Faraday: Synthetic Smart Meter Data Generator For The Smart Grid

ICLR 2024 Workshop:
Tackling Climate Change with Machine Learning

Centre for Net Zero

Powered by **Octopus Energy**

Synthetic Smart Meter Data can accelerate our research and transition towards net zero

- The global transition to net zero transition involves increasing the share of electricity produced from renewable sources and concomitantly electrifying heating and transport.
- This presents new challenges to electricity grids, e.g. new demand peaks, constraints, and mismatches between demand and supply. Households with Low Carbon Technologies (LCTs) can help address this challenge through automation.
- Granular household-level electricity data, especially of households with low carbon technology (LCTs) allow us to build better bottom-up grid models of future energy systems.
- Access to data, however, is limited and restricted due to privacy and governance. Synthetic smart meter data can circumnavigate this issue whilst preserving patterns in real data.

Faraday is a generative AI model that can output synthetic smart meter data conditioned on LCT ownerships of households, the property type and the energy efficiency of the property.

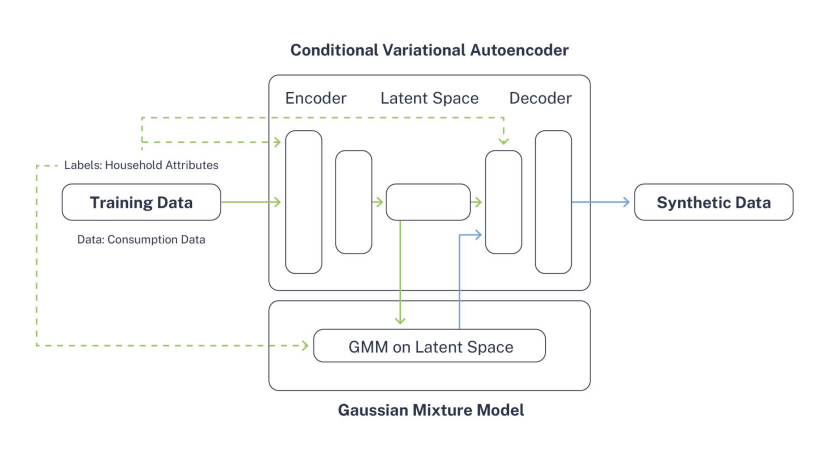
Faraday is trained on 300M+ smart meter readings

- Octopus Energy is one of the largest electricity suppliers in the United Kingdom.
- The training dataset consists of smart meter readings of 20K households over a 1 year period June 2021 June to September 2022.
- The training dataset also consists of additional labels such as the property type, the energy efficiency rating, and the type of LCTs that they own.

Centre for Net Zero's unique access to Octopus Energy's data allows us to train a generative model conditioned on information that is vital for a future grid system where LCTs are prevalent.

Variational Autoencoder + Gaussian Mixture Modelling

- The encoder maps real data into 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 with the decoder.



Modifications to VAE

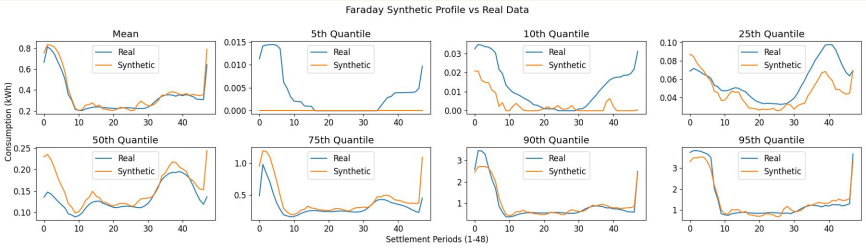
MMD Loss instead of KL-Divergence Loss to handle the non-normal distribution of smart meter data.

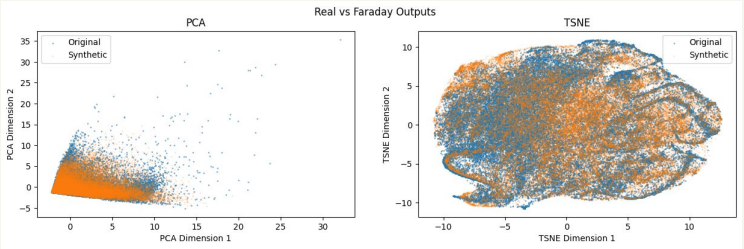
Loss terms also include quantile losses at 5th And 95th quantiles to improve representation at the extremes, an interest for grid researchers.

