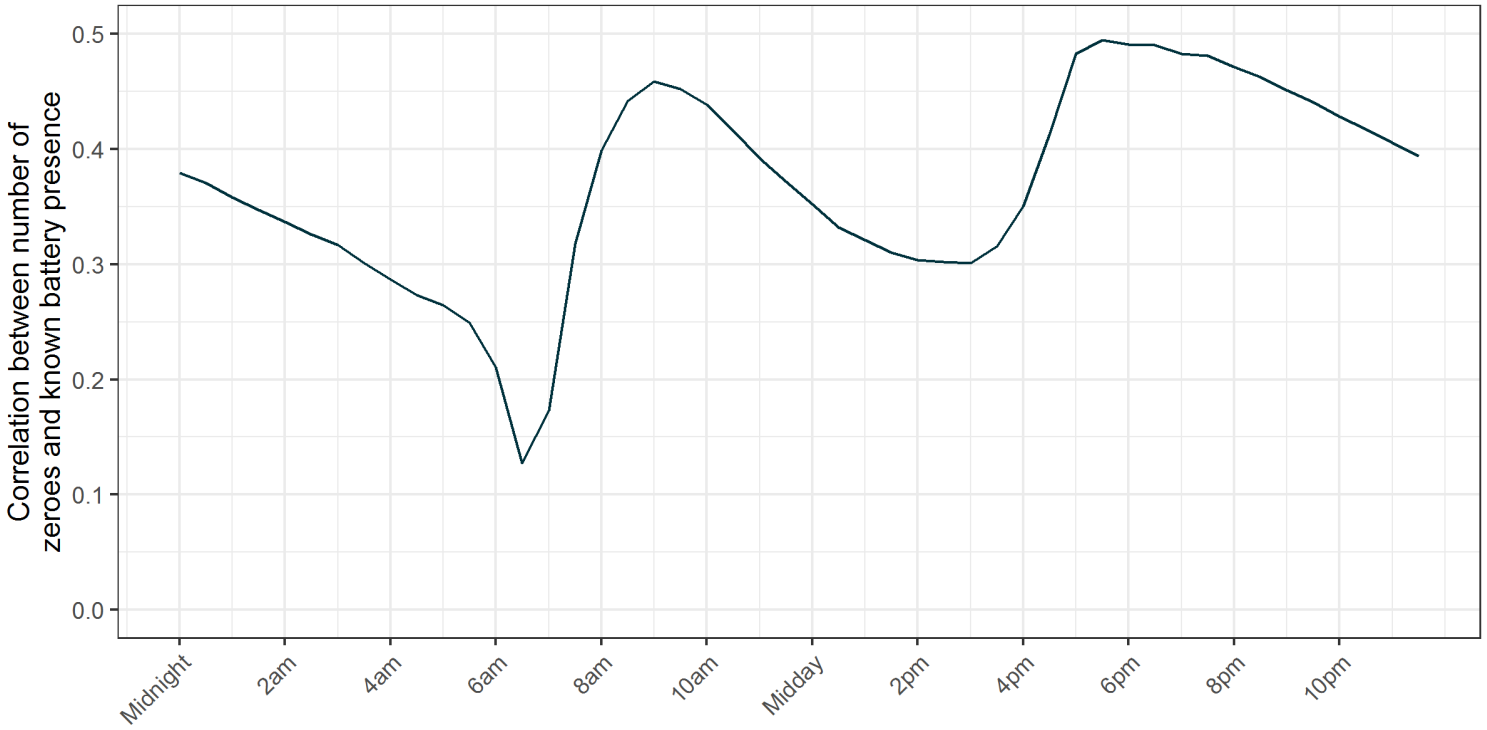
‘Typical’ battery household



‘Typical’ battery household



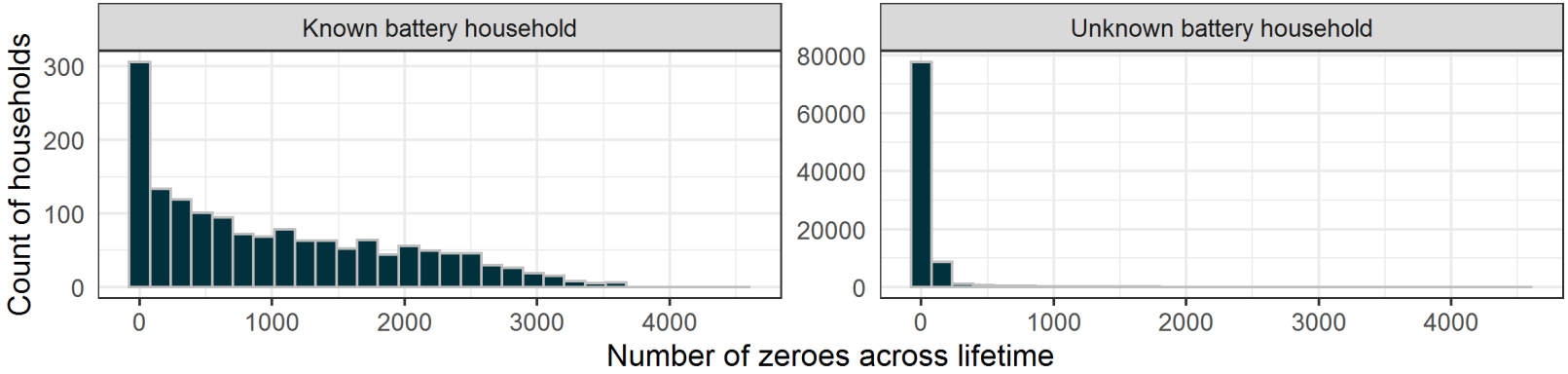
Whitebox - Zeroes



Whitebox - Zeroes

