Predicting residential occupancy and hot water usage from high frequency, multi-vector utilities data

Strata Data Conference NY
13 Sep 2018

Agenda

- Introductions

- Future Energy Systems

- The Challenge

- Our Analytical Approach

- Solution to Data Compression and Data Feature Extraction

- Results

- Key Findings and Learnings

Introductions
Introducing the client, the consultancies and myself

Baringa

A £400m industry and UK government partnership in low carbon energy. Our role is to act as a conduit between academia, industry and the government to accelerate the development of low carbon technologies.

ASI Data Science believe AI is for everyone. ASI empower organisations to become more data-driven by providing first class software, skills, and advisory solutions

Baringa Partners is an independent business and technology consultancy, we help businesses run more effectively, navigate industry shifts and reach new markets. We use our industry insights, ideas and pragmatism to help clients improve their business

Cristóbal Lowery
Senior Manager

cris.lowery@baringa.com

Cris is one of Baringa's leaders on modelling and machine learning, running the centre of excellence of two dozen specialists. He also holds a double first from Imperial College London in AI and Mechanical Engineering.

Future Energy System Challenges

The global energy industry is profoundly changing and technology will be a key enabler in reducing and shifting energy consumption, reducing the costs of the **$16 Trillion** transformation and achieving a lower carbon future.

The Need

Global energy shortfall and electrification of the energy system. Economic and population growth particularly in emerging countries drive significant power demand growth
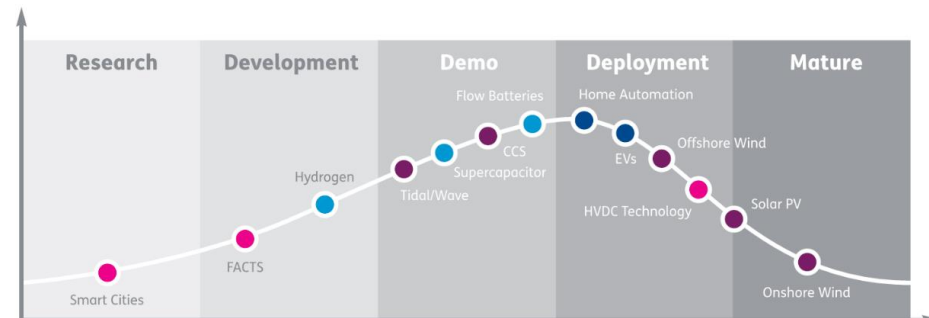
The Commitment

