MONASH University

MONASH BUSINESS SCHOOL

Forecasting the future of the power industry

What can you learn from smart meter data?

Rob J Hyndman

MONASH University

MONASH BUSINESS SCHOOL

Forecasting is the future of the power industry

What can you learn from smart meter data?

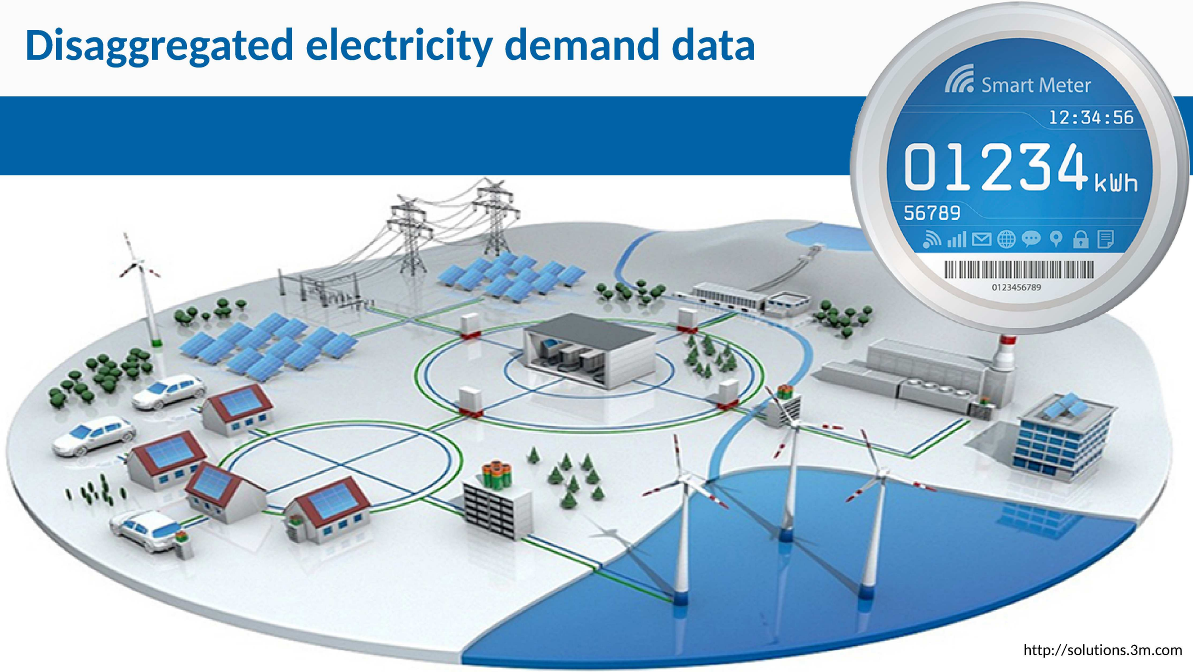
Rob J Hyndman

Outline

Disaggregated electricity demand data

http://solutions.3m.com

Disaggregated electricity demand data



http://solutions.3m.com

Smart metre data

vec.ausnetservices.com.au

AusNet services

HELP

Request your Smart Meter Data

NOTE: You will need your electricity bill handy to complete this form

Once you submit this form you will be sent an email with instructions on how to have your smart meter data redirected to the Victorian Energy Compare website where you can compare energy retail offers and claim your $50 Power Saving Bonus

First Name

Enter the First Name provided to your energy retailer

Last Name

Enter the Last Name provided to your energy retailer

NMI (National Meter Identifier)