- Mean zeroes for May-August 2019 per ***unknown*** *battery household*
  - 71.5

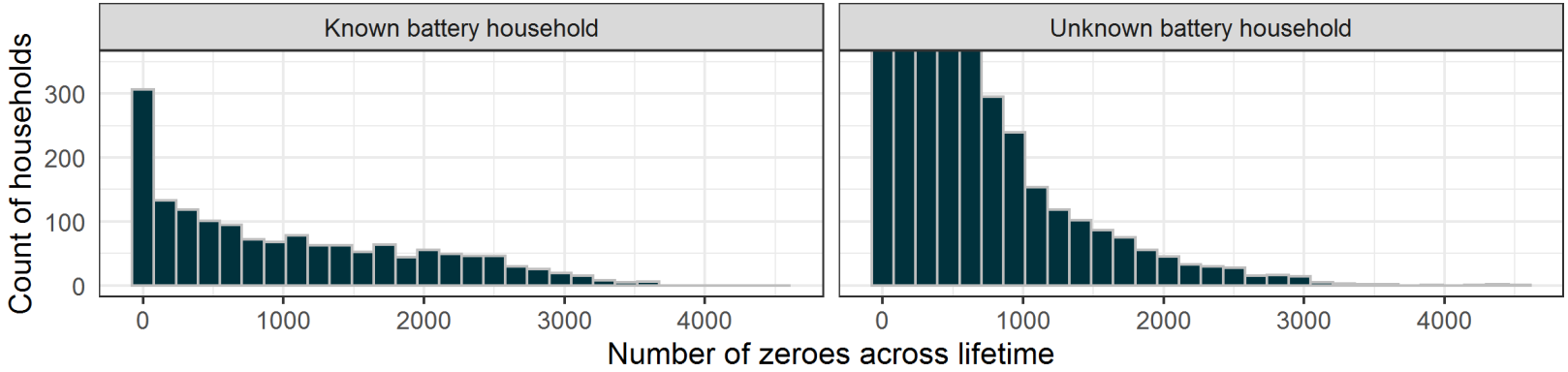
- Mean zeroes for May-August 2019 per ***known*** *battery household*
  - 1,012.7