A low carbon future. Paris Agreement ratified commitments to hold global warming to no more than 2C above pre-industrial levels.
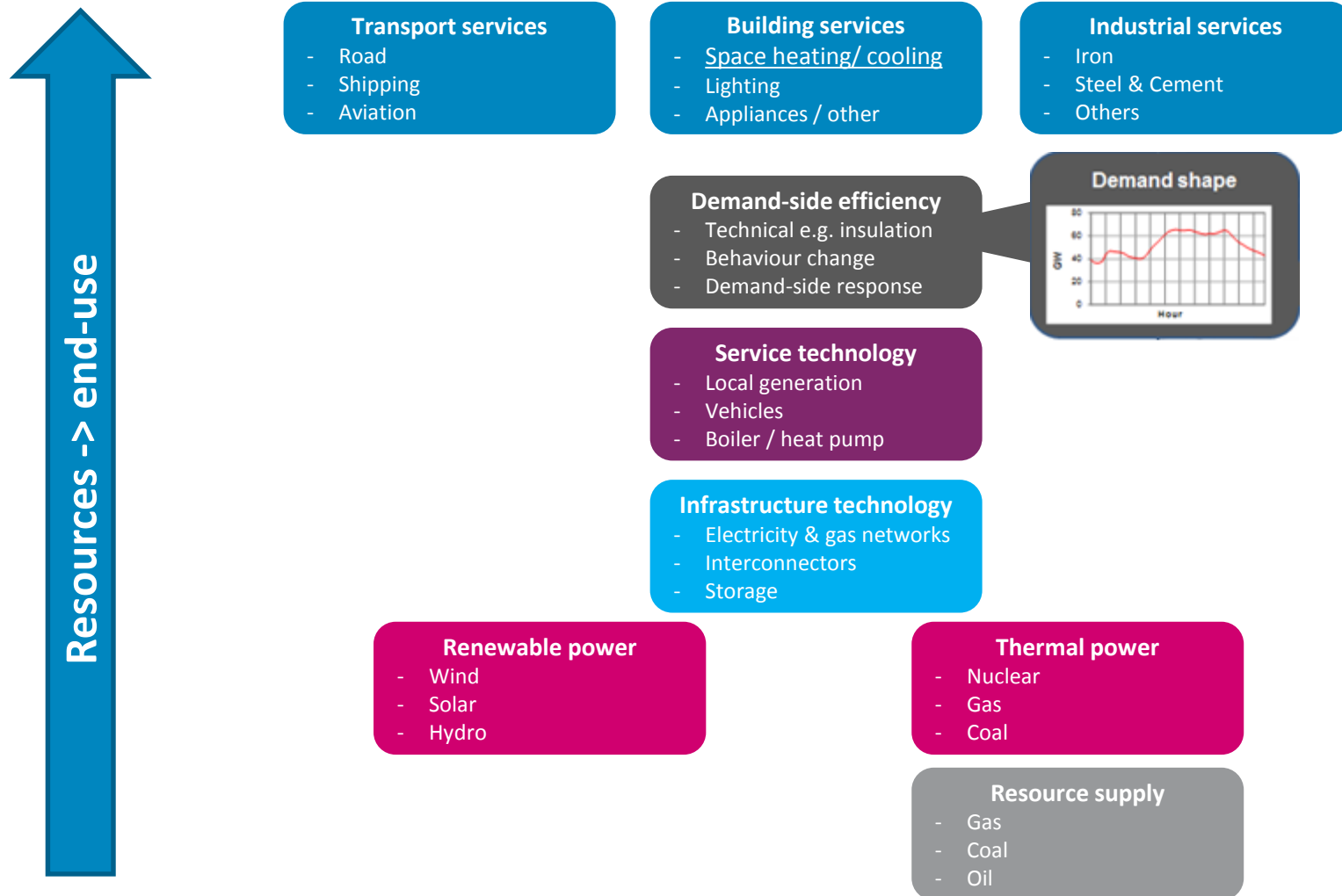
The Enablers

Technology advances, regulatory chance and economics.

Weakest | Weak | Strong | Strongest

Research | Development | Demo | Deployment | Mature

Smart Cities, FACTS, Hydrogen, Tidal/Wave, Supercapacitor, CCS, Flow Batteries, Home Automation, EVs, HVDC Technology, Offshore Wind, Solar PV, Onshore Wind

Future Energy System Challenges

Key technological opportunities will lie in energy storage, demand-side response, active control of appliance amongst others. Here we investigate residential heating of spaces and water

Resources -> end-use

Transport services
- Road
- Shipping
- Aviation

Building services
- Space heating/ cooling
- Lighting
- Appliances / other

Industrial services
- Iron
- Steel & Cement
- Others

Demand-side efficiency
- Technical e.g. insulation
- Behaviour change
- Demand-side response

Demand shape

Service technology
- Local generation
- Vehicles
- Boiler / heat pump

Infrastructure technology
- Electricity & gas networks
- Interconnectors
- Storage

