FARADAY: SYNTHETIC SMART METER DATA GENERATOR FOR THE SMART GRID
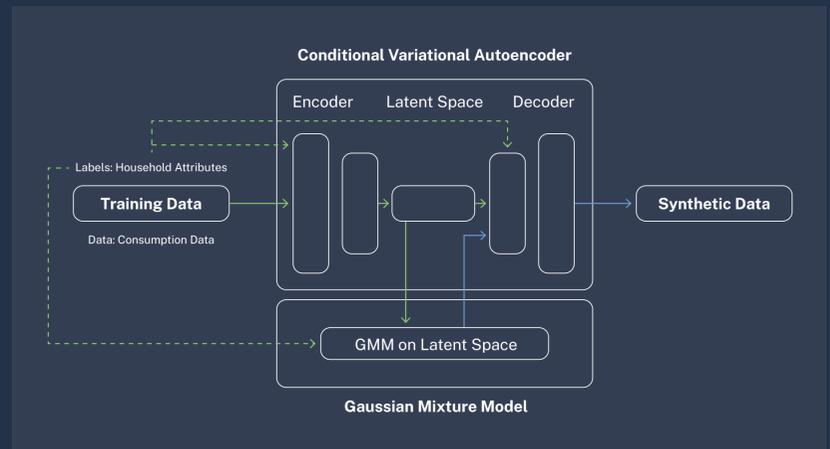
Sheng Chai sheng.chai@centrefornetzero.org
Gus Chadney gus.chadney@centrefornetzero.org

Centre for Net Zero
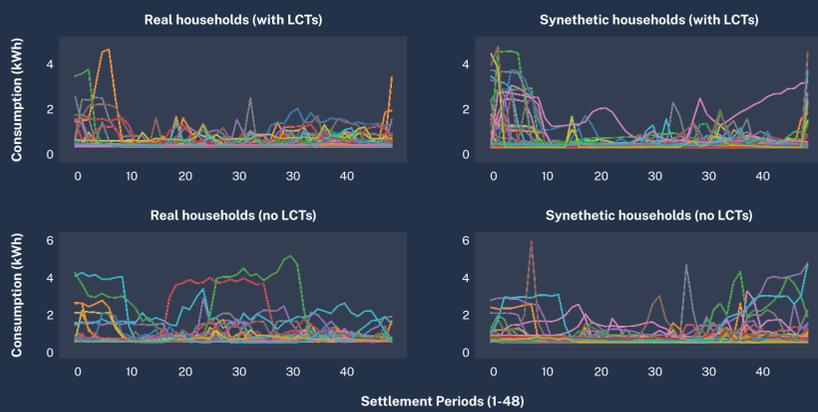Powered by Octopus Energy Group

A GENERATIVE MODEL BASED ON VARIATIONAL AUTOENCODER AND GAUSSIAN MIXTURE

The encoder maps real data into a latent space. A Gaussian Mixture Model (GMM) is then trained on the distribution of the latent space. During inference, random samples are drawn from the GMM and decoded by the decoder to generate synthetic samples.

Conditional Variational Autoencoder
Encoder Latent Space Decoder
Labels: Household Attributes
Training Data
Data: Consumption Data
Synthetic Data
GMM on Latent Space
Gaussian Mixture Model

EXAMPLE LOAD PROFILES GENERATED

Synethic profiles capture patterns seen in real data. Non-LCT households are characterised by peaks around evenings, whilst LCT households are characterised by peaks in overnight periods when electricity is cheap.
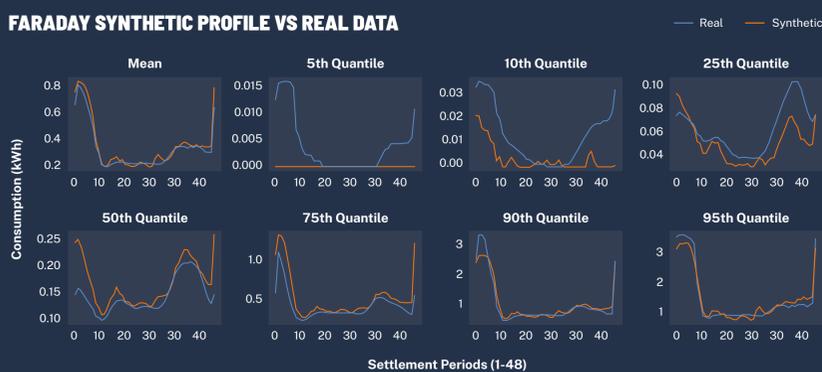