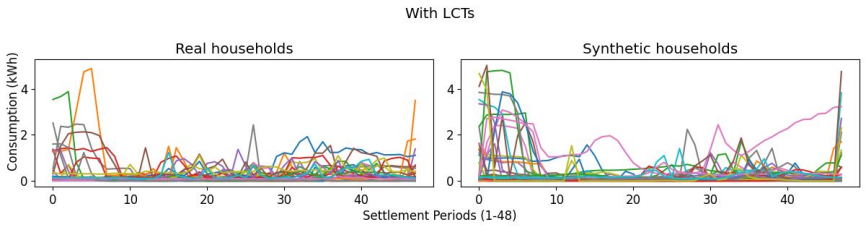
Model is trained with 4 T4 GPUs on Vertex AI

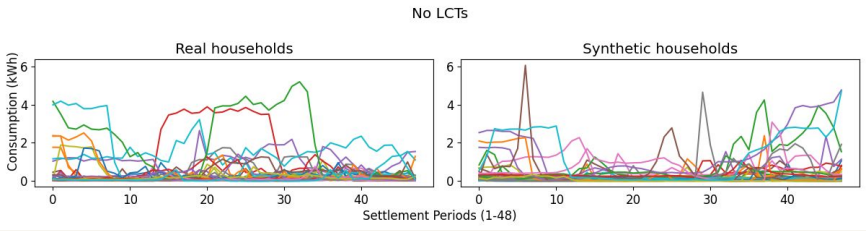
Total training time is ~ 8 hours for both VAE and GMM over 250 epochs on 300M smart meter readings.

Faraday produces high fidelity outputs







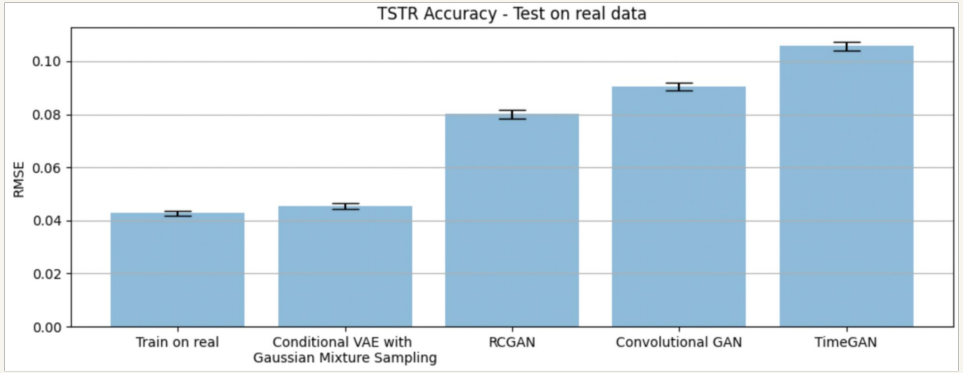


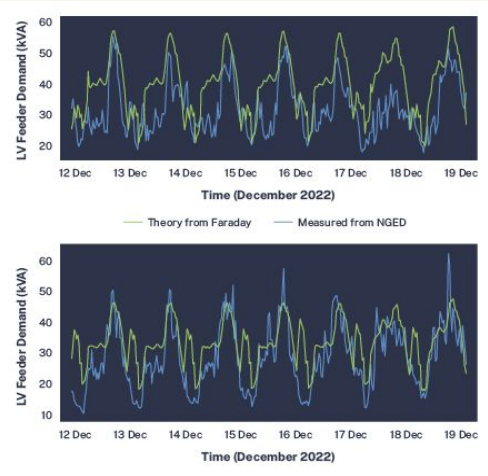
Faraday outputs statistically similar to real data.

When conditioned on whether they own LCTs, Synthetic data captures general patterns in real data:

- Households with LCTs have peaks in early morning hours when electricity is cheap
- Households with no LCTs have more evenly distributed load profile

Faraday produces high utility outputs





Train on Synthetic, Test on Real (TSTR) evaluation produces similar results for real vs synthetic outputs

Two competing forecasting models are trained to predict 2022 consumption from real and synthetic 2021 consumption respectively.

