WPD INNOVATION
Transforming the electricity network

centre for sustainable energy

CATAPULT Energy Systems

Presumed Open Data: Data Science Challenge

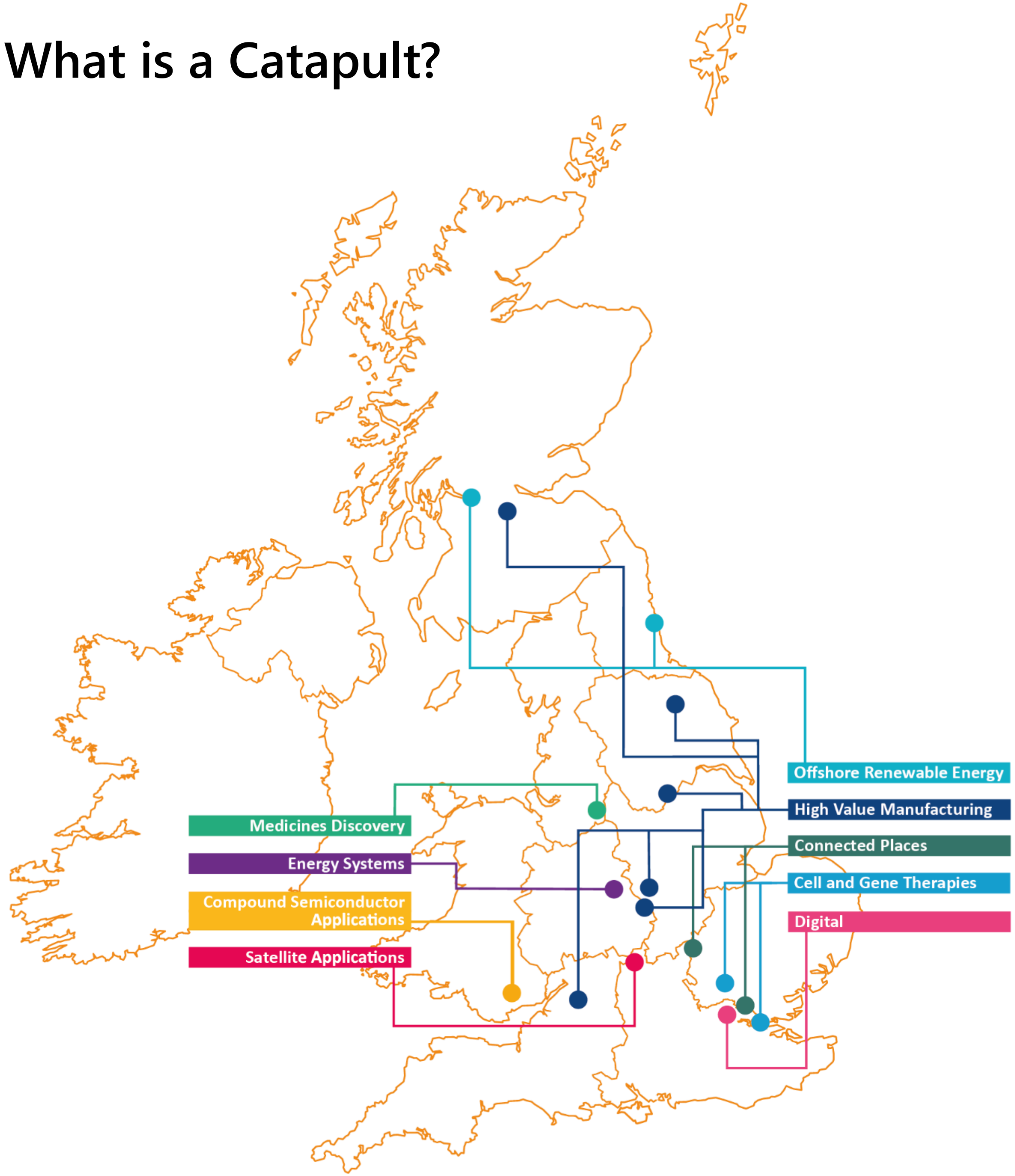
Dr Stephen Haben

Data Scientist

Thursday 28th January 2021

@EnergySysCat   LinkedIn

What is a Catapult?