Renewable power
- Wind
- Solar
- Hydro

Thermal power
- Nuclear
- Gas
- Coal

Resource supply
- Gas
- Coal
- Oil

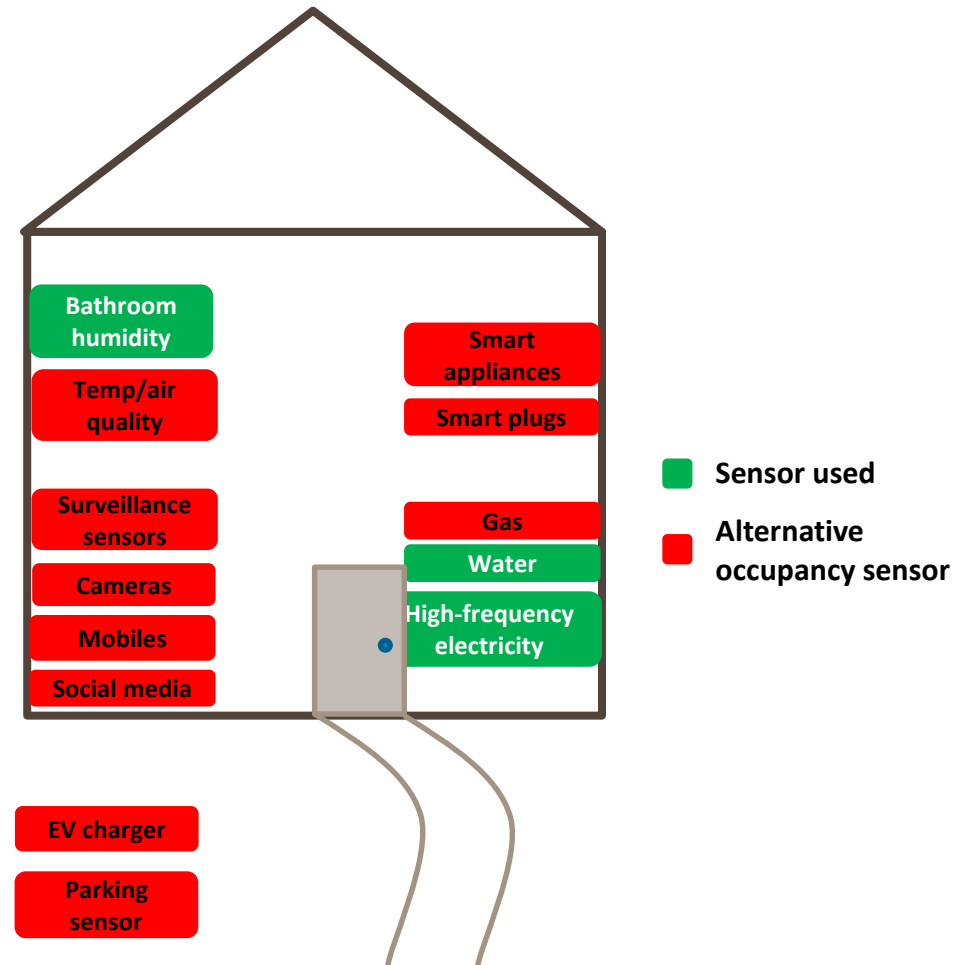# Energy Technologies Institute's Challenge

A key energy service is household water and space heating, accounting for 80% of household energy consumption in Europe. Can we predict residential heating needs using machine learning?

Key requirements:

- Predict residential heating needs, moving from pull to push mode
- Non-intrusive technologies: multi-vector whole house consumption data & high frequency electricity data
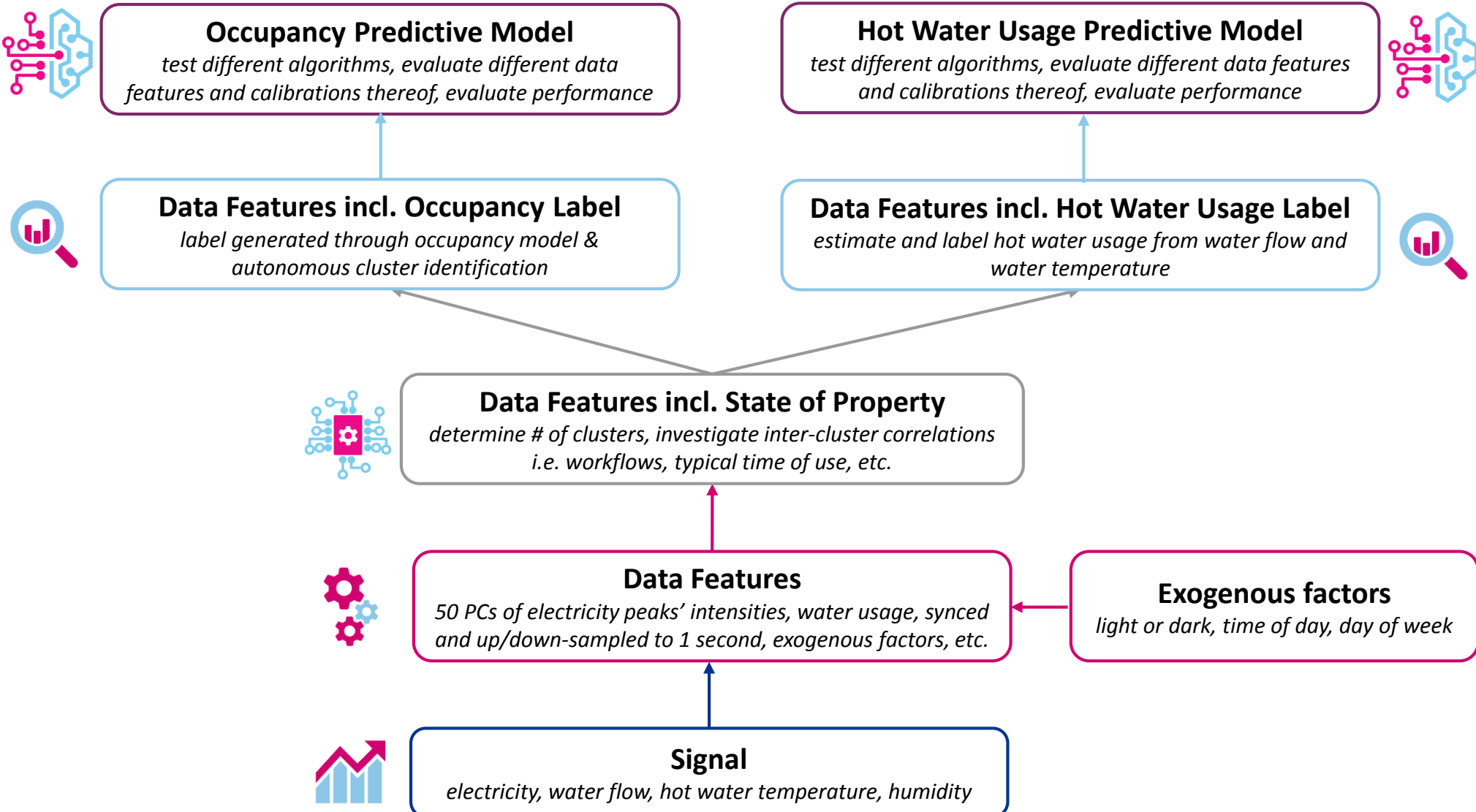- 10 min to 72 hour time frame & uses priors

