13.2.6: - Accuracy of NDM Algorithm - use of Weather Data – Advanced Machine Learning

Summary of this Findings Pack

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	Non-Daily Metered energy is estimated each day using the NDM allocation algorithm. The goal of this analysis is to see whether we can improve the prediction of NDM energy using new, innovative tools and algorithms. Improving NDM allocation accuracy should reduce day-on-day UIG volatility and base level UIG at Allocation, meaning less movement of energy and revenue at reconciliation and therefore less commercial risk for our customers. In this pack, we explore various approaches to improving NDM allocation accuracy using cutting edge machine learning tools.
Data Tree References	EUC, Energy, Annual Quantity, Weather, CWV

Findings Status	**Ongoing**
UIG Impact Peak Volatility %	**20% reduction**
UIG Impact Annual Average %	**70% reduction**
Confidence in Percentages	**+/-5% (will vary by LDZ)**

Findings

The results of each line of analysis follow in this pack, but the overall outcome is that using a Neural Network Machine Learning Algorithm improves the accuracy of NDM prediction. The best results are highlighted in green.

Using weather and consumption data between October 2006 and October 2016 combined with the cold weather and corresponding demand data from Winter 2017-18 to train a Neural Network Model produces an algorithm that:

1. **reduces base UIG at allocation by 70%** on average (so **4% UIG would be 1.2% using this algorithm**)
2. **reduces volatility at allocation by 20%** on average (so **+- 10% daily UIG change would be +- 8%**).

We also learned that using more historic consumption and weather data improves the prediction accuracy (contrary to the current model which limits the analysis to three rolling annual periods), so this model could become more accurate over time as we gain access to richer sources of daily consumption data.

Implementing a Neural Net would result in a 'black box' algorithm that would allocate NDM energy. The algorithm would be too complex to share directly, but if implemented then Xoserve could provide the algorithm to our customers using an API or web service, or similar.

Analysis in this Pack

Further considerations

Whilst these models show a significant improvement in predicting EUC01B demands, there a number of considerations which could affect the success of this modelling approach if it were to be used in practice:

Limitations of the Analysis	Limitations in actual application
• Models developed for EUC01 only – other EUCs would likely require bespoke models • Analysis doesn't include the new EUCs for I&C and Prepayment sites • Sample sizes of the higher EUCs may not be sufficient – or representative of actual usage in that EUC • Models were tested against Gas Year 2016 only – for pre-Nexus months UIG values have been simulated using Nexus formulae.	• Use of a Neural Network can improve the prediction of NDM Demands – but does not predict any other UIG causes • Complexity of the model – would make its replication in Shipper/Supplier systems much harder. Would effectively become an Xoserve provided service. • Unexpected changes in behaviours or other demand drivers – would not be reflected in the models unless the deployed model was set to continually, iteratively retrain itself using incoming data.

Potential Next Steps

- **Generate Neural Network models for all of the other EUCs as well as EUC01.**
- **Train the models for all LDZs simultaneously, but with the additional parameter indicating which LDZ each of the training sample sites belongs to. This may give some of the benefits expected by using an increased amount of data, but without the drawbacks seen when developing the Mainland UK model.**
- **The additional meter location parameters (latitude, longitude elevation and population density) could be used as inputs for the LDZ and EUC specific NN models.**

Summary of Findings: Machine Learning EUC01 with XGBoost

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	It may be possible to improve the NDM demand estimation model through the use of machine learning (ML). This may be because the dependence of gas usage on weather or AQ is complex and difficult to capture in a 'hand crafted' model. We also have additional weather data available which might help with the prediction. This data can easily be incorporated into a ML model. To test this hypothesis we will train ML models to predict NDM demand for domestic meters in EUC01, as the findings in item 13.3.2 showed that EUC01 makes the largest contribution to baseline UIG. Having previously used AQ data for the NDM sample set for gas years 2014, 2015, and 2016, we expanded the data to cover back to 2011, and additionally several months of data over a cold winter for 2017. This permits training on gas years 2012-2015 and on winter 2017 (as a significant portion of the weather data for 2011 is missing) and testing on 2016. More data, particularly more weather data, should improve the model. It is likely that recent temperature history, or temperature time derivative features will improve the model as people's behaviour, and thermal masses of buildings typically introduce time lags. We will include a historic and time derivative temperature features into ML models.
Data Tree References	Wind, temperature, CWV