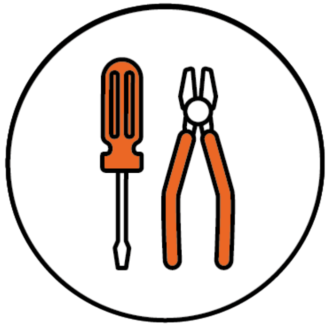


WPD INNOVATION
Transforming the electricity network

centre for sustainable energy

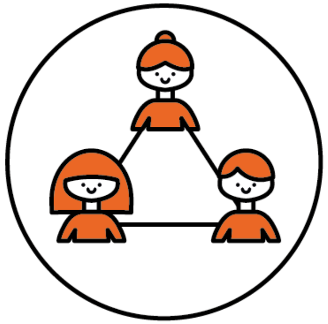
CATAPULT
Energy Systems

Established and overseen by Innovate UK



Technical capabilities, equipment, and other resources

Solve key problems and develop new products and services



Bridge the gap between stakeholders in the sector

Open up opportunities for innovators, in the UK and globally

About Energy Systems Catapult

Mission: Unleash innovation and open new markets to capture the clean growth opportunity

200 Innovation experts

Hubs in Birmingham and Derby

We work with Innovate UK

Established, overseen and part-funded by Innovate UK. Independent from Government. Not for profit

Bridge the gap between stakeholders in the sector

Supporting innovators

Research

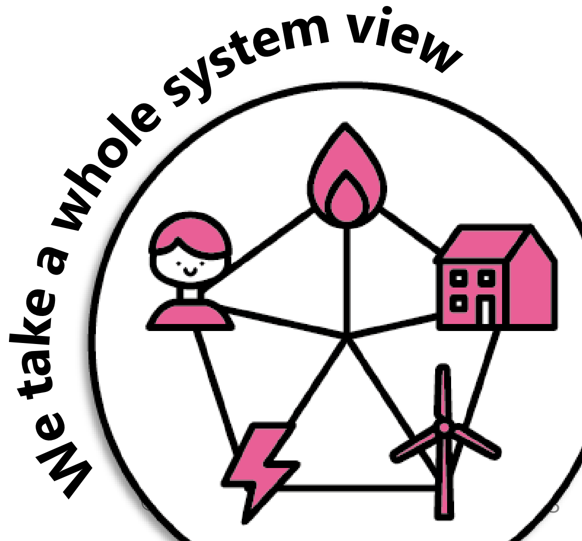
Trials

Systems engineering

01010
10101
01010

Digital

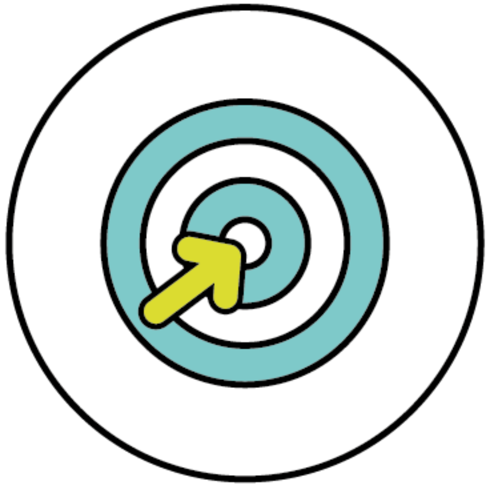
Modelling and simulation



Contents

Overall aims of the challenge

- Focus on application.
- Demonstrate Value in increasing data accessibility:
  - What techniques can improve performance?
  - What datasets demonstrate the most value?
  - What value can be engineered from the data?
  - How much data is needed?
  - What other datasets may be required?



- What it is not: not a demonstration of perfect realism or a ready to implement solution.

Challenge Motivation

Increased penetrations of distributed generation (e.g. solar PV) and low carbon technologies such as heat pumps, electric vehicles etc. will increase uncertainty and levels of demand on the distribution network.

"Smart grids could reduce additional reinforcement costs by £2.5B-£12B by 2050[1]."

Batteries are one viable solution which could avoid expensive network reinforcement.

[1] Delivering UK Energy Investment: Networks, DECC, January 2015

Challenge Scenario

- A 6MWh/2.5MW battery is connected to a primary distribution substation and a 5MW solar farm in Devon, southwest England.

- Design the control of a storage device to support the distribution network to:
  - Maximise the daily evening peak reduction.
  - Using as much solar photovoltaic energy as possible.