Key differences:

- Not appliance disaggregation, which has high error rates and different predictive purpose i.e. predictive maintenance or detect intruder
- Not intrusive and avoids using smart plugs
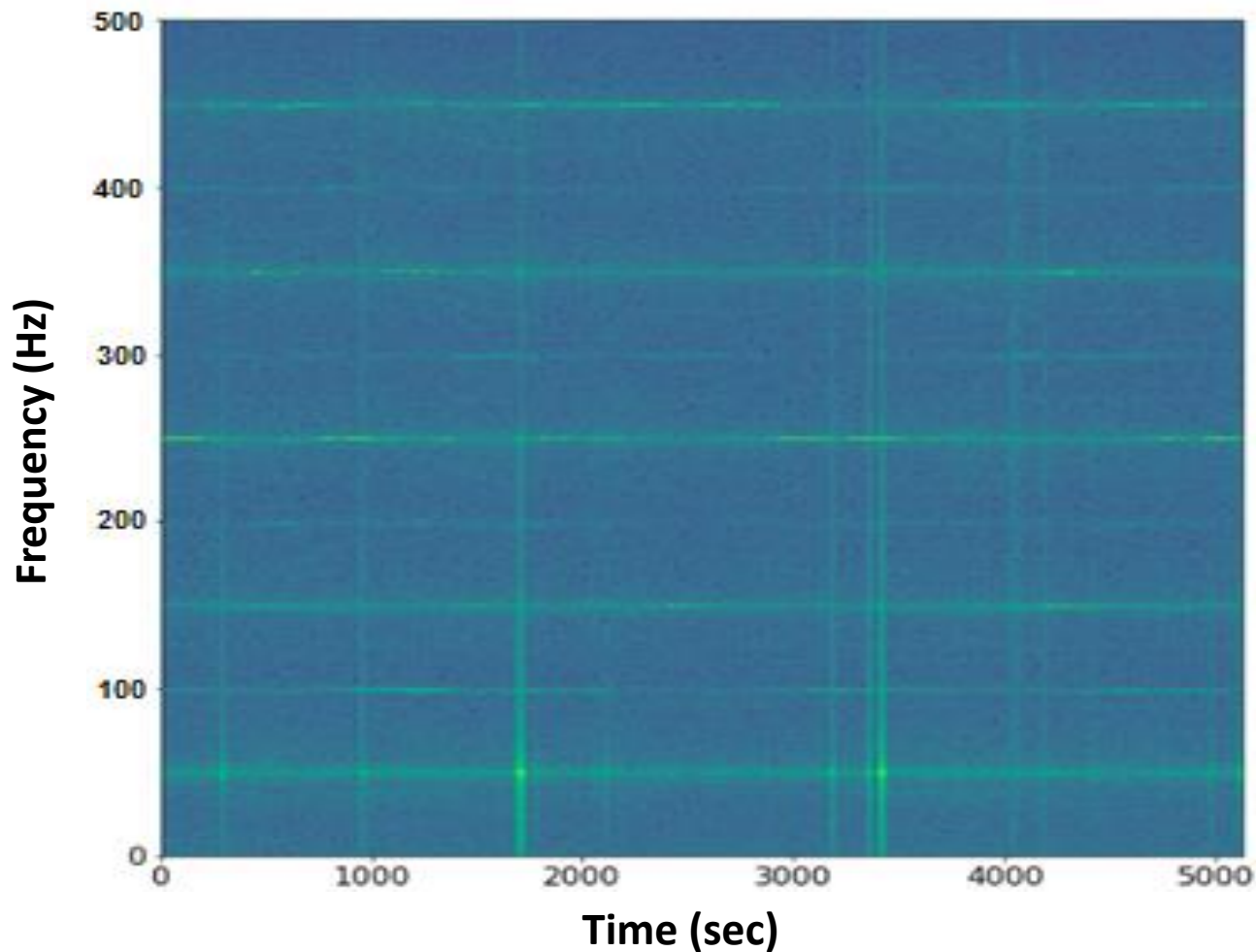- Not programmable home energy management system