630XXXXXXXX

Meter Number

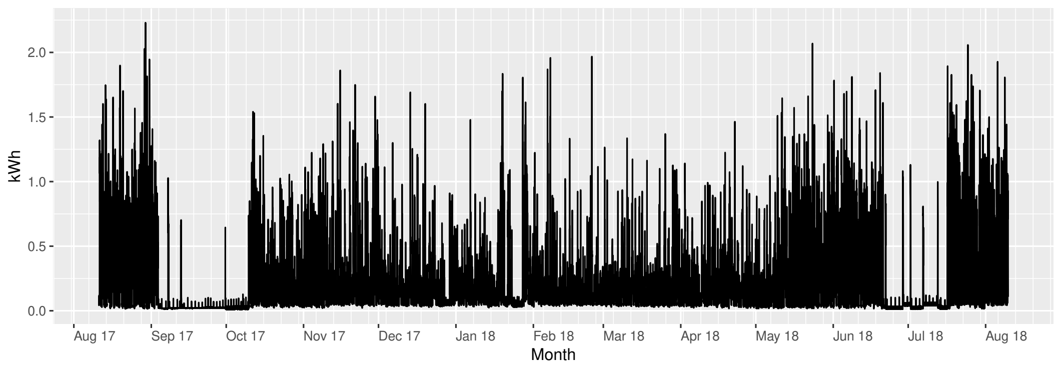
Phone Number

Postcode

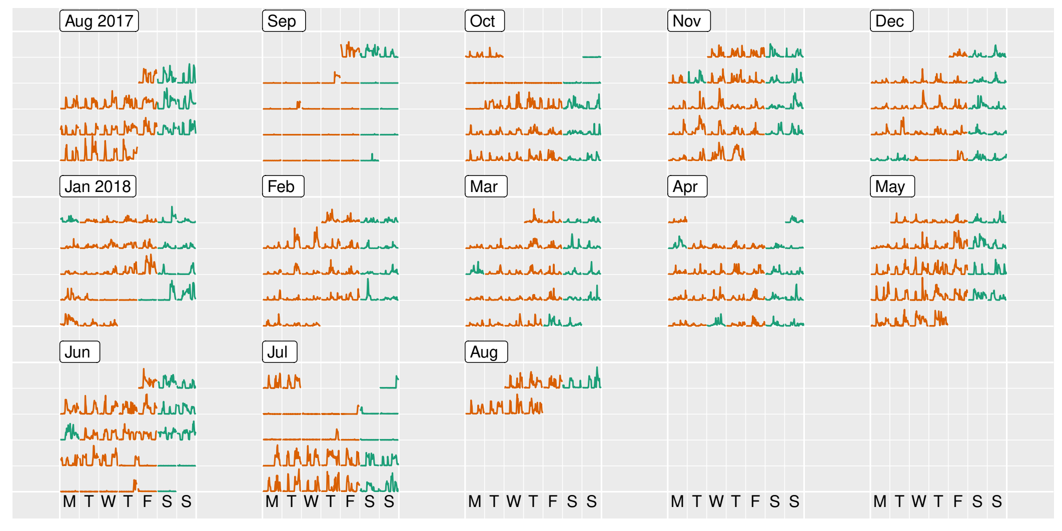
Email Address

Confirm Email

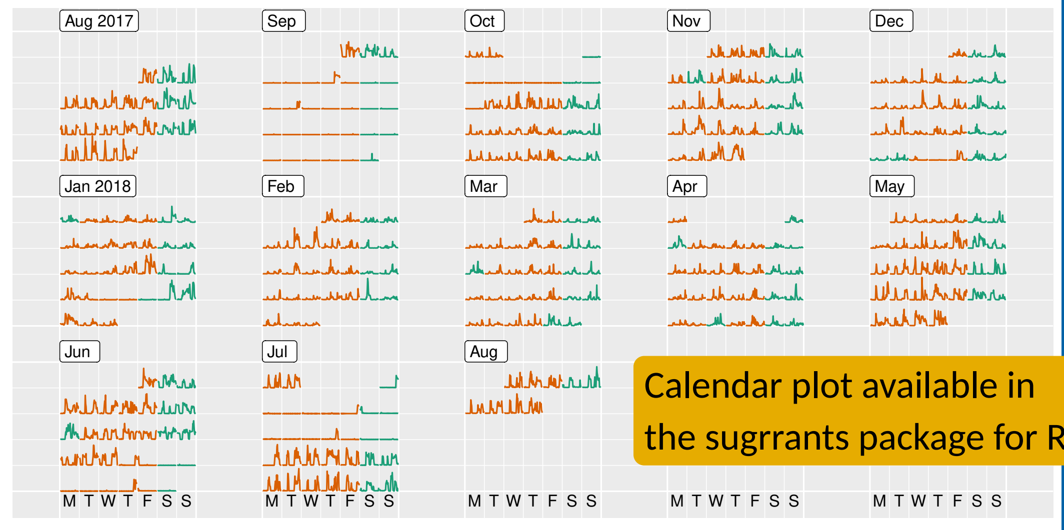
George’s data



George’s data

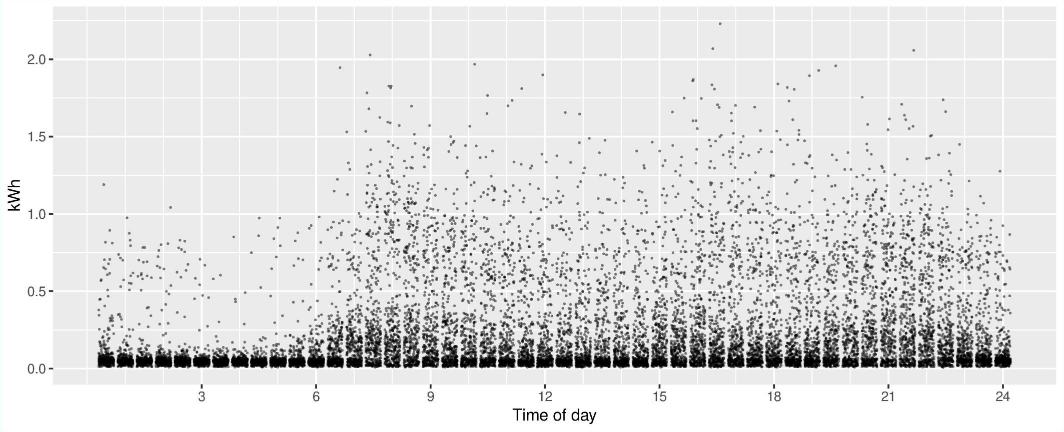


George’s data

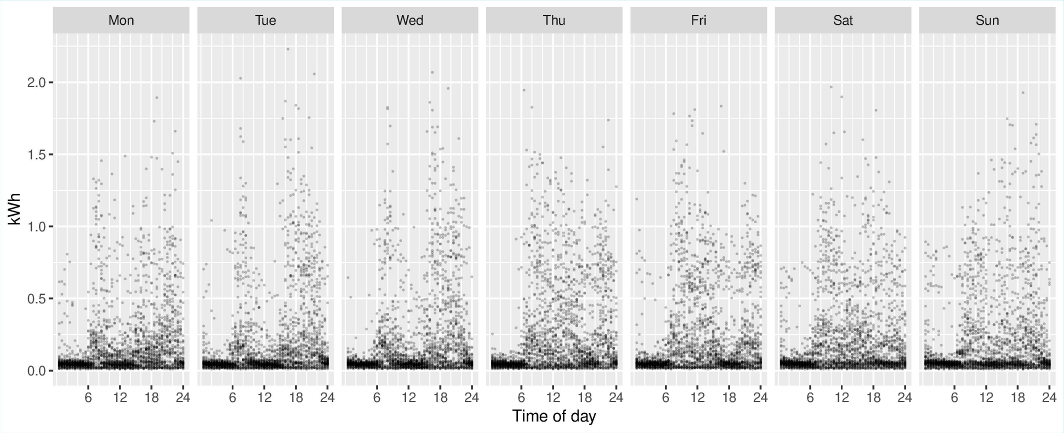


Calendar plot available in the sugrrants package for R.

