UIG Task Force

13.2.8: Demand Uncertainty Estimation

Demand Uncertainty Estimation

- Task Description:
  - Estimating NDM Energy Consumption is a complex problem. Any prediction, including those calculated by Machine Learning models, will include an element of uncertainty.
  - The uncertainty can be broadly grouped into two categories:
    - Systematic uncertainty – that is, actual behaviour that can be measured in the custom that is inherently unpredictable and uncertain. This uncertainty is irreducible and will be influence any prediction.
    - Prediction Uncertainty – that is, uncertainty in the accuracy of the estimate generated by the algorithm or model in question. This uncertainty can potentially be reduced by tuning the model and / or input parameters.
  - This task is designed to help quantify and explain that uncertainty. It could also ultimately inform a realistic expectation of the UIG uncertainty level that could be observed in the daily energy prediction.

- Summary of Results:
  - The Analysis suggested daily upper and lower bounds for typical NDM consumption for a given Meter Point in EUC1 throughout the Gas year.
  - When used to test the actual consumption for the NDM sample meter points, the uncertainty bounds flagged a number of meter points that were used to train the NDM Models which have atypical consumption patterns (consumption that falls outside of the 'normal' range for more than 25% of observations)
  - Removing these Meter Points from the training set improves the performance of the ML models when predicting energy at LDZ level
  - These uncertainty estimates could be used as a Sample Meter Point validator and improve the quality of training data used to train the NDM algorithm.

Measuring uncertainty in machine learning

When measuring uncertainty it should be noted that every estimate has errors due to:

- Inherent randomness in the target value
- Model uncertainty/parameter uncertainty

These errors could be expressed as

a) A fitted distribution

b) An interval (e.g. a 95% interval)

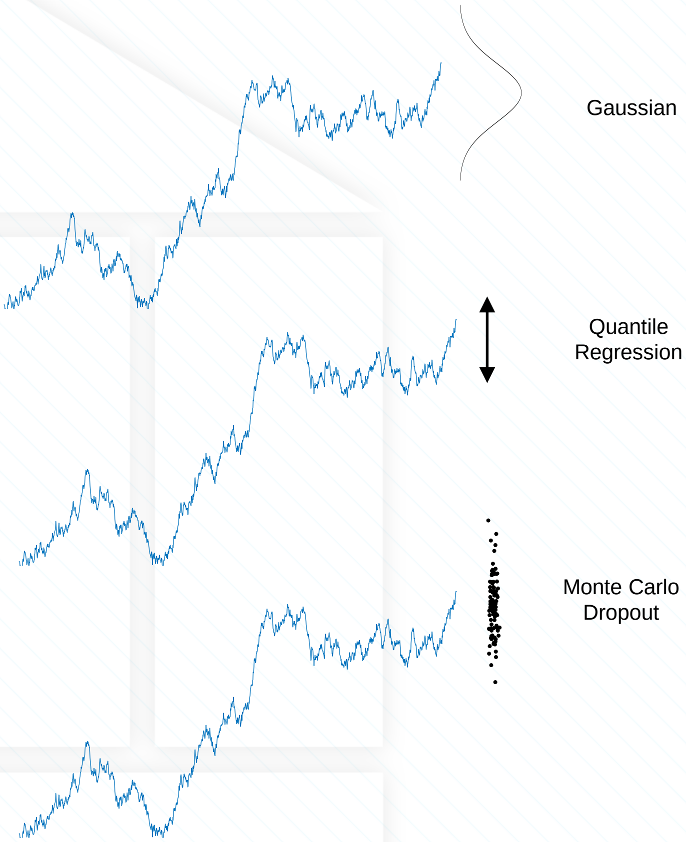
c) A Monte Carlo empirical distribution of possible outcomes