Findings Status	**Ongoing**
UIG Impact Peak Volatility %	**15-20%**
UIG Impact Annual Average %	**10-25%**
Confidence in Percentages	**+/-5% (will vary by LDZ)**

Findings	**Approach to analysis**
It was determined that usage in the sample set could be predicted better than the current NDM sample model; using the additional weather inputs various measures of UIG and UIG volatility on the NDM sample set are 80-90% of their values using the current NDM model. • A strong predictor for the ML model was CWV. • More data improves the ML model. • Temperature derivative features have a small impact, but including the average temperature over the last few days has a larger impact on a model which does not include CWV. ***What does this mean?:*** *It is likely that ML model would predict usage outside the sample set more accurately than the current model, however this has not been verified. In order to obtain a step change in performance it is likely that more data would be required.*	We tested XGBoost model performance on gas year 2016 data. We compare training on 2014 & 2015, training on 2012, 2013, 2014 & 2015, and training on 2012-2015 + winter 2017. Additionally we compared training with and without temperature history and derivative features on 2012-2015 + winter 2017. We also compared raw weather to CWV and SNCWV.

Supporting evidence 1: Results table – Lower is better.

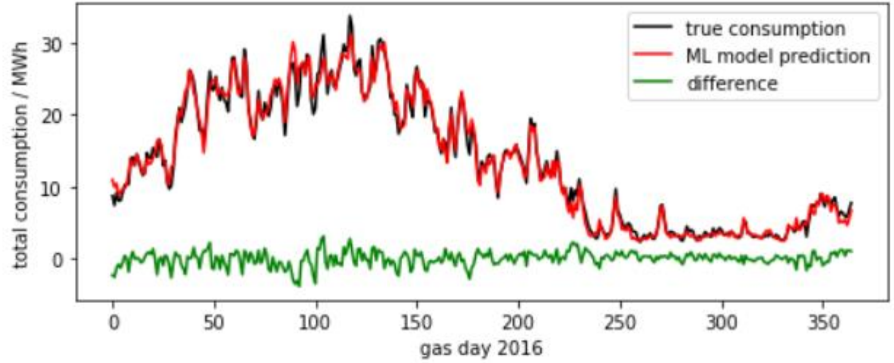
- Test year is gas year 2016 with UK-Link AQs. Trained with AQs (one per meter, gas year) calculated from real usage and WAALPs
- Model A: uses AQ, holiday indicators, day of week, month of year, and a set of raw weather inputs including a temperature gradient feature, temperatures from the previous day and the mean average temperature of the past 3 days.
- Model B: uses all the same inputs as model A, but with the CWV added. This turns out to be a very useful input, suggesting that it may be possible to produce a better model by using fewer better engineered features (like the CWV) rather than expecting the ML algorithm to generate these from raw inputs. This is typically true for data limited problems.
- Separate models were trained on:
    - Years 2014 and 2015
    - Years 2012 to 2015
    - Years 2012 to 2015 and winter 2017 (w17)
- We considered several different error metrics of base error and volatility error, but the model was trained to minimise the daily usage error on each meter. It would be possible to train a model to minimise these other metrics, (however this is difficult and unlikely to result in a step change in performance).