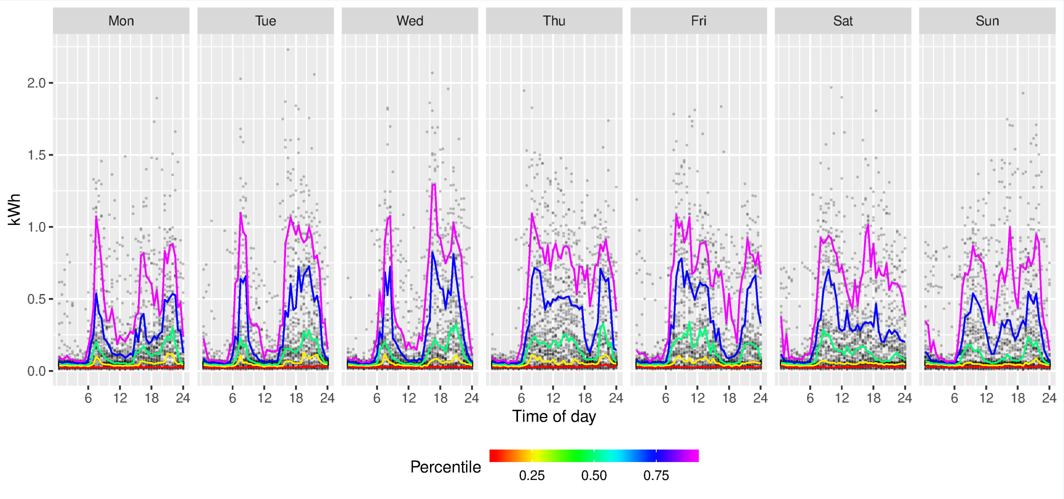
George’s data



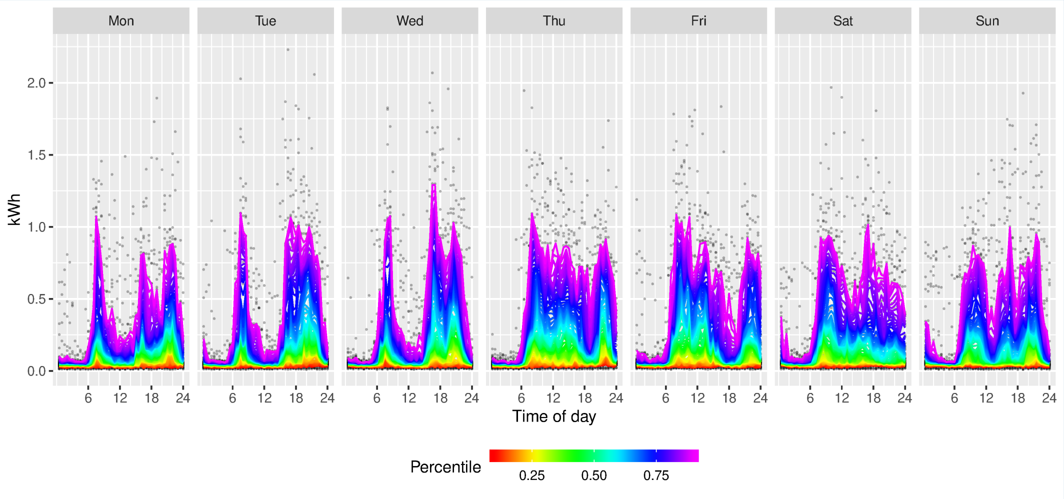
George’s data



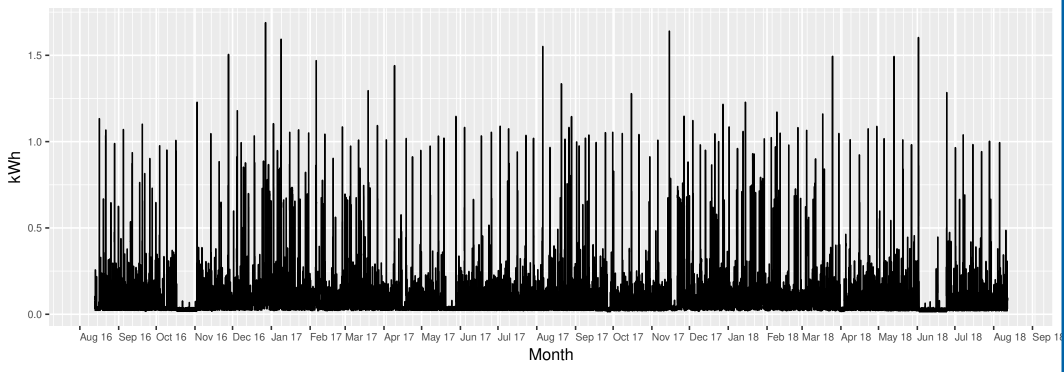
George’s data



George’s data

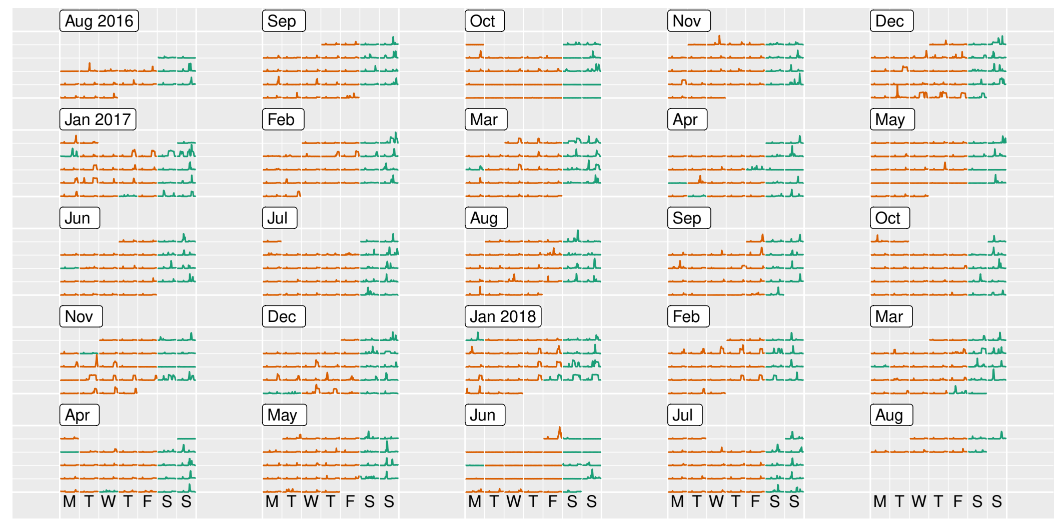


Clare’s data



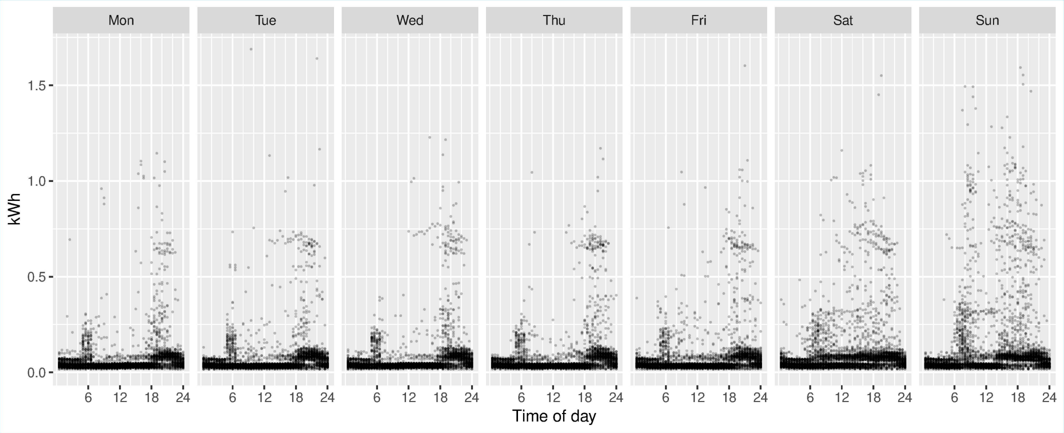
kWh
1.5
1.0
0.5
0.0
Aug 16
Sep 16
Oct 16
Nov 16
Dec 16
Jan 17
Feb 17
Mar 17
Apr 17
May 17
Jun 17
Jul 17
Aug 17
Sep 17
Oct 17
Nov 17
Dec 17
Jan 18
Feb 18
Mar 18
Apr 18
May 18
Jun 18
Jul 18
Aug 18
Month

Clare's data

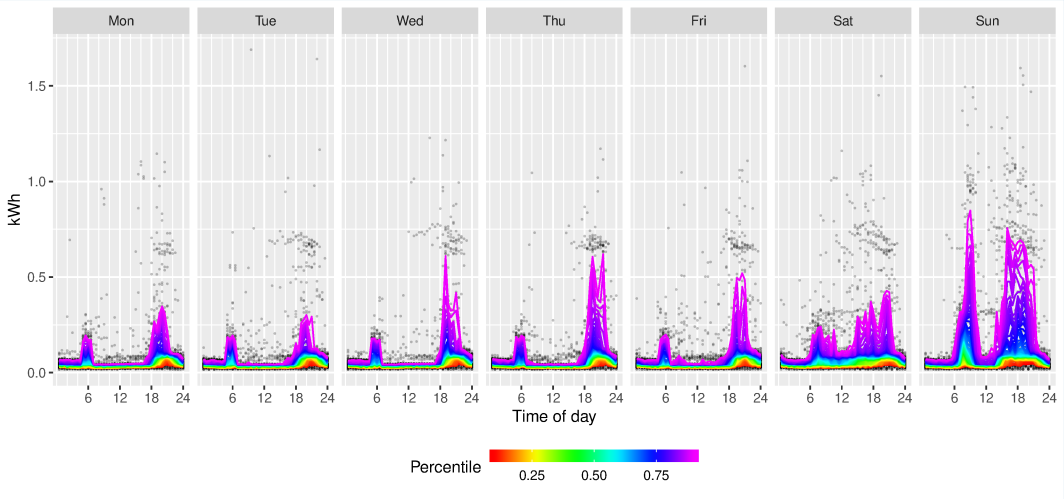


Aug 2016
Sep
Oct
Nov
Dec
Jan 2017
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan 2018
Feb
Mar
Apr
May
Jun
Jul
Aug
M T W T F S S
M T W T F S S
M T W T F S S
M T W T F S S
M T W T F S S

Clare’s data

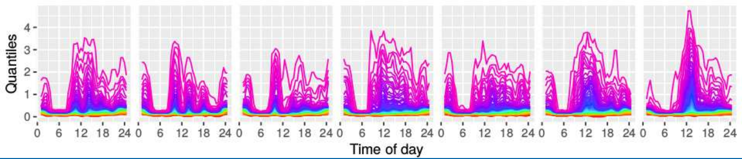


Clare's data



Mon
Tue
Wed
Thu
Fri
Sat
Sun
kWh
1.5
1.0
0.5
0.0
6
12
18
24
Time of day
Percentile
0.25
0.50
0.75

Percentiles conditional on time of week

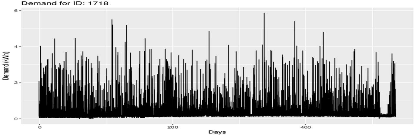


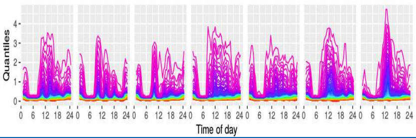
- Percentiles for each household and each half-hour of the week.