Our analytics partner have developed prototypes representing the below:

a) Fitting a Gaussian (or the related log-Gaussian)

b) Quantile regression

c) Monte Carlo dropout



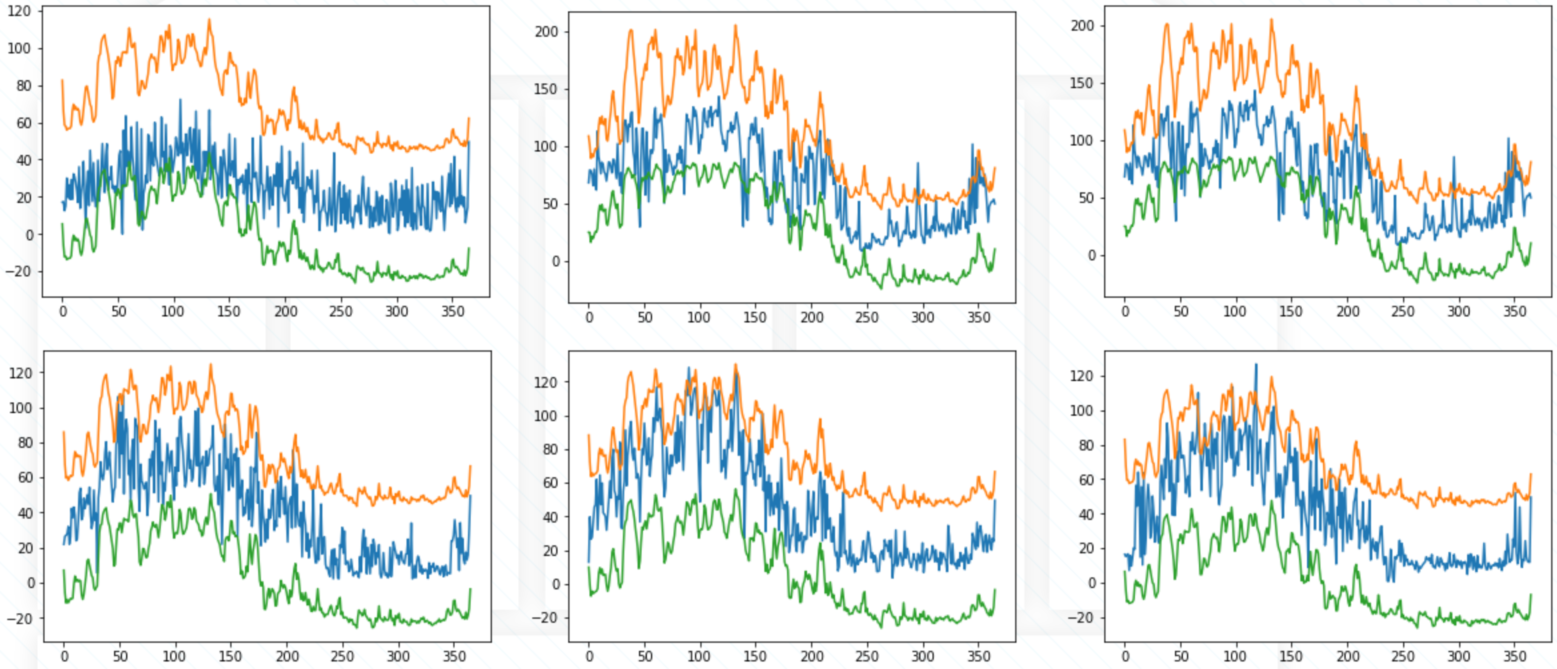
Terms of reference	
Gaussian	Sometimes referred to as normal **distribution**, this is a mathematical function that defines the probability of a number in some context falling between any two real constants.
Quantile Regression	A type of regression analysis used in statistics and econometrics
Monte Carlo dropout	A method used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables

Sample charts representing fitting a Gaussian to single meter points - domestic meter points only

The orange series represents the upper boundary where the top 2.5% of consuming sites fall outside that specific boundary on a given day

The blue series represents the actual consumption

The green series represents the lower boundary where the bottom 2.5% of consuming sites fall outside that specific boundary on a given day



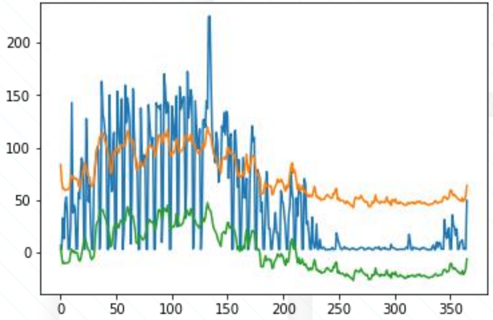
Sample charts fitting a Gaussian to single meter points, domestic only, highlighting cases which don’t fit