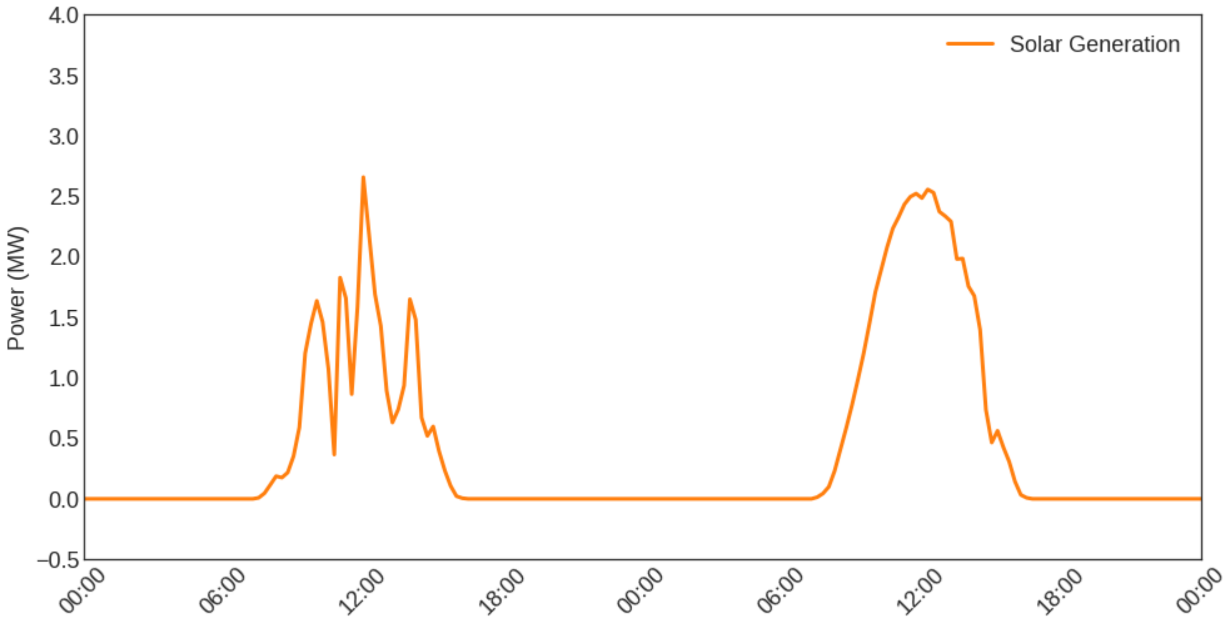
- This will be done for each day for the week following the current challenge date.

- In other words it is ***a constrained optimisation/control problem under uncertainty.***

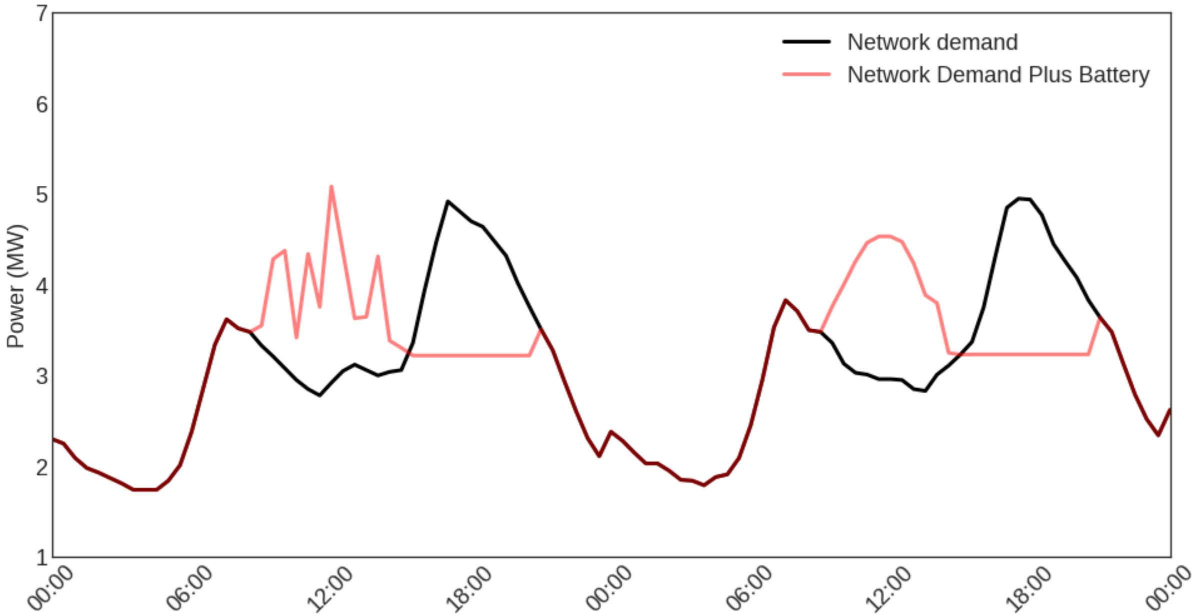
- There will be **four** assessed weeks as part of this challenge.