- Provides a unique fingerprint of typical usage for a given household.
- 336 probability distributions per household.
- Avoids missing data issues and variation in series length
- Avoids timing of household events, holidays, etc.
- Allows clustering of households based on probabilistic behaviour rather than coincident behaviour.
- A more complicated version also allows it to change across the year.

Finding anomalous smart metres

Irish smart metre data

- 500 households from smart metering trial:
  14 July 2009 – 31 December 2010.
- Electricity consumption at 30-minute intervals.
- Heating/cooling energy usage excluded.

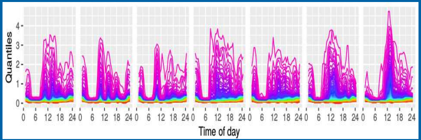




The time series of 535 $\times$ 48 observations per household is mapped to a set of 7 $\times$ 48 $\times$ 99 percentiles giving a bivariate surface for each household.

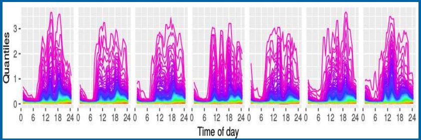
Finding anomalous smart metres

Can we compute pairwise distances between all households?



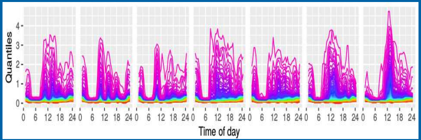
← ? →

Distance

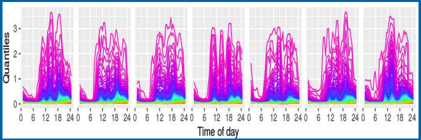


Finding anomalous smart metres

Can we compute pairwise distances between all households?