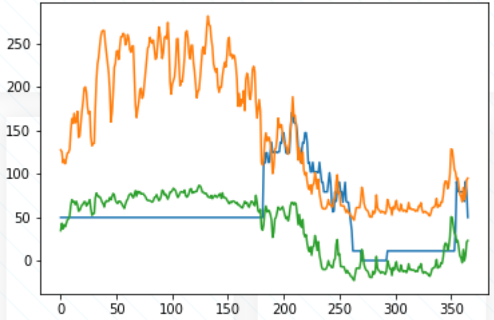
It was important to include meter ‘outliers’ in order to discover whether errors are the result of the specific site or separate external factors. The below charts show examples of meter points that fall outside normal patterns of behaviour. We wanted to discover whether including these would enhance or diminish the performance of the uncertainty estimator and the demand estimation algorithm itself.

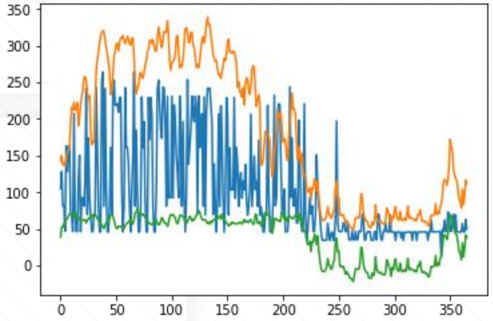
The orange series represents the upper boundary where the top 2.5% of consuming sites fall outside that specific boundary on a given day

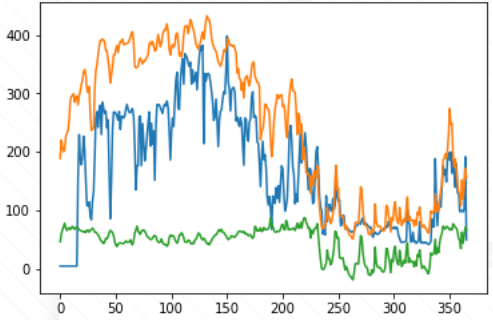
The blue series represents the actual consumption

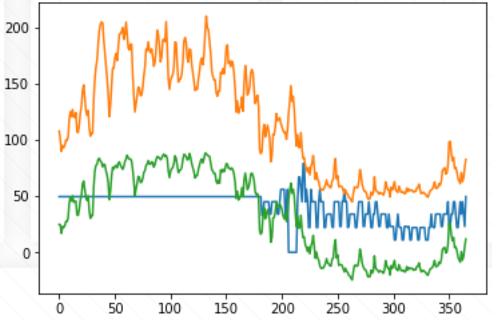
The green series represents the lower boundary where the bottom 2.5% of consuming sites fall outside that specific boundary on a given day

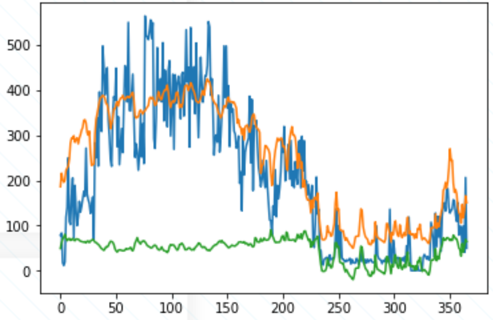












Combining Uncertainties

The subsequent phase of analysis work involved combining uncertainties whilst incorporating the following considerations.

- The estimate of the *total* consumption in an LDZ, or any other collection of users, is the sum of the individual estimates
- Uncertainties don't combine in the same way
  - N completely dependent errors sum as N
  - N completely independent errors sum as sqrt(N)
- For EUC1 NDM, total demand is probably somewhere in between these two
  - Some prediction errors will be specific to a single home or business (e.g. this family is away, that family has guests staying)
    - The model overestimate usage by 20% for some users, underestimate by 20% for some others and the total can end up more accurate than any individual estimate
  - Some prediction errors will be common to all users (e.g. an extreme weather event at the limits of the model's training)
    - If the model systematically overestimates by 2%, the total will overestimate by 2%