Example: 2 days perfect foreknowledge

Solar Generation



Demand with and without Storage



Data sets

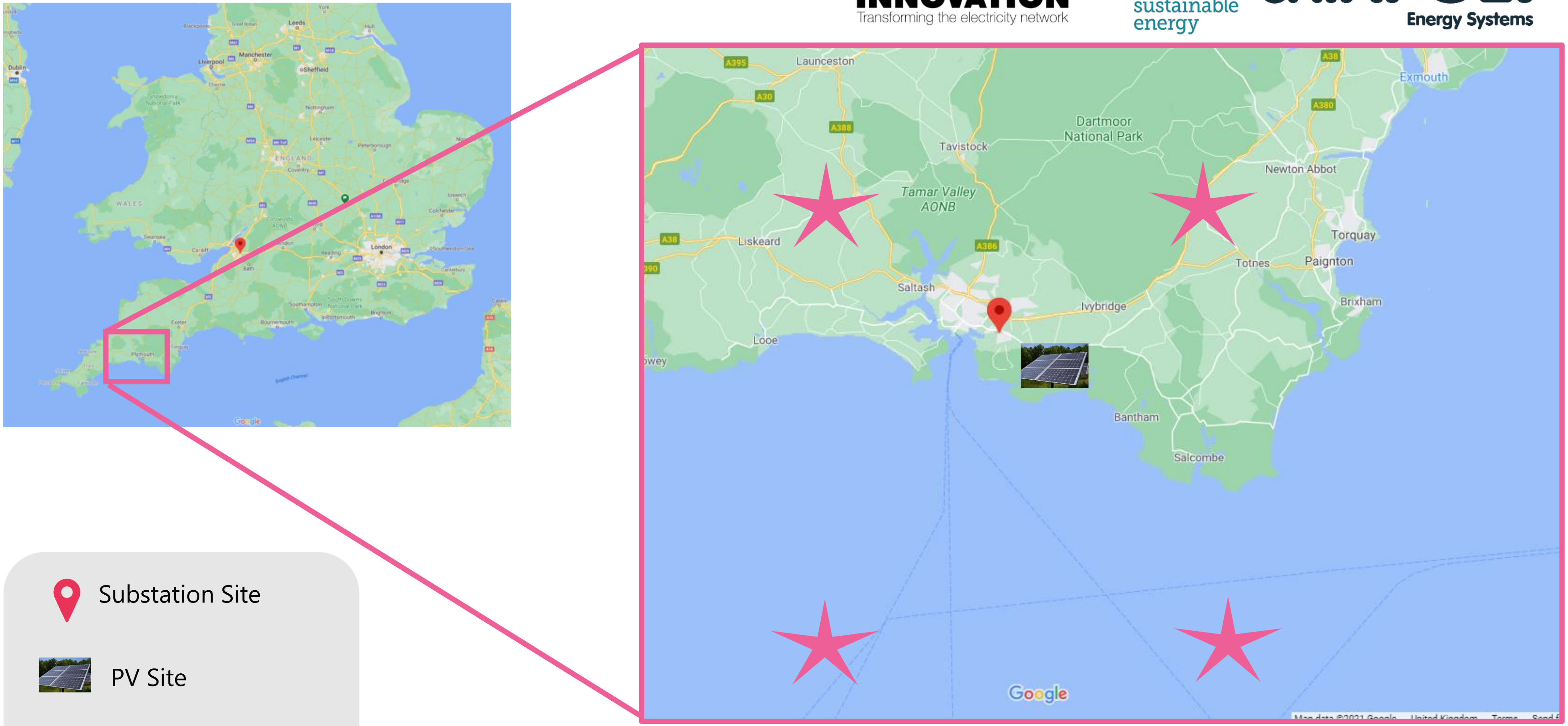
Three core data sets for each part of this task.

For each task there will be:

1. Half hourly demand data up to just before the start of the task week.
2. Data for solar site near substation up to just before the start of the task week.
3. Weather reanalysis data up to the end of the task week.

N.B. All data timestamps are Coordinated Universal Time (UTC).