Real households (with LCTs) — Synthetic households (with LCTs)
Real households (no LCTs) — Synethic households (no LCTs)
Consumption (kWh)
Settlement Periods (1-48)

FARADAY IS TRAINED ON 300M+ SMART METER DATA READINGS FROM OCTOPUS ENERGY
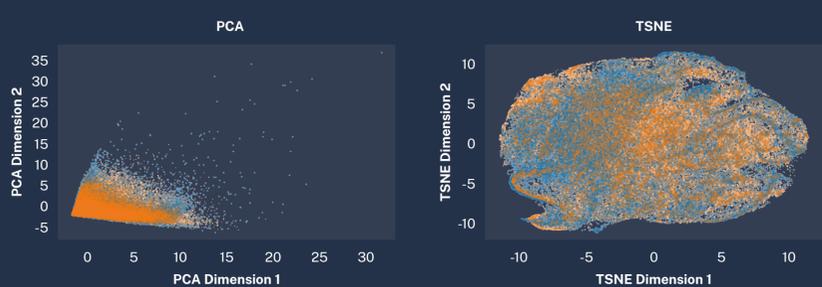
20K households over 1 year period (2021 to 2022) from Octopus Energy
one of the largest electricity suppliers in the United Kingdom. Contains features such as the type of Low Carbon Technology (LCTs) households own, the property type and energy efficiency.

octopus energy

FARADAY OUTPUTS REPRESENT THE STATISTICAL DISTRIBUTION OF REAL DATA FAITHFULLY

FARADAY SYNTHETIC PROFILE VS REAL DATA

— Real — Synthetic

Mean 5th Quantile 10th Quantile 25th Quantile
50th Quantile 75th Quantile 90th Quantile 95th Quantile
Consumption (kWh)
Settlement Periods (1-48)

REAL VS FARADAY OUTPUTS

● Original ● Synthetic

PCA
PCA Dimension 2 / PCA Dimension 1
TSNE
TSNE Dimension 2 / TSNE Dimension 1

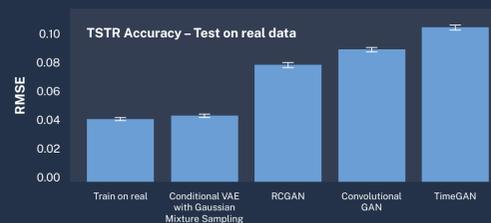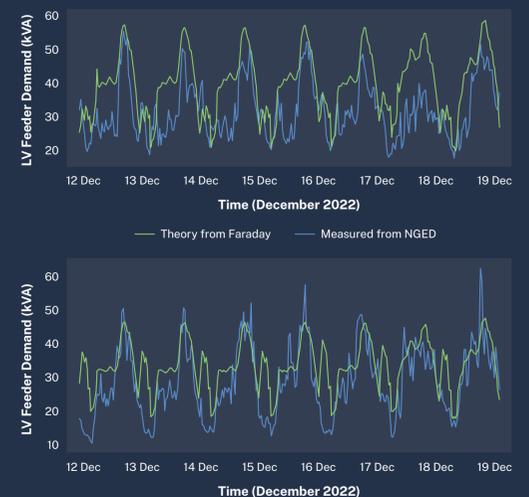## PEAKS IN FARADAY OUTPUTS MATCHES REAL-WORLD SUBSTATION MEASUREMENTS IN TIME AND MAGNITUDE

Faraday outputs aligned in the magnitude and timing of peaks measured when compared to real substation readings in Birmingham UK.

LV Feeder Demand (kVA)
Time (December 2022)
— Theory from Faraday — Measured from NGED

TSTR Accuracy – Test on real data
RMSE
Train on real | Conditional VAE with Gaussian Mixture Sampling | RCGAN | Convolutional GAN | TimeGAN

Faraday is also validated with downstream forecasting applications using TSTR. Forecast built using synthetic data performed similarly to forecast built with real data, demonstrating its utility.

FUTURE WORK ON PRIVACY

Faraday currently has implicit privacy protection e.g. K-anonymity and only outputting daily profiles. More research on privacy e.g. differential privacy, or privacy evaluation methods like membership inference attack would allow us to output profiles of longer time period (e.g. weekly/monthly) which would have higher utility.

Faraday is currently available for alpha testing as an API and a web app. For access, please get in touch with faraday@centrefornetzero.org