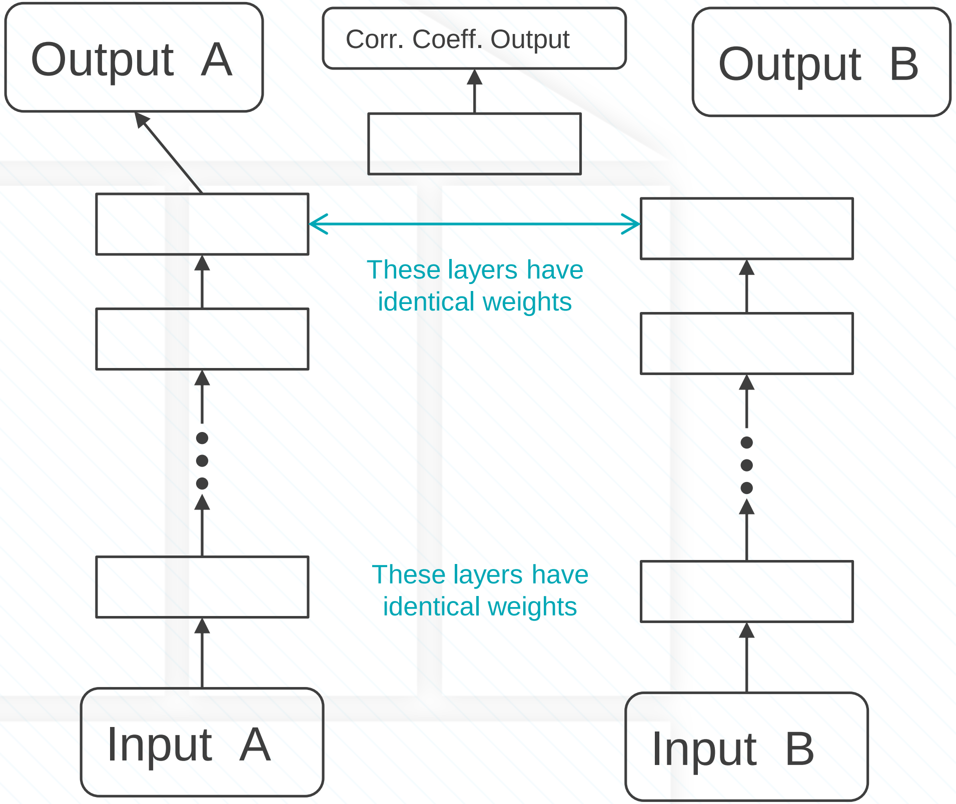
In simple terms, the above approach can be explained as working with a single meter point in the first instance in order to understand how uncertainties combine. Multiple meter points are then included to discover correlations over a large data set. This is a beneficial approach as combining uncertainties results in better estimates.

The analysis work on combining uncertainties gave rise to the following questions and findings as detailed below.

- What is the uncertainty on the *total* estimate, if we have bounds on individual users?
  - We could assume completely independent errors
  - Leads to bounds which are generally too narrow, getting worse as we add more estimates to the sum
- We could assume completely correlated errors
  - Leads to bounds which are too wide
  - Some of the individual-specific errors “average out”
- We could assume there is some degree of correlation
  - Can produce “intermediate” bounds that are not too narrow (like independent errors)
  - Need some measure of correlation: there is an additional parameter to fit with this method
- Our investigation so far has found
  - There is error correlation (so the first approach is not suitable)
  - The errors are not 100% correlated (so the second approach is too pessimistic)
  - The correlation coefficient varies, but we can model that variation and use that model to produce reasonable bounds

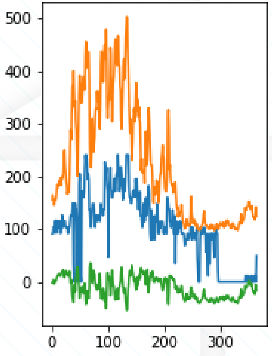
Pairwise estimate

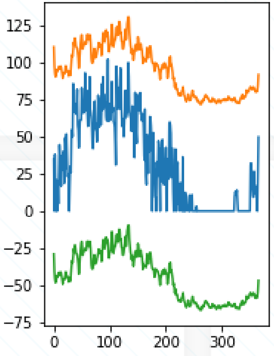
- Can we build a NN model that also estimate the correlation between two users?
- We can fit a **joint** distribution the same way we can fit a **single** sample distribution
  - Train on random *pairs* of meters from the training data
- We train a model that, for each day, estimates a 2-D Gaussian distribution for a pair of meters *and* a normalized correlation coefficient.
  - Each input goes through several layers with weight tying (so the model is identical for A and B)
  - Output layers for A and B generate mean and s.d. for each
  - An separate output layer takes input from *both* sides and generates a correlation coefficient