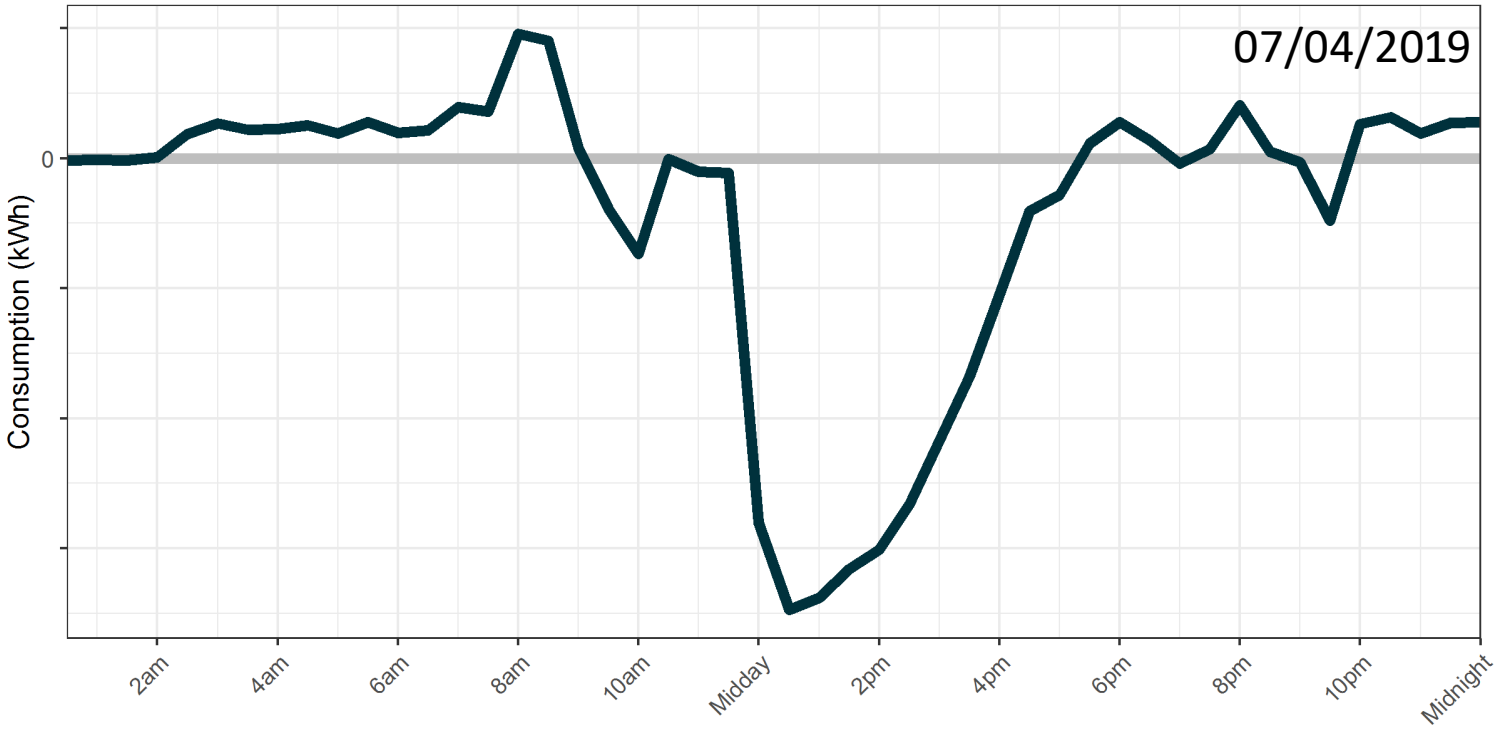
Whitebox - Zeroes

- Mean zeroes for May-August 2019 per ***unknown*** *battery household*
  - 71.5

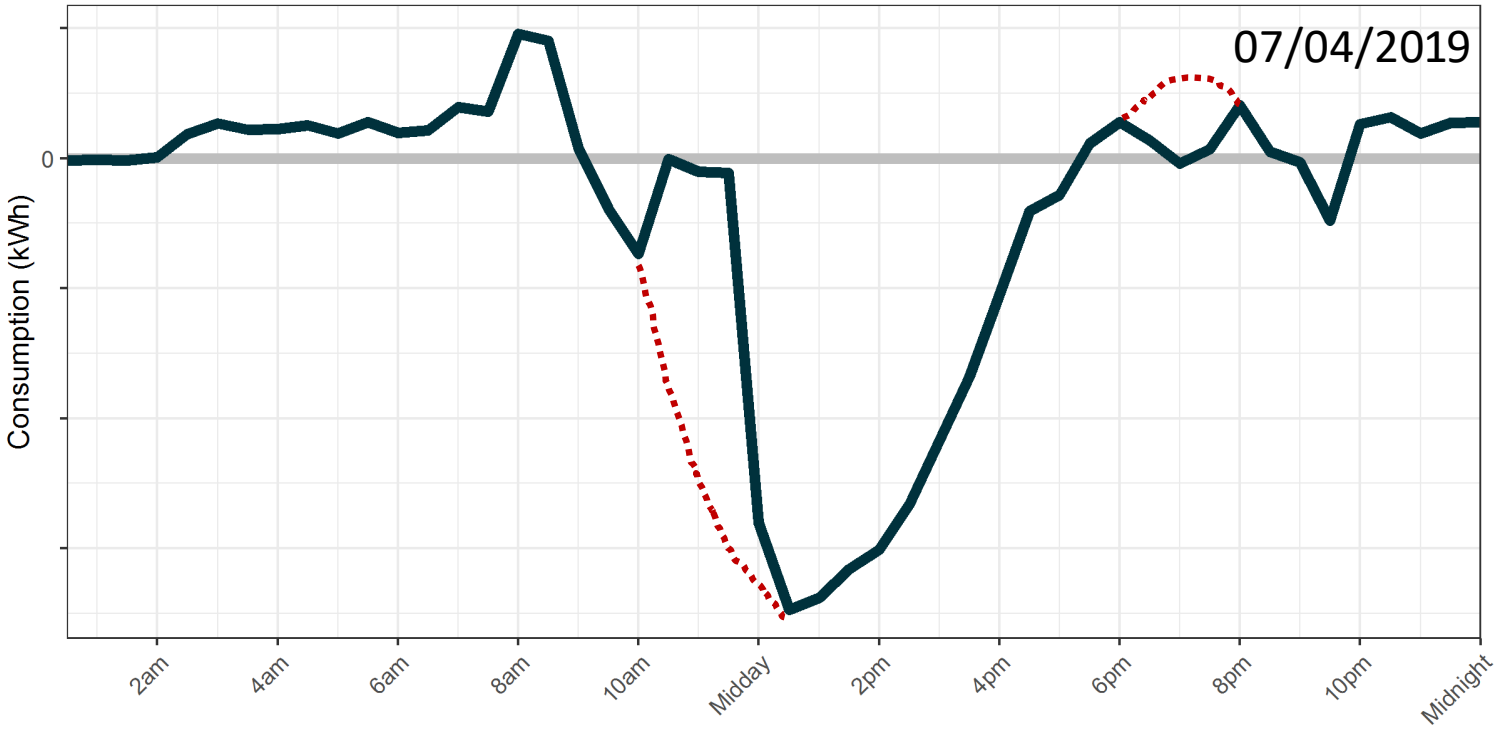
- Mean zeroes for May-August 2019 per ***known*** *battery household*
  - 1,012.7