The Location



Substation Site

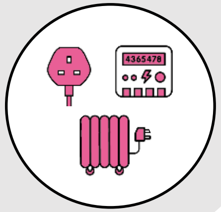
PV Site

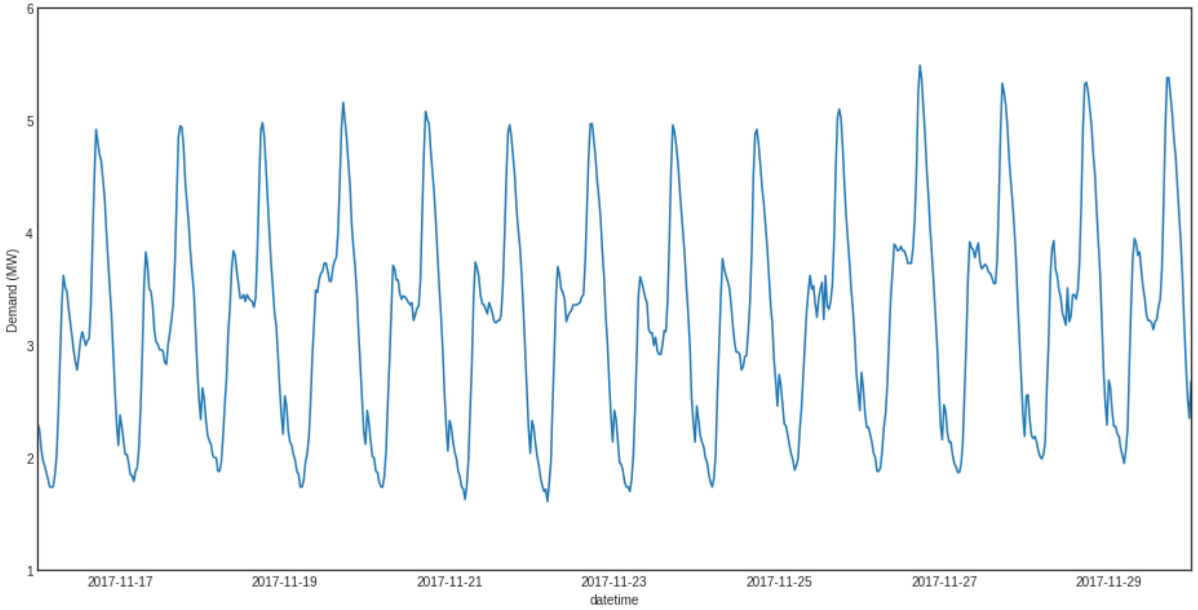
Weather Grid point

Demand Data

Stentaway Primary distribution substation:

- Half hourly average demand data (MW)
- Data starts 3rd November 2017
- Note that demand behaviour may follow GMT/BST but the timestamps are in UTC.
- Main peaks occur in the evening: 4pm to 8pm period.
- Relatively clean though some:
  - Large values
  - Constant values
  - Few missing values.





Photovoltaic data

5MW solar farm near Stentaway:

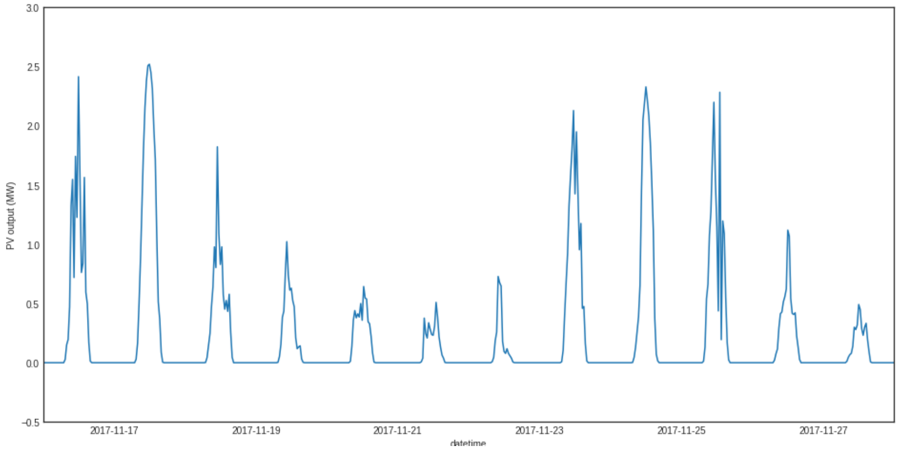
Half hourly data for:

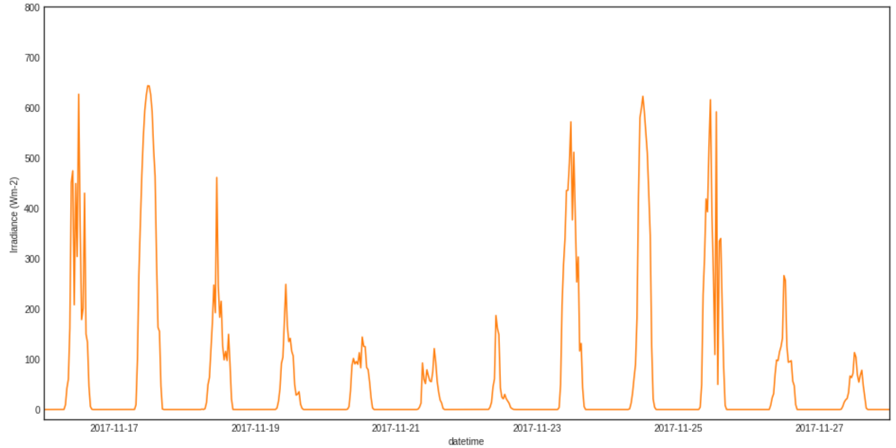
Average solar output (MW)
Average irradiance (Wm^{-2})
Panel Temperature (Celsius)