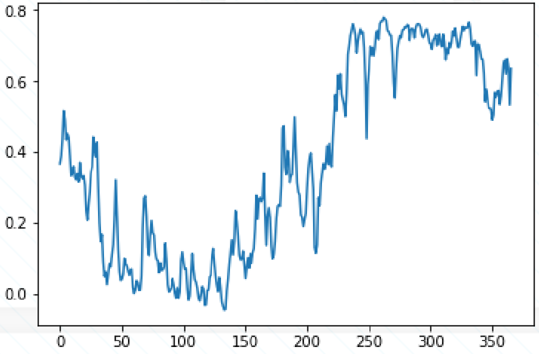
Pairwise results

Individual Readings (with error bounds)
Truth: blue

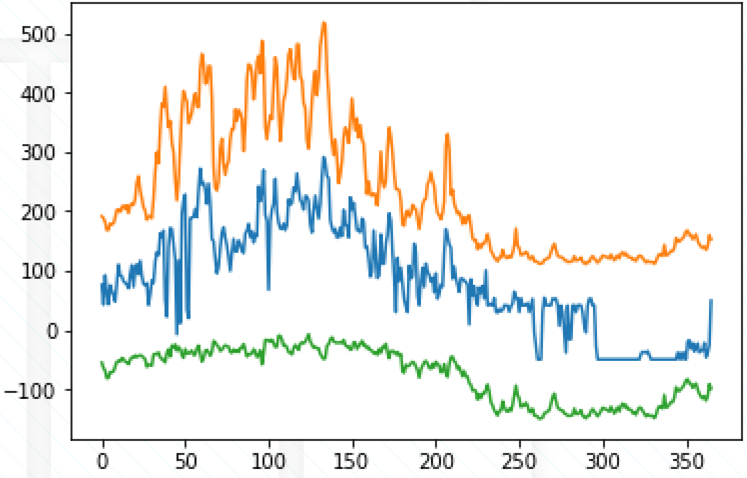




Estimated error covariance showing seasonal variation



Sum of two (blue) with error bounds (green, orange) produced using the covariance to tighten them