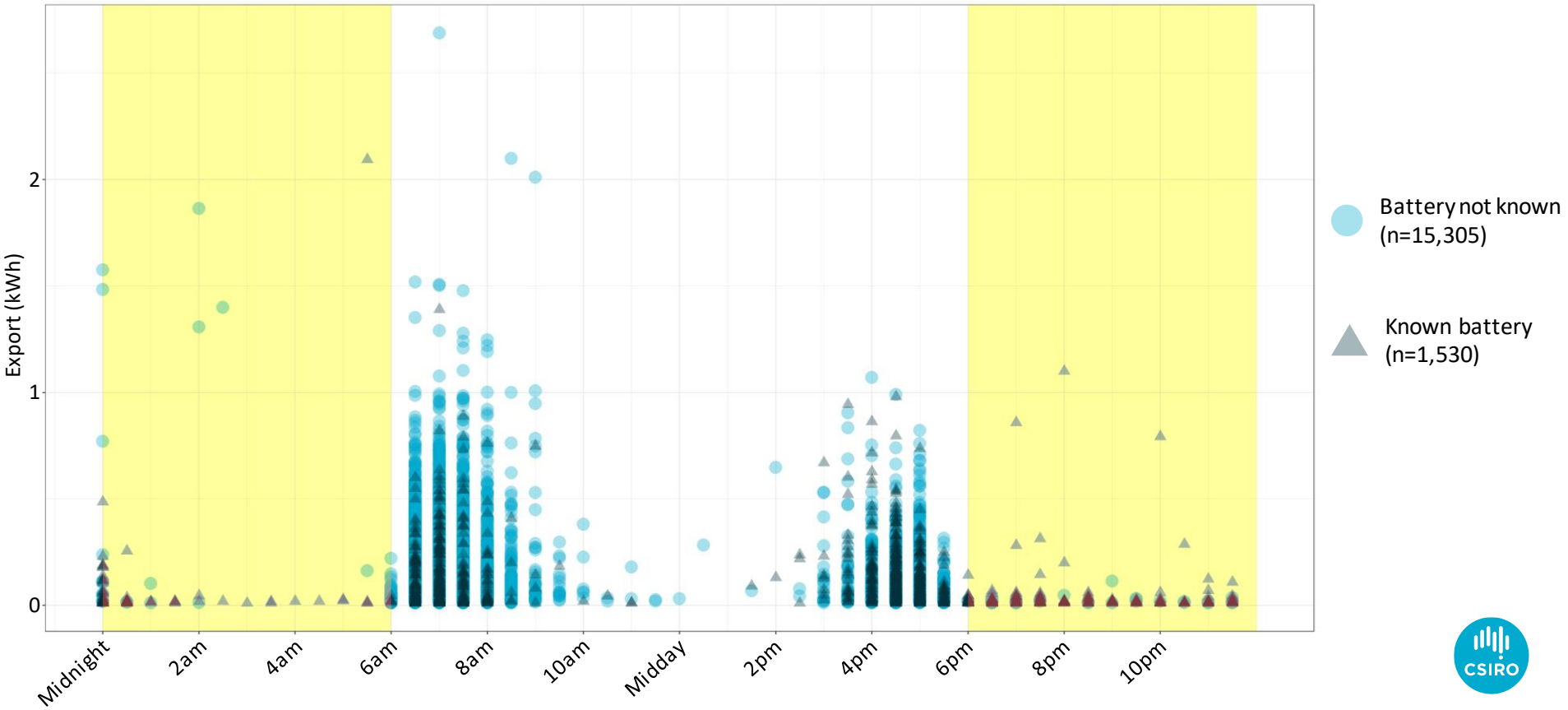
Whitebox – Overnight export



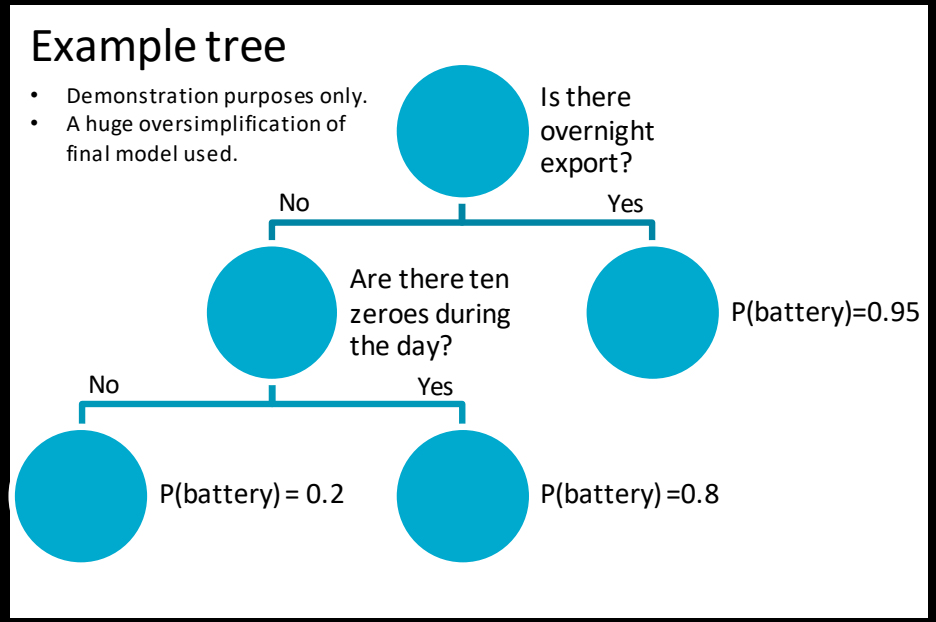
Whitebox – Overnight export



Whitebox - Overnight export



Blackbox – random forest

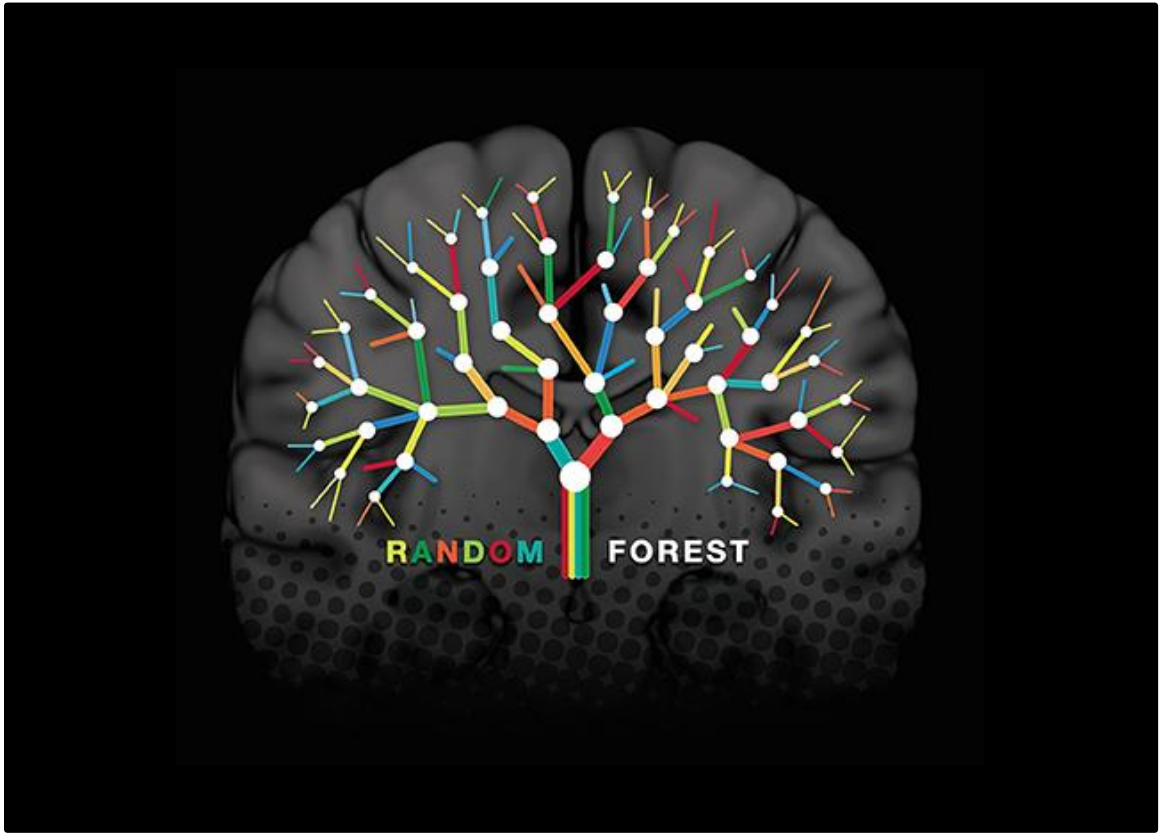


Blackbox – random forest

Consumption Data