Bathroom humidity

Temp/air quality

Surveillance sensors

Cameras

Mobiles

Social media

Smart appliances

Smart plugs

Gas

Water

High-frequency electricity

EV charger

Parking sensor

Sensor used

Alternative occupancy sensor

Analytical Framework

Key steps include feature extraction, labelling target variable and predicting

Occupancy Predictive Model
test different algorithms, evaluate different data features and calibrations thereof, evaluate performance

Hot Water Usage Predictive Model
test different algorithms, evaluate different data features and calibrations thereof, evaluate performance

Data Features incl. Occupancy Label
label generated through occupancy model & autonomous cluster identification

Data Features incl. Hot Water Usage Label
estimate and label hot water usage from water flow and water temperature

Data Features incl. State of Property
determine # of clusters, investigate inter-cluster correlations i.e. workflows, typical time of use, etc.

Data Features
50 PCs of electricity peaks' intensities, water usage, synced and up/down-sampled to 1 second, exogenous factors, etc.

Exogenous factors
light or dark, time of day, day of week

Signal
electricity, water flow, hot water temperature, humidity

Data Compression

Each house generated 600,000 electricity data points a second, but most of the signal appeared to be in the 50 Hz harmonics
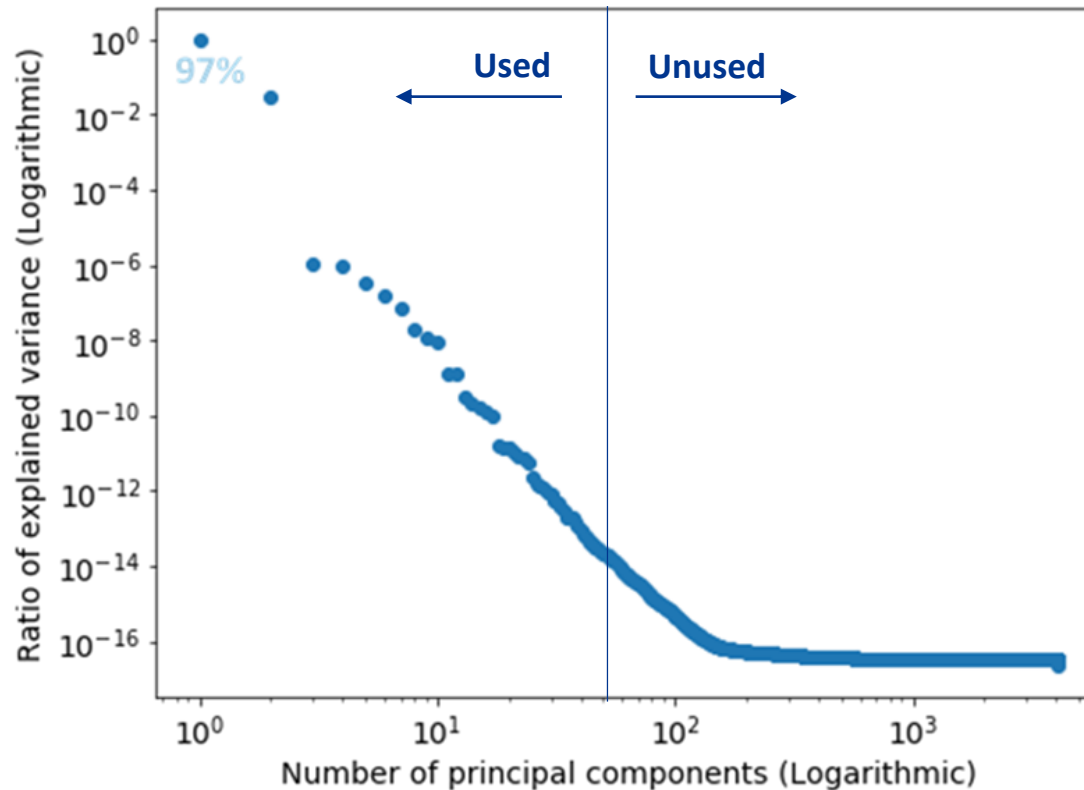
Data Compression

Using a 1 second stride and a 2 second window, Fourier Transforms were applied reducing the dimensionality by 75x

Data Compression

Further 160x compression obtained by applying Principal Component Analysis and keeping top 50 components. Total compression is 12,000x
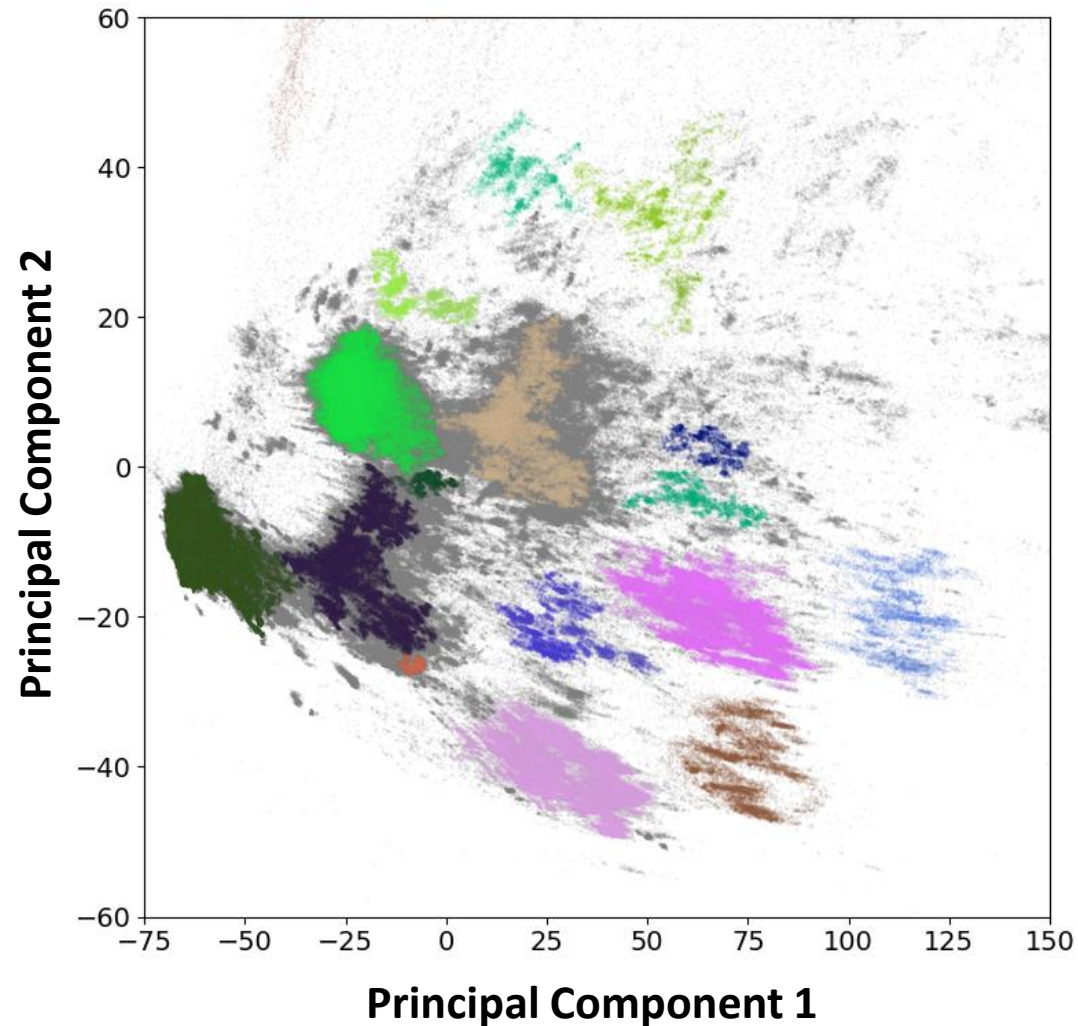