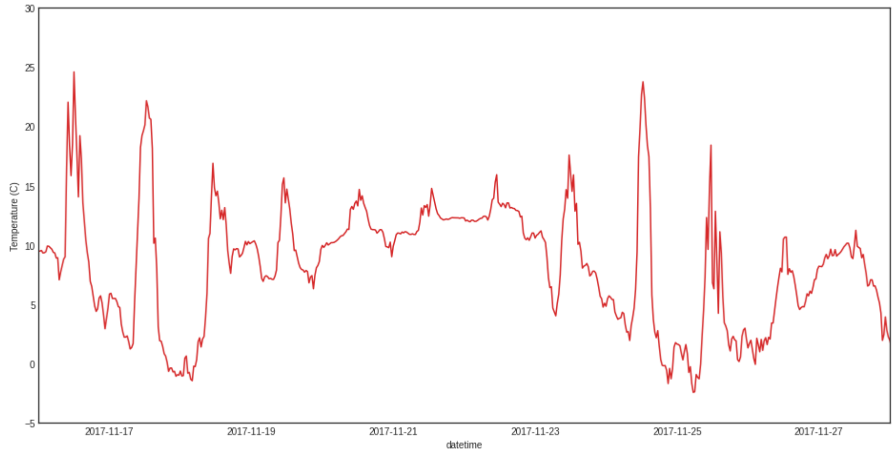
Note panel temperature will typically be hotter than ambient temperature during hot spells.

Starts 3rd November 2017.

Relatively clean but there are a few missing values







Weather Data

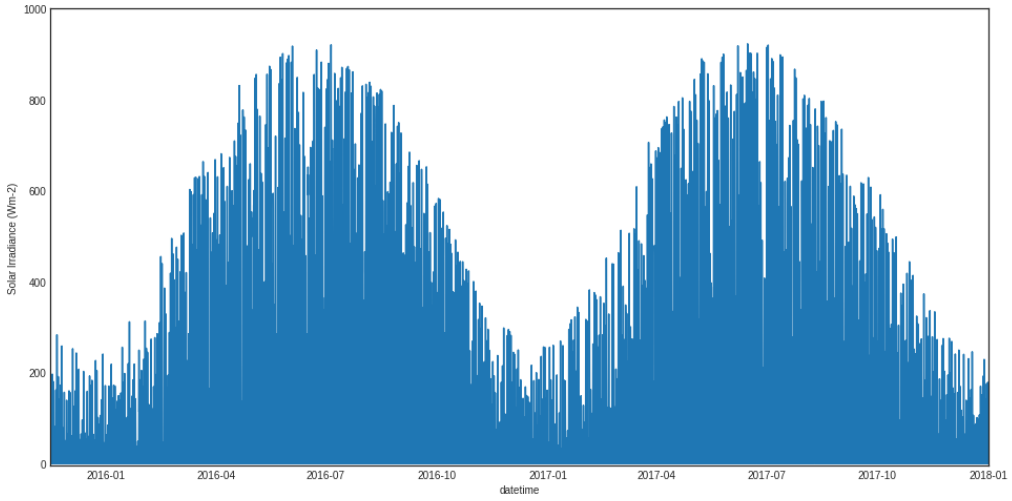
WPD INNOVATION
Transforming the electricity network

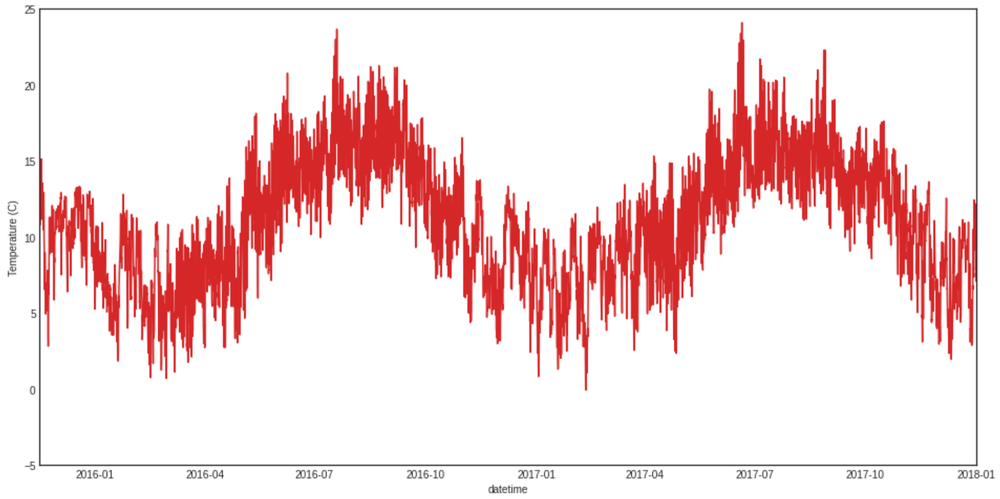
centre for sustainable energy

CATAPULT
Energy Systems

Data from **NASA's MERRA-2** (Modern-Era Retrospective analysis for Research and Applications, Version 2):

- Hourly *reanalysis* data – i.e. reassimilated forecasts using current observations.
  - Irradiance (Wm^{-2})
  - Temperature (Celsius)
- Six sites including:
  - The four surrounding grid points of substation/PV site
- Data set with longest historical data: starting 1st January 2015.