EQL Battery Tags

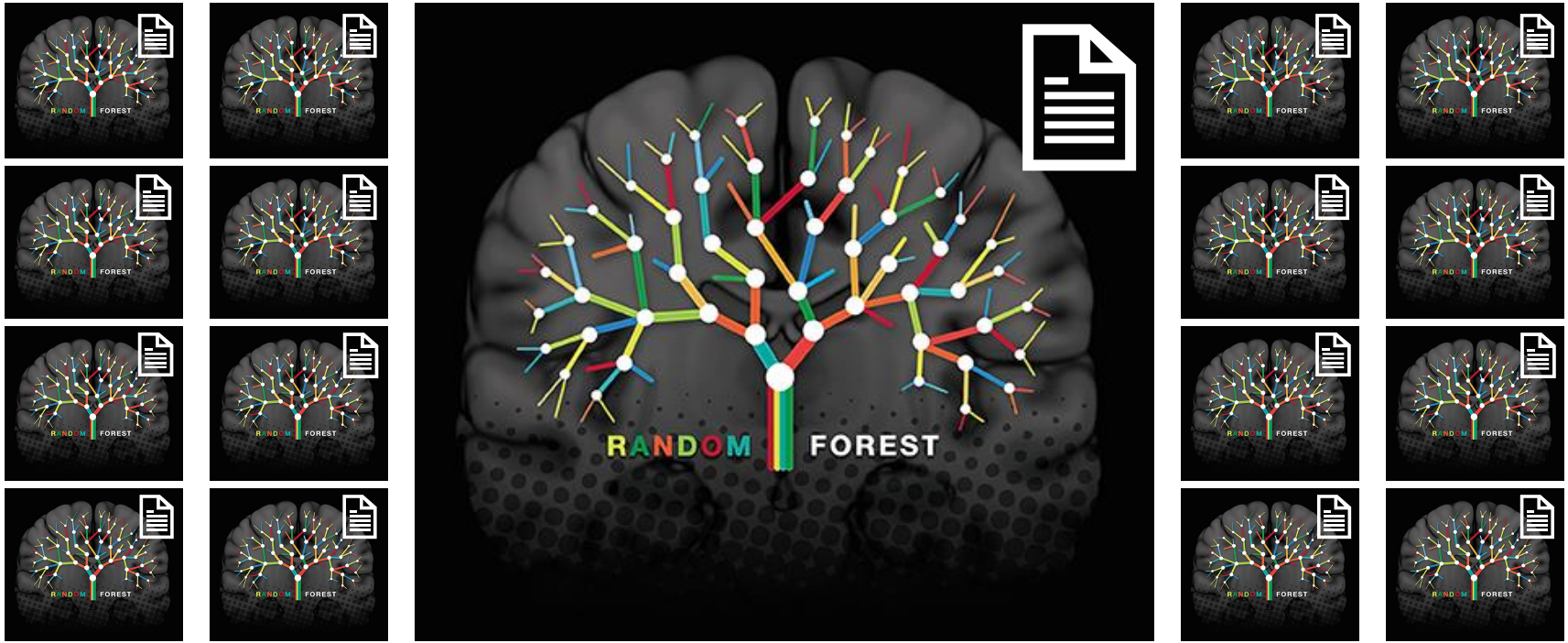
Whitebox metrics



Prediction

This Photo by Unknown Author is licensed under CC BY

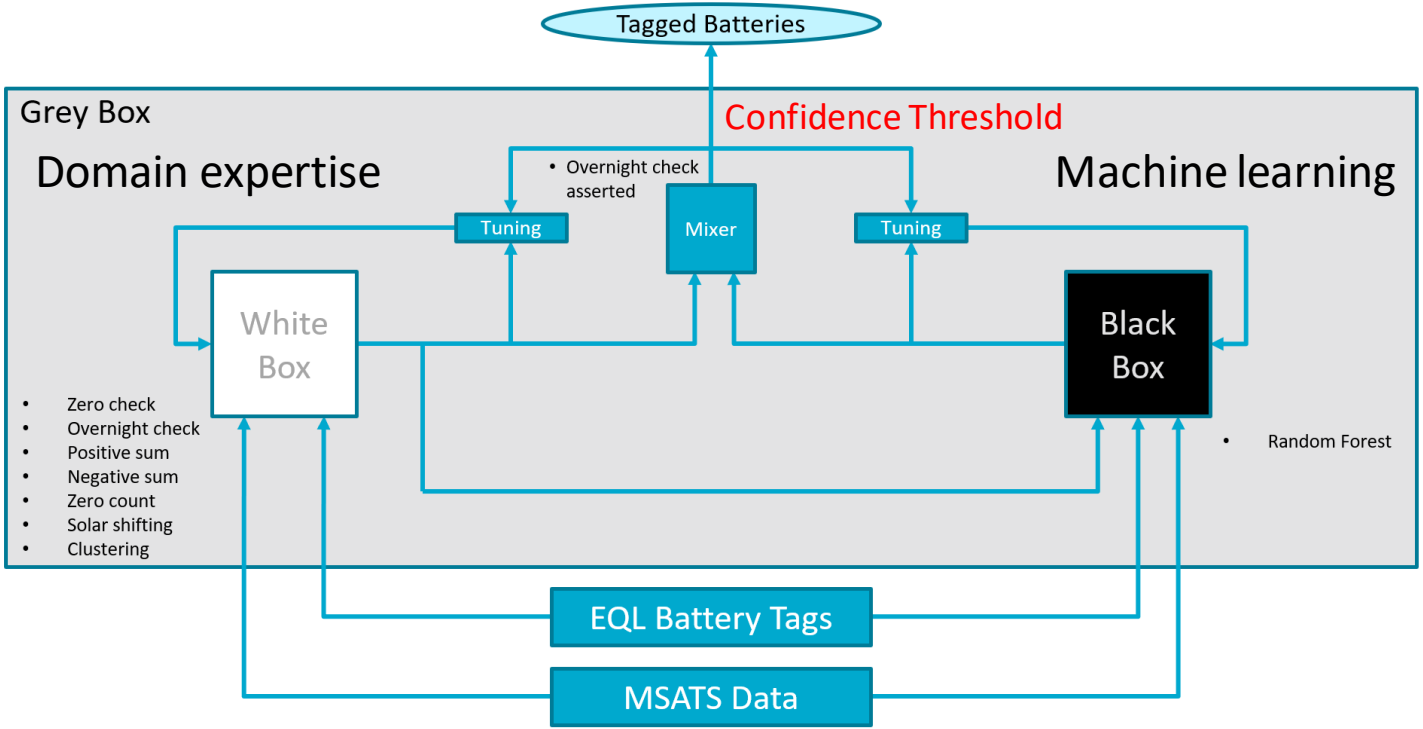
Blackbox – random forest



This Photo by Unknown Author is licensed under CC BY