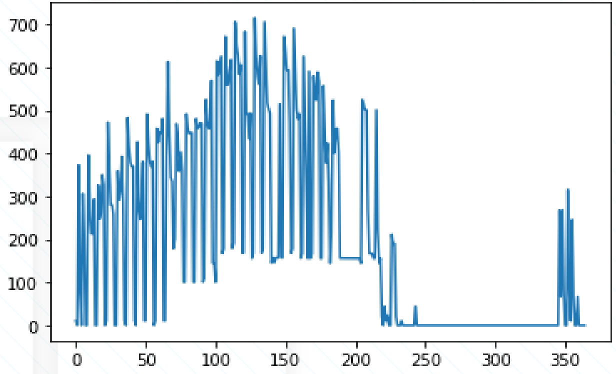


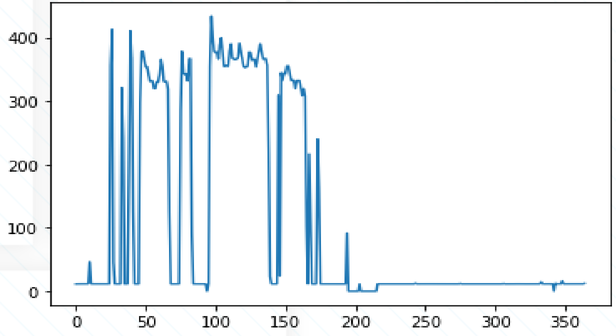
Combining many estimates

- If we assume the errors are Gaussian, the pairwise correlations give us enough information to bound the total
  - This is an assumption that might not hold in all circumstances, and in particular “tail events” might not be well captured by this model (e.g. another Beast from the East might be outside these bounds)
- Early results show it is behaving as expected: as we add 100s of meters, we find that:
  - Assuming independent errors fails: too many points lie outside the estimated interval
  - Assuming fully correlated errors is more conservative than needed: it is considerably wider than needed
  - The partially correlated error model produces intervals that cover over 95% of the true points
- There is a scaling issue as implemented (though we can likely address this)
  - At present it evaluates the correlation coefficient on all pairs of meters
  - We can replace the sum over all pairs of meters with a Monte Carlo evaluation
- There is also an issue with negative correlation coefficients
  - Need to ensure that the covariance matrix is positive definite
  - Could fix this by enforcing all correlation coefficients to be greater than zero (slightly conservative assumption)

Uncertainty estimator as sample-set validator

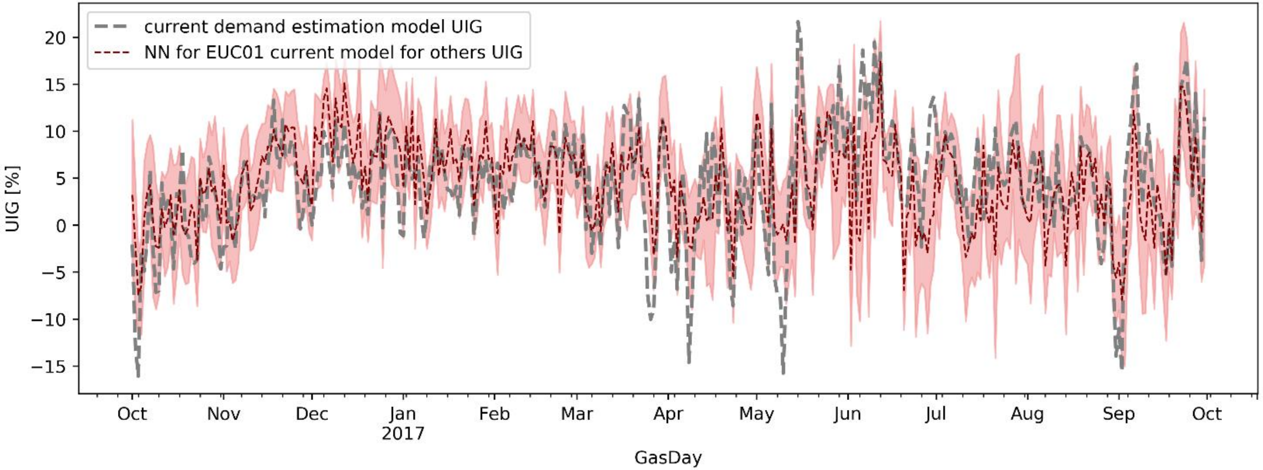
- When prototyping the uncertainty estimator on individual sample-set meter points, it was observed that a number of meter points had a high percentage of readings outside of the bounds of the uncertainty estimator
- This is a useful application of the uncertainty estimator to validate sample set meter points
- Provided details of 61 meter points which had more than 25% of their readings outside of the error interval estimated by a prototype uncertainty estimator (which was set between 2.5% and 97.5%). Following might be worth investigating for each of these meter points:
  - Are any of these sites commercial/industrial rather than residential? E.g. they only use energy Monday to Friday.
  - Are there data quality issues? E.g. the daily meter read is set a fixed value for multiple days.
  - Why are there meters which appear to be duplicates?





Summary

- Have developed a means to estimate uncertainty **on EUC1 only** in the neural network NDM estimate.
- This provides a means to 'validate' the sample set data.
- The overall performance of the neural network might have been improved due to changes in model and removing sample meter points that don't appear to be representative.
- Some of the understanding and approaches developed here are potentially useful when applied to long term reconciliation.