Challenge in detail

For each day of the task week, the aim is to reduce the evening demand peak over the period **3.30PM to 9PM** by discharging a battery storage device at the most optimal times.

This should be done charging the storage with as much solar energy as generated from the **before 3.30PM** on that day.

More precisely, if $\boldsymbol{L}_d = (L_{d,1}, L_{d,2}, \dots, L_{d,48})^T$ be the actual future substation demand over day d of the test week ($d = 1, \dots, 7$).

Here $L_{d,k}$ is the average power (in MW) over the k^{th} half hour of day d. So $k = 1$ would mean the period from midnight to 00:30 AM on the current day, d.

The challenge therefore requires determining a charging schedule for the battery for each day given by

$$\boldsymbol{B}_d = (B_{d,1}, B_{d,2}, \dots, B_{d,48})^T$$

where $B_{d,k}$ is the average power (in MW) over the k^{th} half hour of day d, to minimise the peak demand over the evening period (the half hours $k = 32 \ to \ 42$) given by

$$\min_{\boldsymbol{B}_d \in \mathcal{B}_d} \left\{ \max_{k \in \{32,\dots,42\}} (L_{d,k} + B_{d,k}) \right\}$$

Where $\mathcal{B}_d$ are all feasible charging/discharging profiles for the battery storage device.

Battery Constraints

The optimal battery schedule must also satisfy some constraints (These define the feasible set of profiles $\mathcal{B}_d$).

The first constraint is on the maximum import and export of energy. In this case

$$-2.5MW = B_{min} \leq B_{d,k} \leq B_{max} = 2.5MW$$

Secondly the battery cannot charge beyond its capacity, $C_{d,k}$, (in MWh):

$$0 \leq C_{d,k} \leq C_{max} = 6MWh$$

The total charge in the battery at the next time step $C_{d,k+1}$ is related to how much is currently in the battery and how much charged within the battery at time k, i.e.

$$C_{d,k+1} = C_{d,k} + 0.5B_{d,k}$$

Finally to simplify the problem, charging ($B_{d,k} \geq 0$) is only allowed from midnight to 3.30PM ($k = 1\ to\ 31$). And discharging ($B_{d,k} \leq 0$) is only allowed from 3.30PM to 9PM $k = 32\ to\ 42$. Further, the battery must start empty at the start of each day in the test week. I.e. $C_{d,1} = 0$ for $d = 1, \ldots, 7$.

Rules for charging PV generation

Finally there is a rule for charging the storage device: Solar generation is always stored first and what is left comes from the grid.

I.e. when $B_{d,k} \geq 0$ be the requested power to be charged into the storage device. If $P_{d,k}^{Total}$ is the total PV average power generated by the solar then one of two scenarios occurs:

1. If $B_{d,k} \leq P_{d,k}^{Total}$, then all the energy comes from the solar generation.
2. If $B_{d,k} > P_{d,k}^{Total}$, then $P_{d,k}^{Total}$ is stored in the battery from solar generation and $B_{d,k} - P_{d,k}^{Total}$ comes from the grid.

Assuming these rules denote the amount of power charged from the PV at time k on day d as $P_{d,k}$, then the proportion of this energy over the day is given by:

$$p_{d,1} = \frac{\sum_{k=1}^{31} P_{d,k}}{\sum_{k=1}^{31} B_{d,k}}$$

Scoring

For each day ($d = 1, \dots, 7$) of the current task week a score is calculate given by:

$$S_d = R_{d,peak}(p_{d,1}C_1 + p_{d,2}C_2)$$

Where

- $R_{d,peak}$ is the peak reduction (as a percentage) during the evening period on day d defined by

$$R_{d,peak} = 100\left(\frac{\max_{k \in \{32,\dots,42\}}(L_{d,k}) - \max_{k \in \{32,\dots,42\}}(L_{d,k} + B_{d,k})}{\max_{k \in \{32,\dots,42\}}(L_{d,k})}\right)$$

- $p_{d,1}$, $p_{d,2}$ are the proportion of energy stored in the battery from solar energy and from the grid respectively on day d. Notice $p_{d,1} = 1 - p_{d,2}$.

- $C_1 = 3$, $C_2 = 1$ are weights for the solar and grid energy, respectively. These weights are based on the relatively lifetime GHG emissions intensity of solar and electricity from the grid.