Further reading on random forests can be found here: https://towardsdatascience.com/understanding-random-forest-58381e0602d2

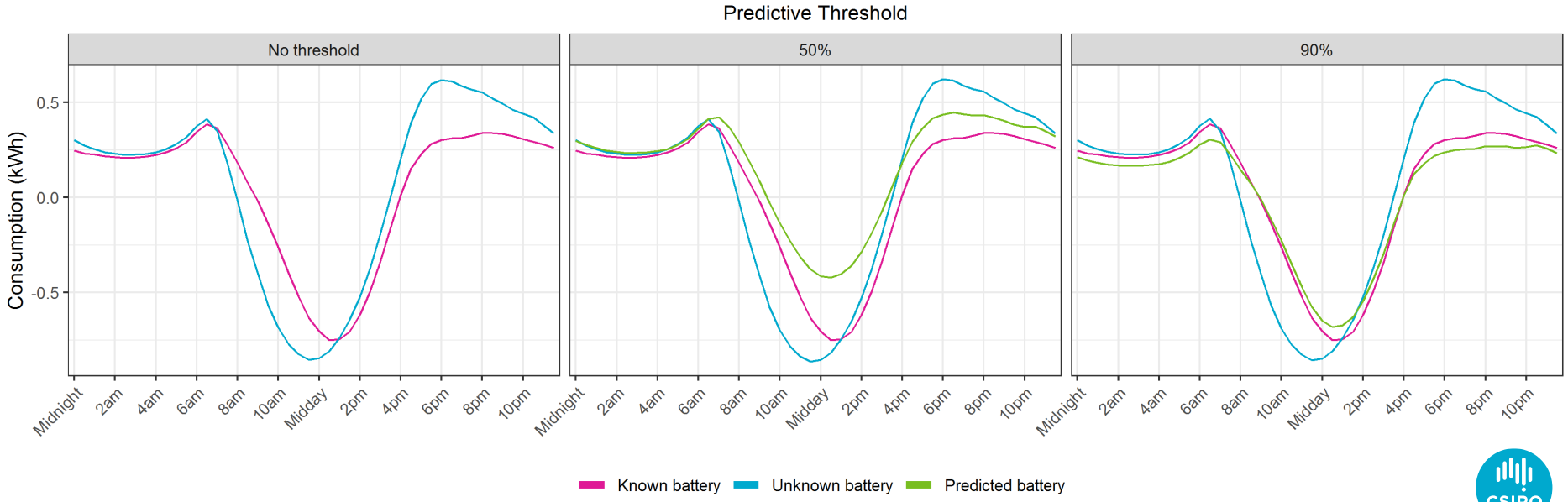
Methodology



Predicting batteries

- Households can be grouped by the threshold predicted for them
- Different groups will exhibit different diurnal behaviours