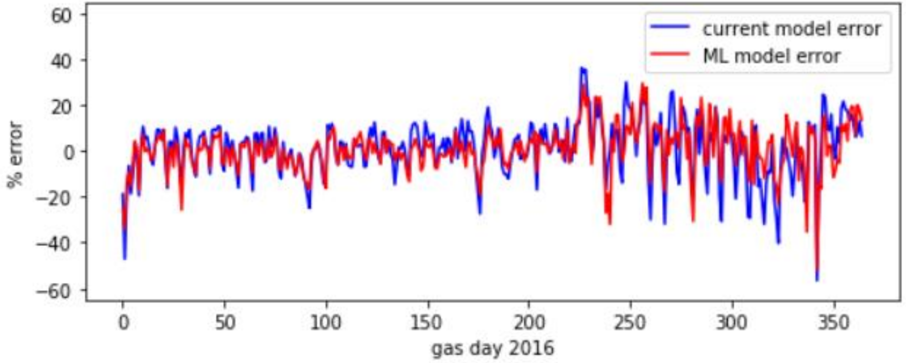
			Measures of base UIG	Measures of UIG volatility		
LDZ	New model training years	Model	Mean daily error (kWh)	RMS error on total daily usage (kWh)	StD. on total daily error (kWh)	Mean (abs) day-to-day change in error (kWh)
EA		Current	570	1,086	924	791
	14,15	XGB A	802	1,395	1,142	1,036
	12→15	XGB A	541	1,084	940	827
	12→15, w17	XGB A	586	1,105	937	810
	12→15,w17	XGB B	561	940	754	715
SC		Current	283	711	652	537
	14,15	XGB A	402	851	750	635
	12→15	XGB A	278	703	646	529
	12→15, w17	XGB A	245	660	613	499
	12→15,w17	XGB B	214	590	550	440
EM		Current	226	1,378	1,359	1,051
	14,15	XGB A	412	1,487	1,429	1,116
	12→15	XGB A	37	1,229	1,228	920
	12→15, w17	XGB A	64	1,188	1,186	894
	12→15,w17	XGB B	40	1,004	1,004	754

Supporting evidence 2: Plots of ML results

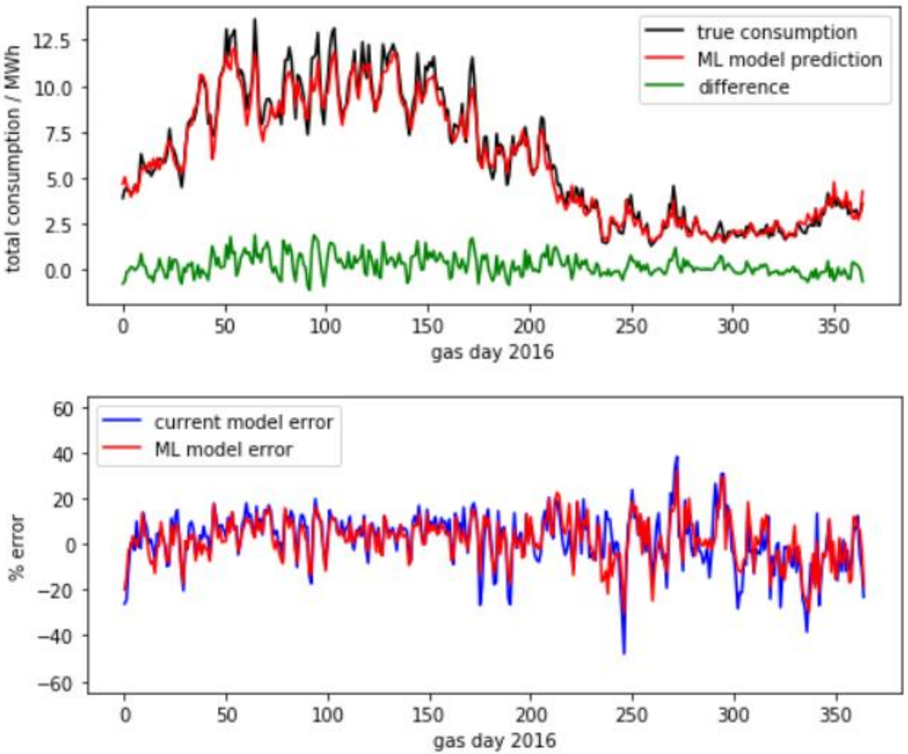
EM, XGB 12-15 Winter 17 Model B

- For the best model (trained on all available data and using CWV as a feature (model B), the results are shown here for 3 LDZs for the 2016 test year.
- The errors on the ML model and the current NDM model are also compared on the lower chart.
- When CWV is included in the model inputs, the prediction error for the ML model follows the current NDM model closely, but the volatility spikes are smaller suggesting that the ML model reacts more quickly to weather changes.