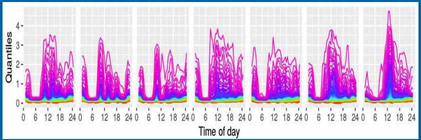
← ? →
Distance



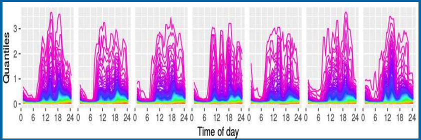
- Jensen-Shannon measure gives distance between two densities

Finding anomalous smart metres

Can we compute pairwise distances between all households?



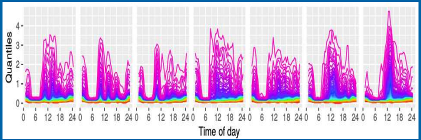
$\leftarrow$? $\rightarrow$
Distance



- Jensen-Shannon measure gives distance between two densities
- Distance between household *i* and household *j*:
  Δ_{ij} = sum of 7×48 JS distances.

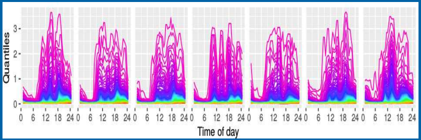
Finding anomalous smart metres

Can we compute pairwise distances between all households?



$\leftarrow$? $\rightarrow$

Distance

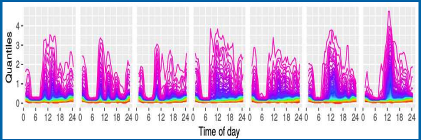


- Jensen-Shannon measure gives distance between two densities
- Distance between household *i* and household *j*:

  Δ_{ij} = sum of 7×48 JS distances.
- Similarity between two households: $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$

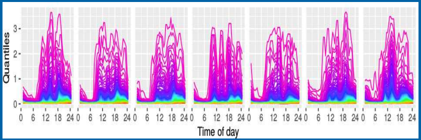
Finding anomalous smart metres

Can we compute pairwise distances between all households?



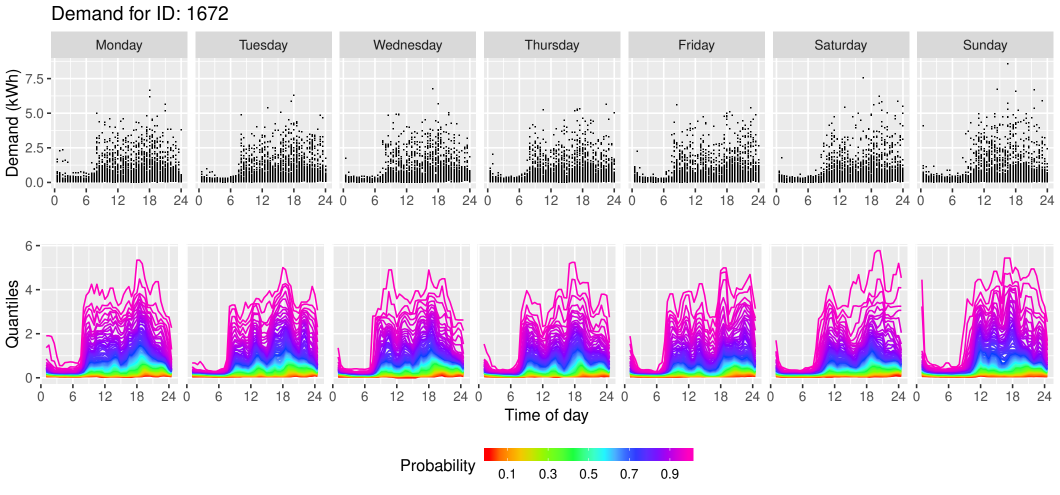
$\leftarrow$? $\rightarrow$

Distance

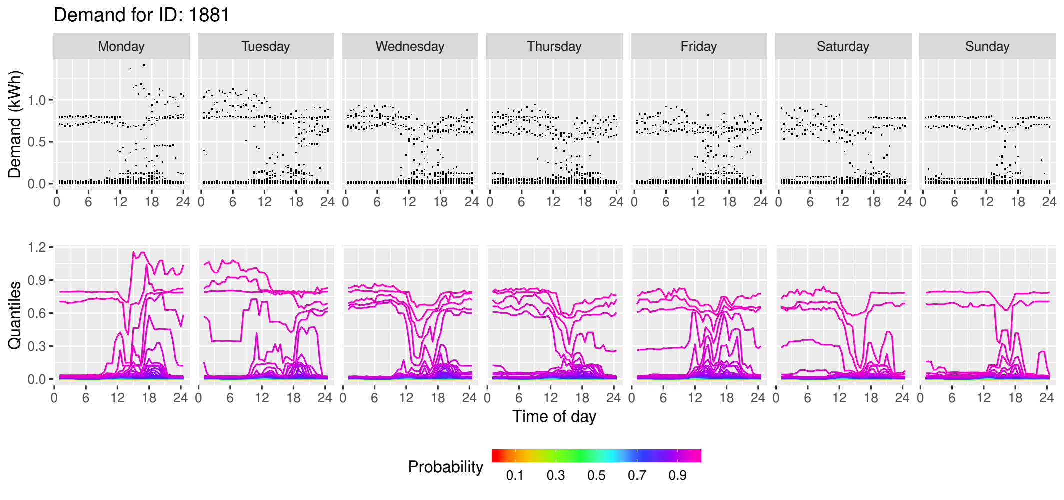


- Jensen-Shannon measure gives distance between two densities
- Distance between household *i* and household *j*:
  Δ_{ij} = sum of 7×48 JS distances.
- Similarity between two households: $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$
- Household typicality: $f_i = \sum_j w_{ij}$.