- As we increase our certainty, the average diurnal profile of predicted household looks increasingly like the average diurnal profile of known battery households



Queensland predicted counts

Threshold	Count	Extrapolated	Total Batteries	Percentage
70%	1,220	3,432	6,986	7.66%
80%	989	2,928	6,482	7.10%
90%	768	2,481	6,035	6.61%
CER small-scale systems (2014-2019)			5,329	5.84%

Threshold – The probability of the household having a battery, according to our grey box model. Choosing which threshold to use is a subject of further research.

Count – The number of households predicted as having a battery that weren't previously known

Extrapolated – The number of households we think would have batteries that weren't previously known by applying proportions to non-interval metered households

Total batteries – The total battery estimate for Queensland based on the EQL numbers and the extrapolated numbers

Percentage – The percentage of Queensland solar households that have batteries based on the total batteries estimate and the ~91k solar households estimated by this study.

Identifying batteries

Further research

Expand work

- Use updated consumption data
  - Can look across all seasons, rather than just Winter.
- Use the DER register to validate work in other regions
  - Victoria
  - New South Wales
  - South Australia
  - Tasmania

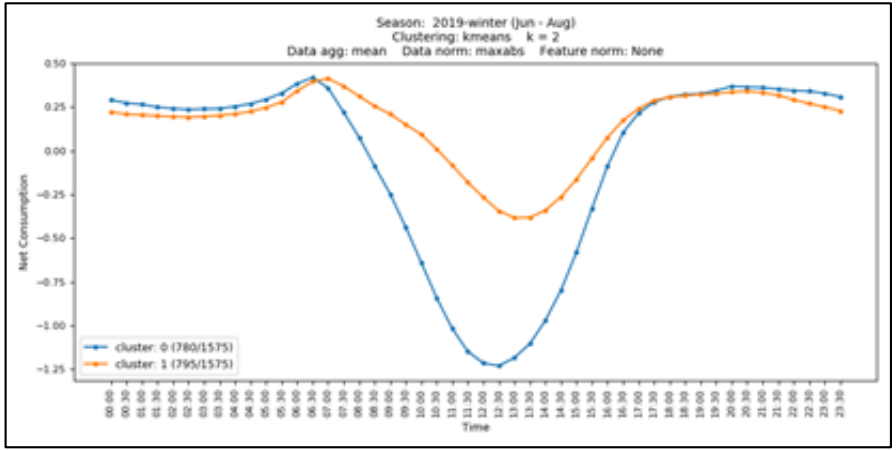
Detect control schemes

- First approach - shape clustering of battery households

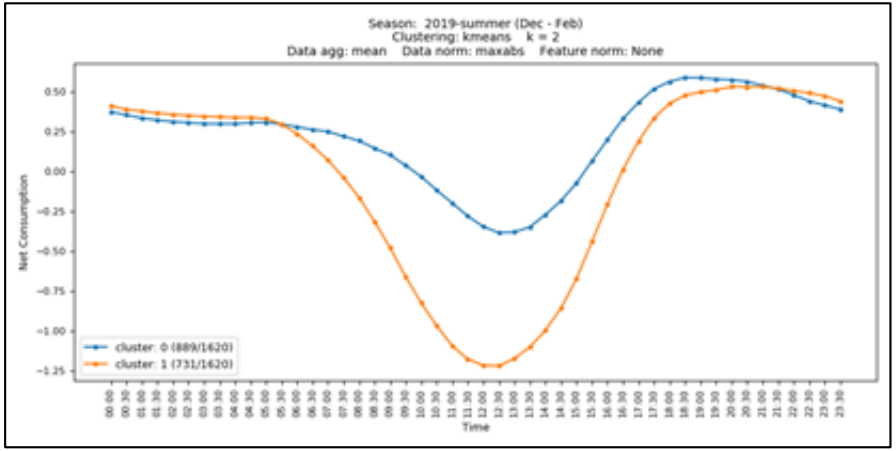
Analysis Component	Trialled Methodology
Data aggregation	Median, Mean
Consumption scaling	Max absolute, Min-Max
Time scaling	Max absolute
Time period	Summer, Autumn, Winter, Spring, Annual
Cluster domain	Time, Frequency
Cluster method	k-means, Spectral