If synthetic data is useful, model trained on synthetic data should perform similarly to model trained on real data.

Faraday outputs aligned with real world substation data in Birmingham, UK

A group of alpha testers from the University of Birmingham compared Faraday outputs to four substations in Birmingham and found that the timing and magnitude of peaks aligned well.

Future Work: Privacy

Faraday currently has implicit privacy protections:

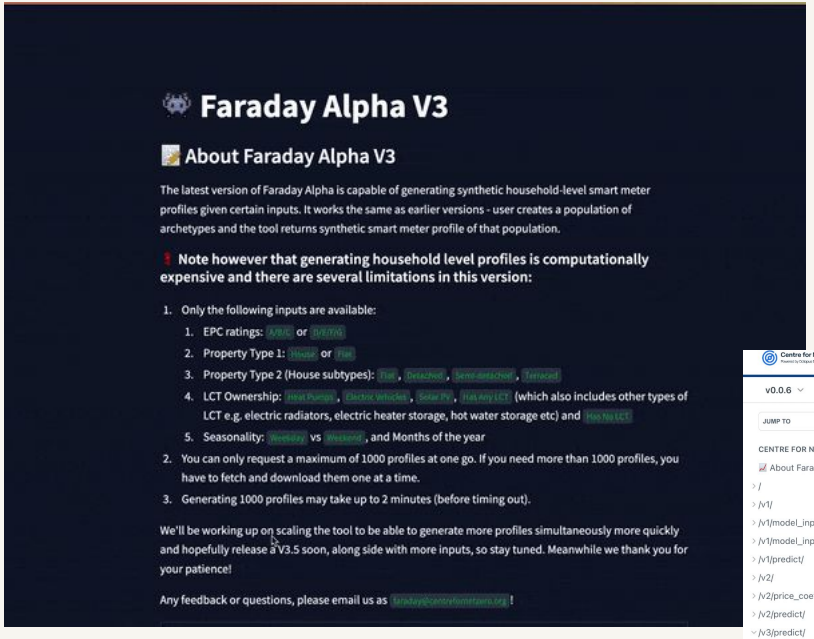
- Only outputs daily profile to limit re-identification risk.
- K-anonymity of 3: minimum of 3 households at the finest grain of the data.
- Model is only exposed as partial black-box setting via an API. Model weights are not published.
- Random noise added to smart meter data prior to training.

More work and research needs to be done on privacy e.g.:

- Differential Privacy and its impact on fidelity/ utility.
- Explicit privacy evaluation tasks e.g. membership inference attacks.

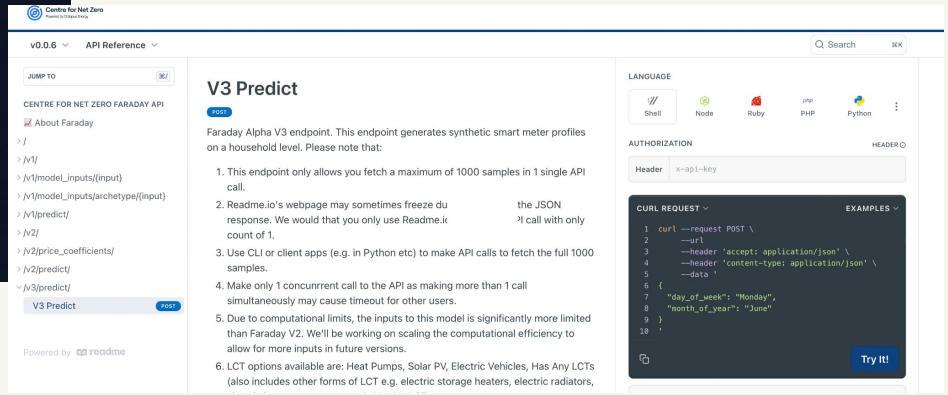
More robust research and privacy protection can allow us to release synthetic data of longer time horizon e.g. weekly/ monthly which would have even higher utility.

Faraday is Available as an API and Web APP



Web App

Contact
faraday@centrefornetzero.org
for access and information regarding your use case



API Documentation Site