Most typical household



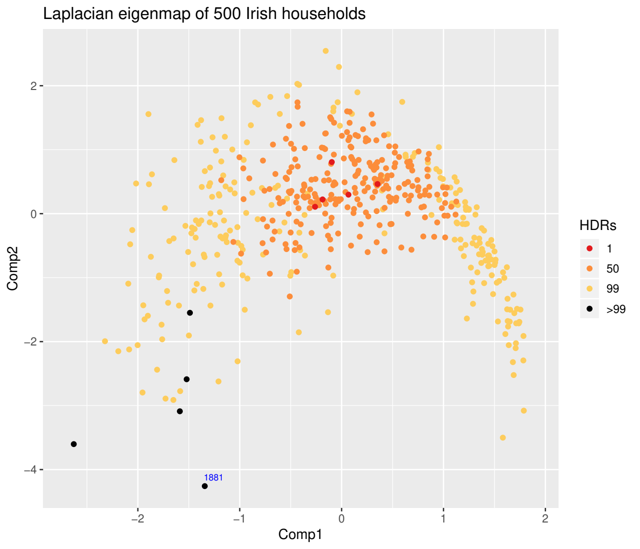
Most anomalous household



Visualization via embedding

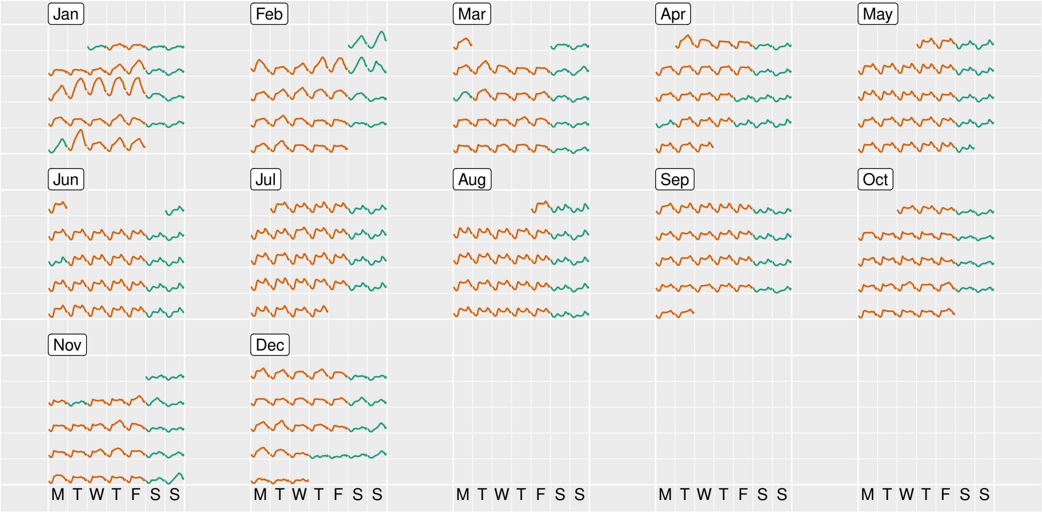
Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.

- With more data, we might be able to find clusters of similar households.
- Anomalous households may be due to:
  - malfunctioning equipment
  - unusual schedules or behaviours
  - nefarious activity