Preliminary Results

Winter 2019



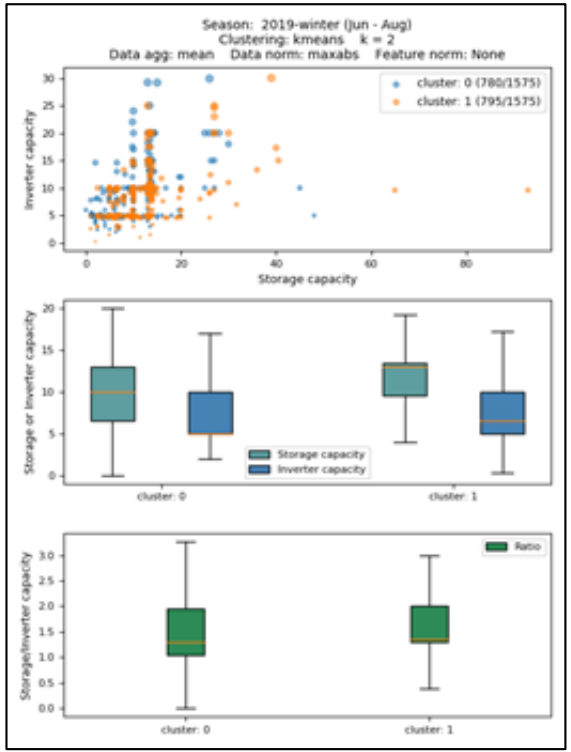
Summer 2019/20



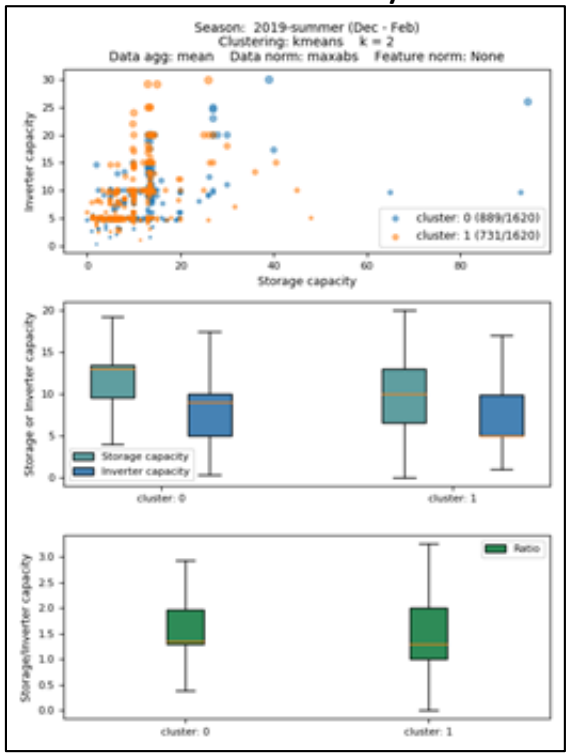
Note – These are preliminary results. These graphs are in progress images our researchers are working with. The results on these slides should only be used to guide discussions for future research, and not for sector decision making.

Preliminary Results

Winter 2019



Summer 2019/20



Note – These are preliminary results. These graphs are in progress images our researchers are working with. The results on these slides should only be used to guide discussions for future research, and not for sector decision making.

Going forward – Detecting control schemes

- Explore a few more approaches with clustering
  - Include white-box features
  - Include known battery properties in clustering
  - Assess business days vs non business days
  - Look at min/max demand days, high/low wholesale price days
  - Look at outlier households

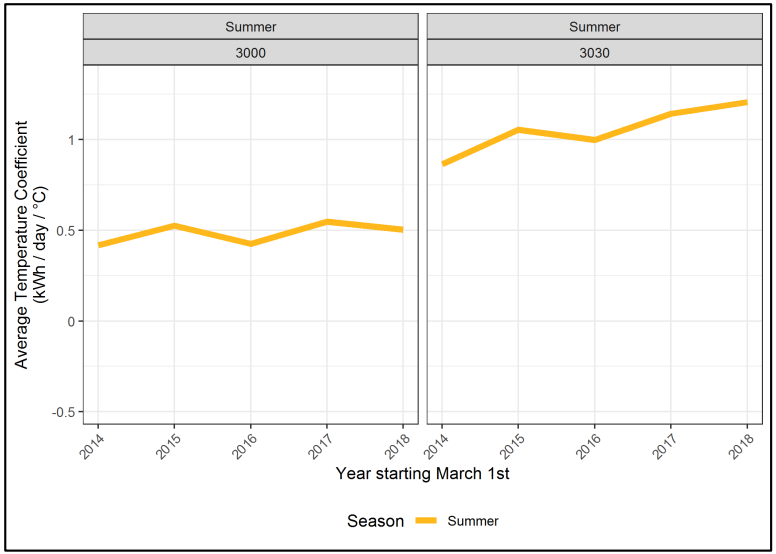
Going forward – Other research

- Determine households properties from data
  - Solar PV size
  - Battery Inverter size
  - Battery Storage size
- Research economical benefits of batteries
- We're also looking for suggestions!

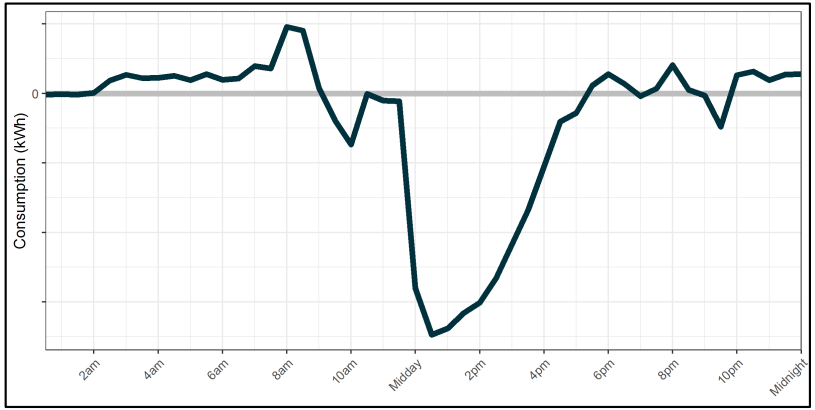
Summary

Two branches of work

Weather sensitivity



Battery tagging



- Questions?
- Comments?
- Suggestions?
- Ideas for future work?

Thank you

Energy
Lachlan O'Neil
Energy Data Scientist

+61 2 4960 6103
Lachlan.ONeil@csiro.au