Model Summary	
Model	Sequential Model with modification for uncertainty estimator
Segment	1 model per LDZ for EUC1 only
Inputs	Standard set with some outliers removed,
Training	GY 2006-2017 (exc. 2016)
Testing	GY 2016

How it works: top level summary

Estimate uncertainty for individual meter points

- **Relatively easy to do**
- Used to flag up discrepancies in sample data set

Correlate uncertainties across multiple users

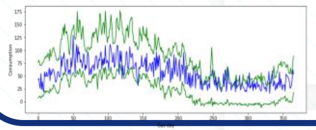
- i.e. how do we combine uncertainties?
- **Impractical to carry out between all pairs**

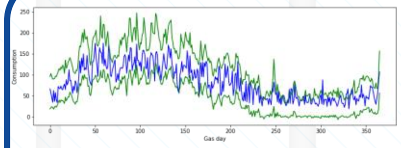
Monte Carlo estimation of distribution across meters

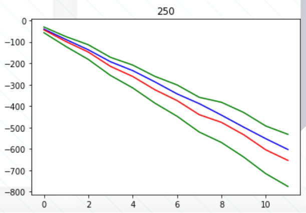
- Error bounds scale up correctly
- Likely provides a conservative approximation of uncertainty

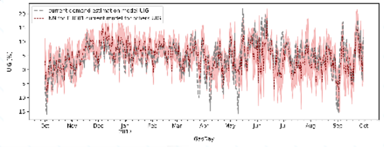
Run against full meter population

- Will depend on representativity of sample data set



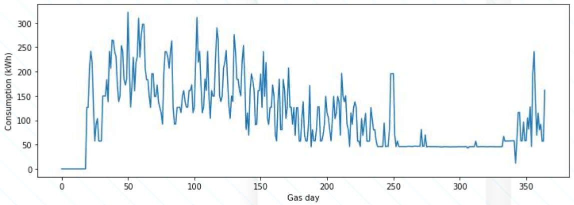


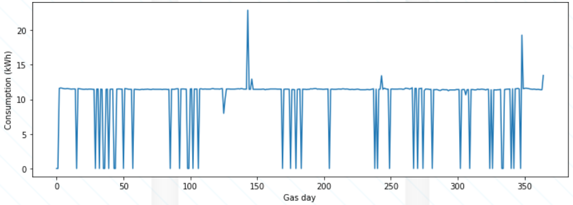


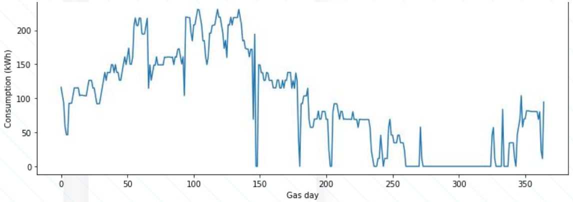


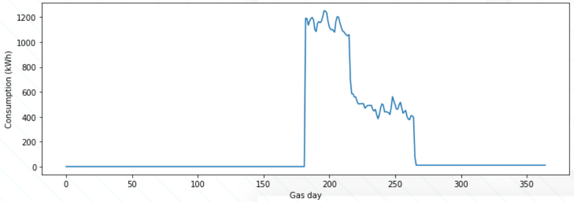
Interesting results

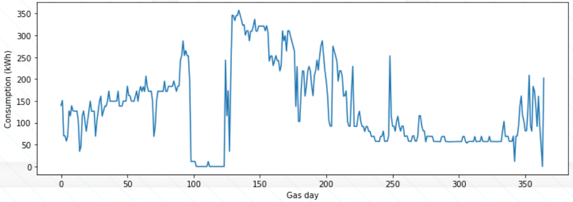
- In trying to produce sensible uncertainty distributions, it was discovered that a number of meter points were adversely affecting the training process due to highly unusual patterns of behaviour. By removing these points from the sample dataset, we were able to improve the performance of both the uncertainty estimator and the demand estimation algorithm itself.
- Examples of these meter point consumption patterns are presented here:

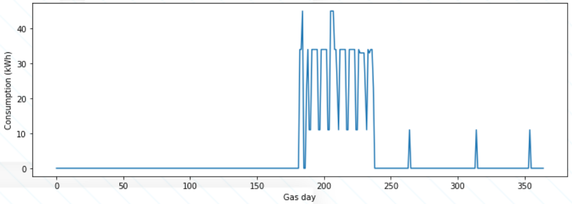






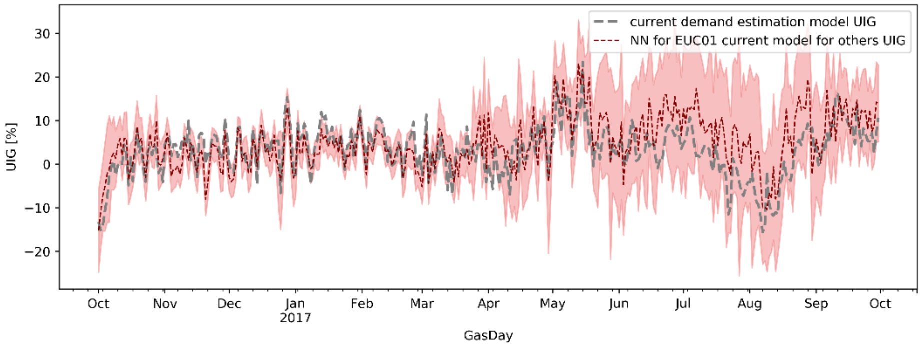




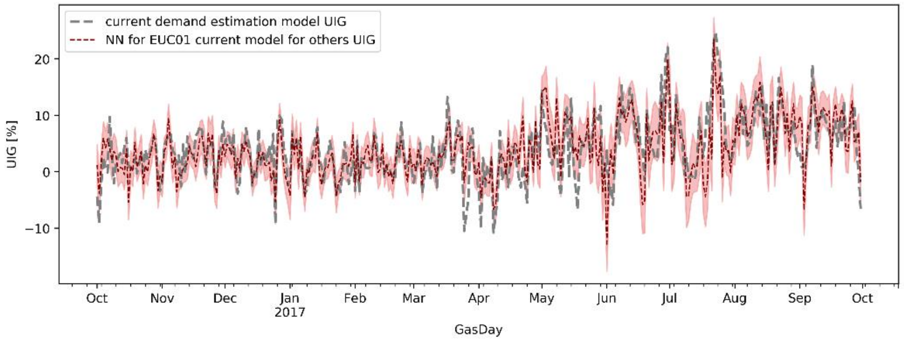


Uncertainty Results

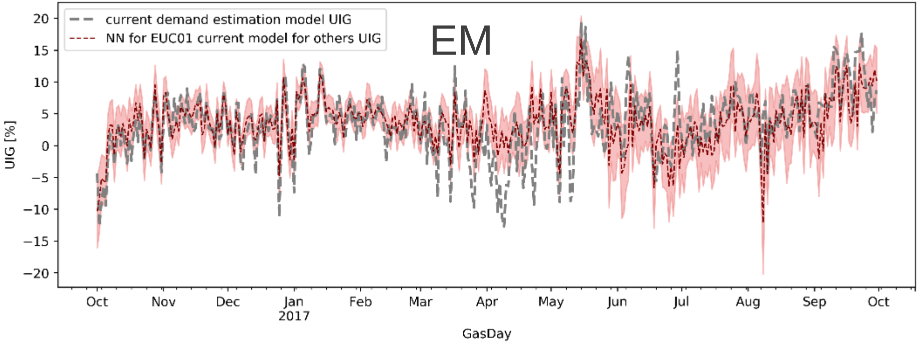
EA



SC



EM



Uncertainty results shown with some points filtered. Results still with sequential NN.

Filtering some points mostly improved means and standard deviations, but EM was an exception.

Filtering mostly improved uncertainties but EA in the summer was an exception.

Further insights and next steps

- There can be high individual fluctuations between meters but these seem to average out overall.
  - What matters most for uncertainty on the total is the common, systematic error
- The relative uncertainty from domestic EUC1 alone can be several percentage points of total demand.
- There is seasonal variation in uncertainty
  - absolute uncertainty is highest in winter, while relative uncertainty is highest in summer
- We expect that improvements to the model (LDZ input, functional model) would also reduce the uncertainty
  - A next step would be to add uncertainty estimation to these models
- This would need to be extended to EUC2+
- Filtering out certain examples seems to improve the uncertainty fit without harming overall performance
  - We might be able to be smarter and just filter out dates that seem to be missing (also addressed by the functional model discussed in the Machine Learning Findings 13.2.6)

xoserve