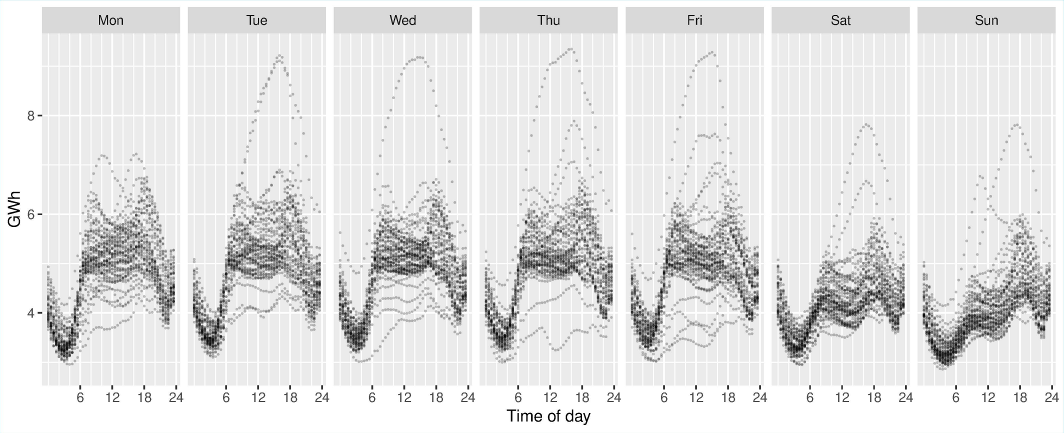


Outline

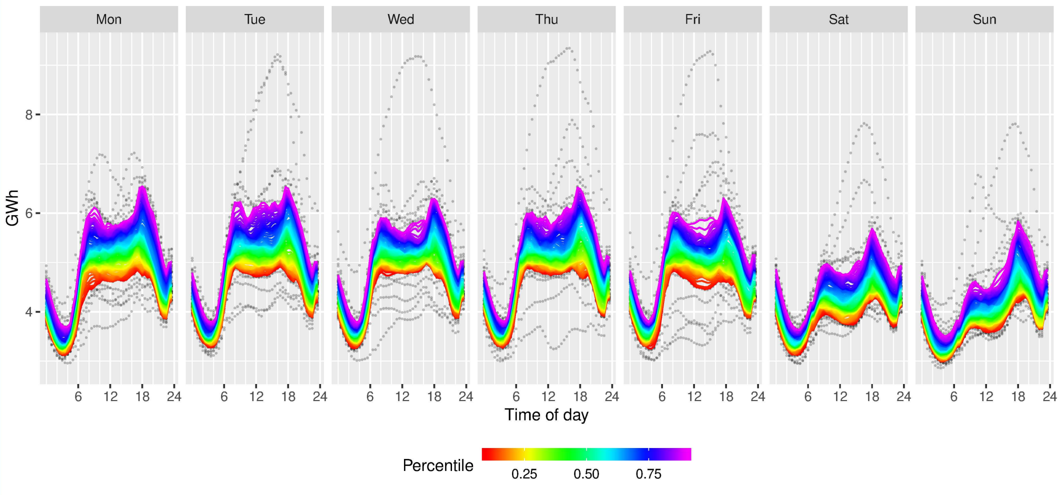
South Australian network demand 2014



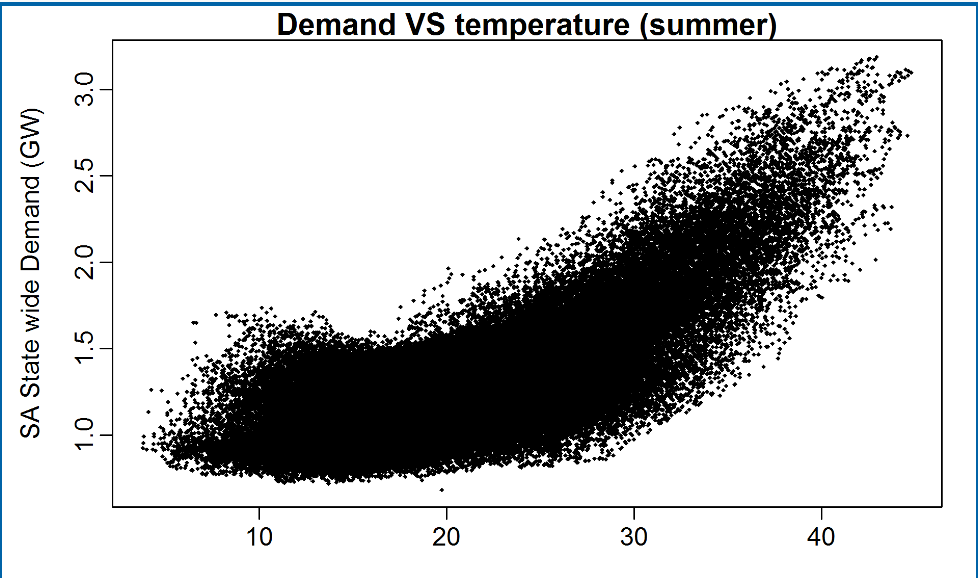
South Australian network demand 2014



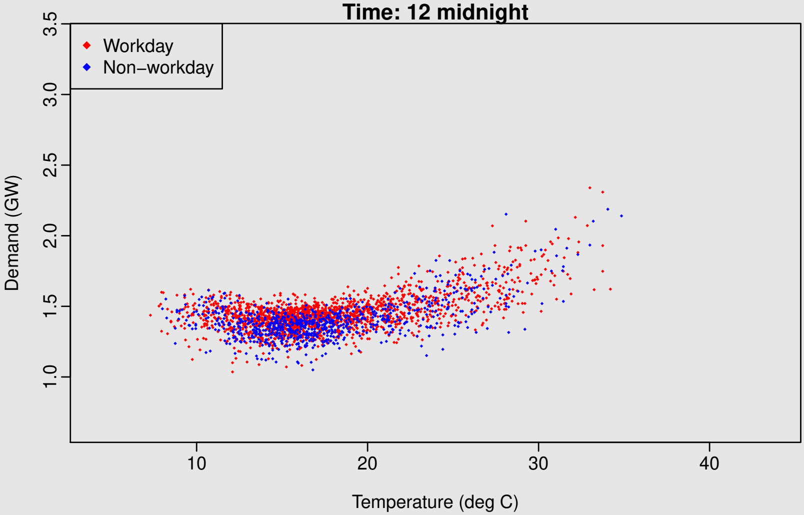
South Australian network demand 2014



South Australian demand data



Temperature data (Sth Aust)



Predictors (inputs)

- Calendar: Time-of-day, day-of-week, time-of-year, holidays
- Prevailing and recent weather conditions (up to a week ago)
- Price, economic measures, population
- Recent demand
- Separate models for each half-hour period, each season, and working/non-working days

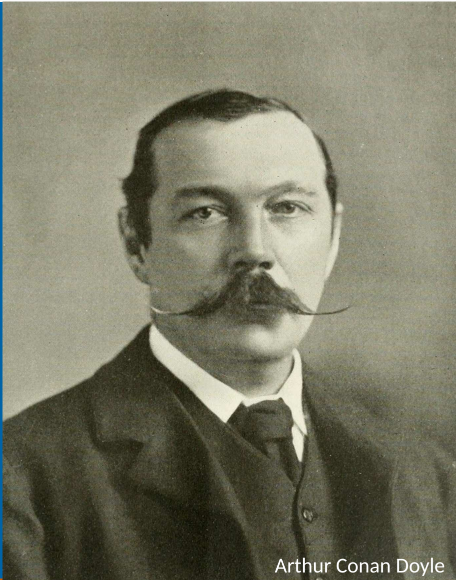
- The same predictors will work for smart metres, but with much larger percentage errors than for networks.
- The best network forecasts in the future will use network data and smart metre data.
- Energy forecasting is inherently probabilistic.

Network demand vs smart-metre demand

You can . . . never foretell what any one man will do, but you can say with precision what an average number will be up to.

— *Arthur Conan Doyle (via Sherlock Holmes)*

- Network demand is more regular and predictable than smart-metre demand.
- Smart-metre demand provides a lot more detail about individual behaviour.
- Reconciled forecasts allow us to combine network forecasting with smart-metre forecasting.



Arthur Conan Doyle

Measures of forecast accuracy

- **Point forecasts:** Mean Square Error
  Mean Absolute Percentage Error
- **Probabilistic forecasts:** Quantile Scores
- We want probability forecasts to be *calibrated* and *sharp*.

Calibration

The percentage of points above each predicted percentile is correct. e.g., 10% of points are above the 90th predicted percentile.

Sharpness

The predicted probability distribution is concentrated around the centre, with short tails. Prediction intervals are narrow.

Evaluating probabilistic forecasts

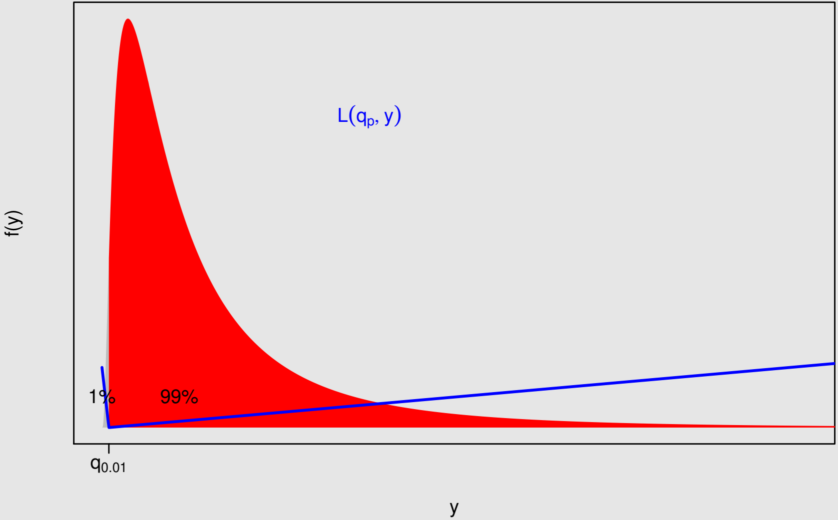
Let q_p be the percentile forecast with probability $1 - p$ of exceedance.