Principal Component 1: Active power @ 50Hz
Principal Component 2: Reactive power @ 50Hz
Principal Component 3: combination of active and reactive power @150Hz and @250Hz

Data Features

HDBSCAN found property to be in one of 25 electrical states 89% of the time, 3 of the states are autonomous

Data Features

Other types of feature engineering that derived high predictive performance are shared below

▲ Using historical and current values of the target variable proved very powerful. To introduce these in a live environment occupancy will need to be labelled on the go, which may present some technical challenges, or should be directly sensed.

▲ Exogenous factors proved powerful and could act as priors, especially when there are data quality issues:
  – Time of day;
  – Weekday;
  – Non-working day;
  – Light/dark based on sunrise and sunset;
  – Mealtime.

▲ Memory proved to be key and is expressed in a few different forms:
  – The consecutive usage, time of usage or absence of a variable
  – Changes to the state over a period of time i.e. increment in humidity;
  – Consumption over a period i.e. hot water usage over a period of 24 hours;
  – Exponentially weight moving average (EWMA) to include water consumption or electricity usage over a time range with a stronger focus on the more recent consumption.

Hot Water Usage Results

Random Forest used for both hot water usage and occupancy predictions

▲ Data features that are consistently powerful are:
  – Hot water usage (auto-regressive feature)
  – Water usage
  – Bathroom humidity
  – Memory of various data features (with lifetime of c. 24 hours or comparable to predictive time horizon)

▲ Electricity principal components have some influence, depending largely on the time horizon

▲ Electricity clusters performed worse than the principal components and had minimal influence

▲ Exogenous factors had minimal influence

Time horizon	Predictive performance [AUC]
10 mins	92%
1 hr	70%
4 hrs	62%
24 hrs	61%
72 hrs	46%

Occupancy Results
Performance appears better than for hot water usage

▲ Key predictive features are:
- Historical and current occupancy (more so for shorter time horizons)
- Bathroom humidity
- Electricity principal components with memory
- Electricity clusters do not perform very well
- Exogenous factors performed well, unlike for the hot water usage model

▲ Performance generally higher for shorter time horizons. The 72 hour model outperforms the 24 hour model, suggesting data size issues.

▲ All results a fair bit above 50% AUC

Time horizon	Predictive performance [AUC]
10 mins	98%
1 hr	89%
4 hrs	78%
24 hrs	63%
72 hrs	69%

Summary of key findings

Heating systems could be improved with this technology and similar use cases are likely to arise in the near future i.e. smart electric vehicle charging

Concept

- A less explored objective of meter consumption analytics is predicting occupancy and hot water use

- This use case could drive more cost effective energy consumption in both absolute terms and through time of use

- Need for such systems will increase as heating decarbonises and makes price swings more volatile

Analytical results

- It is possible to predict hot water usage and occupancy to a good standard using high frequency, multi-vector utility data. More specifically, our initial findings suggest:
    - It is not necessary to disaggregate appliances
    - High frequency electricity data increases predictive power
    - None electrical data (humidity, water use, etc.) have additional predictive information
    - Priors are key predictors of occupancy i.e. time of day
- Given data volumes of high frequency data, compression techniques will be key and Fourier transforms are a logical option in a production environment