The final score for the current task week, is simply the average over each day of the week:

$$S_{final} = \frac{\sum_{d=1}^{7} S_d}{7}.$$

Task Weeks

Each Task will encompass a week period:

- **Task 0:** 2018-07-23 to 2018-07-29 (inclusive) – practice challenge.
- **Task 1:** 2018-10-16 to 2018-10-22.
- **Task 2:** 2019-03-10 to 2019-03-16.
- **Task 3:** 2019-12-18 to 2019-12-24.
- **Task 4:** 2020-07-03 to 2020-07-09.

Submission Process

- **Data hosted:** Is hosted on the WPD Open Data hub at : https://www.westernpower.co.uk/innovation/pod/ (registration required).
- **Data:** In Datasets page on hub, in folder labelled "pod_df_taskX" for task X. Separate file for demand data, PV and weather data. Also has the submission template.
- **Submission Template:** Use template provided, update name to give team name, e.g. "teamname_setX.csv" for Task X.
- **Submissions**: Should be emailed to podchallenge@es.catapult.org.uk (also use this email address for other questions about the challenge).
- **Initial Submission**: For first submission (first challenge or practice challenge) cc in all members of the team.
- **Weekly Leaderboard:** Will be emailed to participants once calculated.

Some Rules

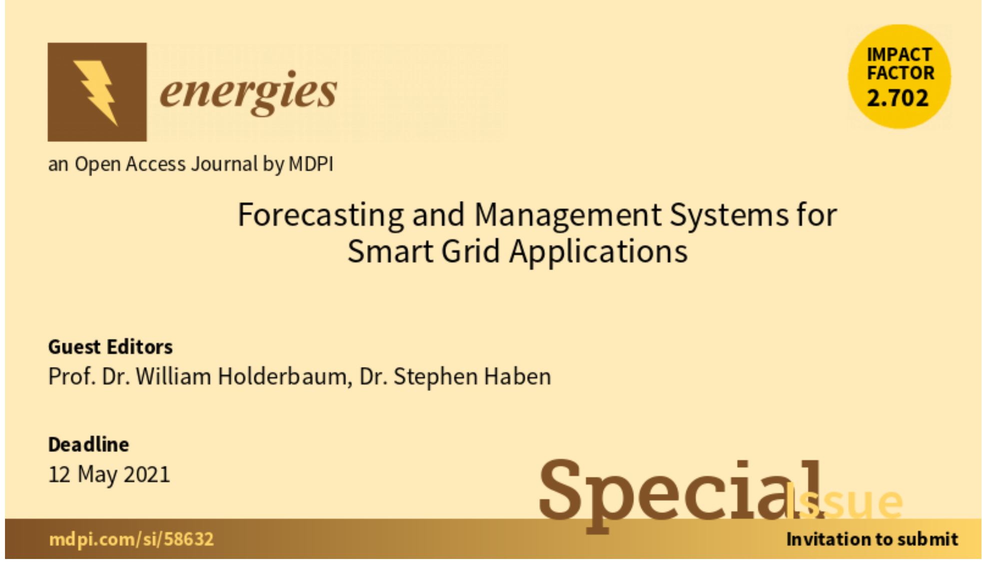
- A check will be made that **constraints are not violated**. Those that do will score zero.
- Each member taking part should **register** at https://www.westernpower.co.uk/pod-data-science-challenge
- **Teams of up to five** are allowed. Need to **notify us who is in your team** via initial submission.
- Need to make a **submission for all tasks** to qualify for prizes.
- Additionally, a **short outline of method** (3 pages maximum) is require to qualify for publication prize and must be submitted **with the final task 4** submission.
- Submissions must **use the template** provided.
- **Cannot use additional data sets** not provided for the challenge.

Important Dates

- **28th January 2021**: Data Science Challenge kick-off event and release of initial data set and practice challenge **Task 0** (submission deadline 11th February).
- **4th February 2021:** Registration Ends and teams must be finalised.
- **12th February 2021:** Start of Task 1 and release of data. (submission deadline 25th February).
- **26th February 2021**: Start of Task 2 and release of data (deadline 4th March).
- **5th March 2021**: Start of Task 3 and release of data (deadline 11th March).
- **12th March 2021**: Start of final Task 4 and release of data (deadline 18th March).
- **18th March 2021**: Submission deadline for Task 4 and for the outline report for methods.
- **End-March 2021**: Winner(s) announced.
- **TBA**: Presentation for challenge winners.

Last thoughts

- **Check the constraints** before submission.
- Note the **battery starts at zero** each day.
- Utilise the **test challenge**.
- Use **daylight savings** information.
- **LinkedIn page** – discussion and forum
- **Energies Journal** open for all entries.



Thanks you and enjoy the challenge!

Stephen Haben

Stephen.haben@es.catapult.org.uk