Pin-ball loss function

$$L(q_p, y) = \begin{cases} (1-p)(q_p - y) & \text{if } y < q_p; \\ p(y - q_p) & \text{if } y \geq q_p. \end{cases}$$

- average over all target percentiles (e.g., 1%, 2%, $\ldots$, 99%), all forecast horizons, and all observed data.
- Takes account of how far percentiles are exceeded.

Evaluating percentile forecasts



Outline

1 Smart metre data analytics

2 Network demand forecasting

3 Global Energy Forecasting Competitions

4 The Victorian smart metre data

Global Energy Forecasting Competitions

- A series of competitions to discover what methods work best in practice.
- Organized by Professor Tao Hong (University of North Carolina) in collaboration with the *International Journal of Forecasting*.

- **GEFCom2012:** load and wind power
- **GEFCom2014:** load, price, wind, solar (probabilistic)
- **GEFCom2017:** hierarchical load (probabilistic)

GEFCom 2012

2012 Global Energy Forecasting Competition

- Forecasting hourly loads for a US utility across 20 zones.
  - Up to 1-week-ahead forecasts.
  - 104 participants
- Forecasting hourly wind power generation at 7 wind farms.
  - Up to 48 hours ahead.
  - 133 participants
- Results published in *International Journal of Forecasting* 2014.
- Data, results and all code publicly available.

GEFCom 2012

2012 Global Energy Forecasting Competition

Load Forecasting Track

1. Singleton and Charlton *(Counting Lab UK)*
2. Lloyd *(University of Cambridge, UK)*
3. Nedellec, Cugliari, Goude *(EDF R&D, INRIA, France)*
4. Ben Taieb and Hyndman *(Monash Uni, Australia)*

Wind Power Forecasting Track

1. da Silva *(DTI Sistemas Brazil)*
2. Mangalova and Agafonov *(Siberian State Aerospace University and Siberian Federal University, Russia)*
3. Nagy *(Budapset University of Technology and Economics, Hungary)*
4. Lee *(University of Texas at Austin, USA)*

GEFCom 2012

2012 Global Energy Forecasting Competition

Load Forecasting Track

1. Singleton and Charlton *(Counting Lab UK)*
2. Lloyd *(University of Cambridge, UK)*
3. Nedellec, Cugliari, Goude *(EDF R&D, INRIA, France)*
4. Ben Taieb and Hyndman *(Monash Uni, Australia)*

Wind Power Forecasting Track

1. da Silva *(DTI Sistemas Brazil)*
2. Mangalova and Agafonov *(Siberian State Aerospace University and Siberian Federal University, Russia)*
3. Nagy *(Budapset University of Technology and Economics, Hungary)*
4. Lee *(University of Texas at Austin, USA)*

What worked?

- Semi-parametric statistical models
- Gaussian process regression
- Gradient boosting
- Feature engineering

GEFCom 2014

2014 Global Energy Forecasting Competition

1. Load forecasting
2. Price forecasting
3. Wind forecasting
4. Solar forecasting

- Rolling forecasts every week in real time
- 99 percentile forecasts submitted
- Evaluation using probability scoring
- Winning methods published in the IJF.
- Data, results and all code publicly available.

GEFCom 2014

2014 Global Energy Forecasting Competition

Load

1. EDF, France
2. EDF, France

Price

1. EDF, France
2. Wrocław University of Technology, Poland

Wind

1. Eigen Analytics, USA
2. Budapest Uni of Technology & Economics, Hungary

Solar

1. Huang,Perry, Australia
2. Budapest Uni of Technology & Economics, Hungary

GEFCom 2014

2014 Global Energy Forecasting Competition

Load

1. EDF, France
2. EDF, France

Price

1. EDF, France
2. Wrocław University of Technology, Poland

Wind

1. Eigen Analytics, USA
2. Budapest Uni of Technology & Economics, Hungary

Solar

1. Huang,Perry, Australia
2. Budapest Uni of Technology & Economics, Hungary

What worked?

- Semi-parametric regression
- Bootstrapping
- Gradient boosting
- kNN regression
- Quantile regression averaging
- Generalized additive tree ensemble

GEFCom 2017

2017 Global Energy Forecasting Competition

Hierarchical load probabilistic forecasts

- Rolling forecasts every month in real time
- Zone and total loads of ISO New England
- 99 percentile forecasts submitted
- Evaluation using probability scoring
- Winning methods to be published in the IJF.
- Data, results and all code publicly available.

- No contestant used reconciliation to align zone and total loads.
- Final results not yet published.

Forecasts about forecasting (Hong et al, IJF 2016)

bit.ly/gefcom2014

The next 10 years

1. Solar power forecasting research will flourish
2. Development of practical error measures for probabilistic forecasting
3. A connection between probabilistic and point energy forecasting
4. Increased use of high res data, temporally, spatially and conceptually
5. The unification of energy forecasting methodologies
6. A diversification of energy forecasting subjects
7. The fusion of energy forecasting problems
8. Interdisciplinary collaborations with other communities
9. Additional energy forecasting competitions
10. Regular conferences in energy forecasting
11. A dedicated publication outlet for energy forecasters
12. A society for energy forecasters

Outline

1. Smart metre data analytics
2. Network demand forecasting
3. Global Energy Forecasting Competitions
4. The Victorian smart metre data

The Victorian smart metre data

The Victorian smart metre data

How might it be used for analysis?

- Identify anomalous behaviour.
- Cluster households to target demand response initiatives.
- Improve network and sub-network demand forecasting through highly disaggregated data.
- Improve network and sub-network demand forecasting by directly measuring solar generation.
- Estimate hourly occupancy rates for urban planning.
- Estimate daily occupancy rates for epidemiological study.
- Better measure television audiences for major events.

Thanks

Di Cook

Kate Smith-Miles

Anastasios Panagiotelis

Souhaib Ben Taieb

George Athanasopoulos

Nick Tierney

Earo Wang

Dilini Talagala

Mitchell O'Hara-Wild

Sevvandi Kandanarachchi

Cameron Roach

Puwasala Gamakumara

Thiyanga Talagala

Sayani Gupta

Further information:
robjhyndman.com