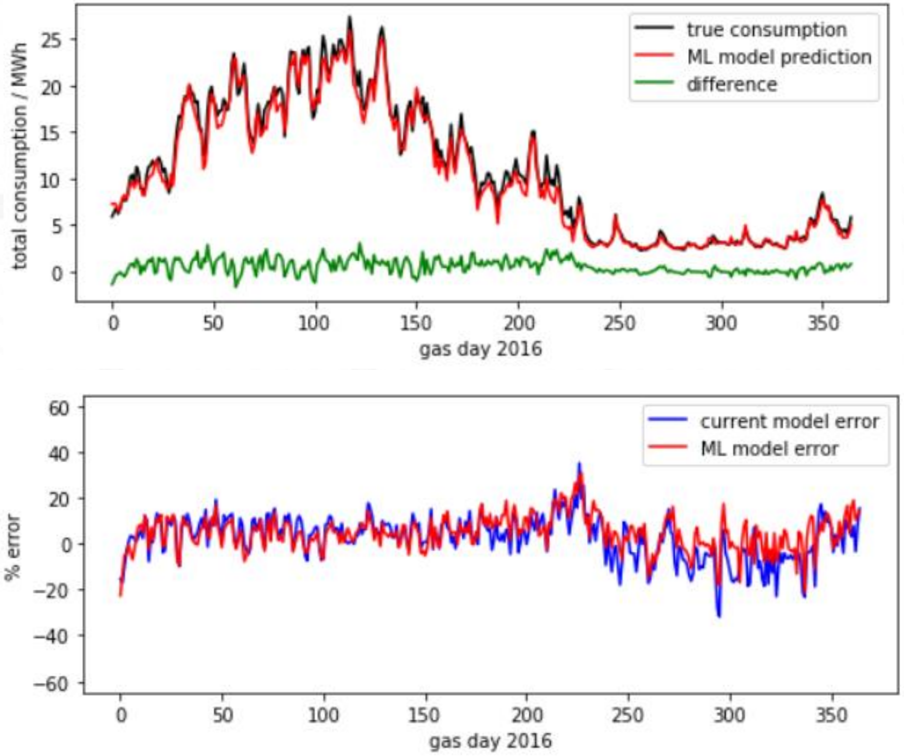




SC, XGB 12-15 incl. Winter17 Model B



EA, XGB 12-15 incl. Winter 17 Model B

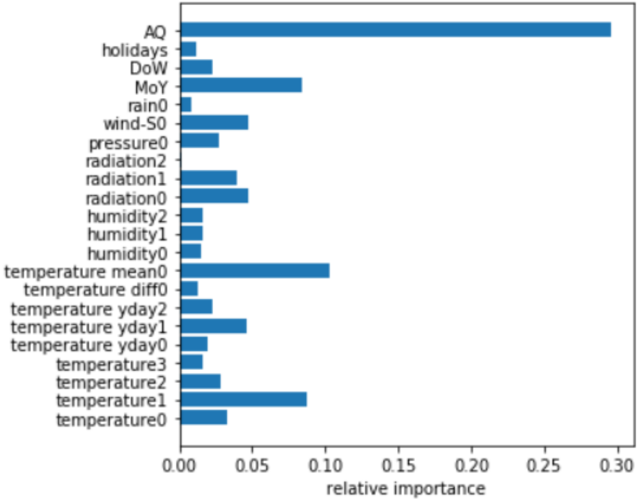


Note lower volatility with ML models

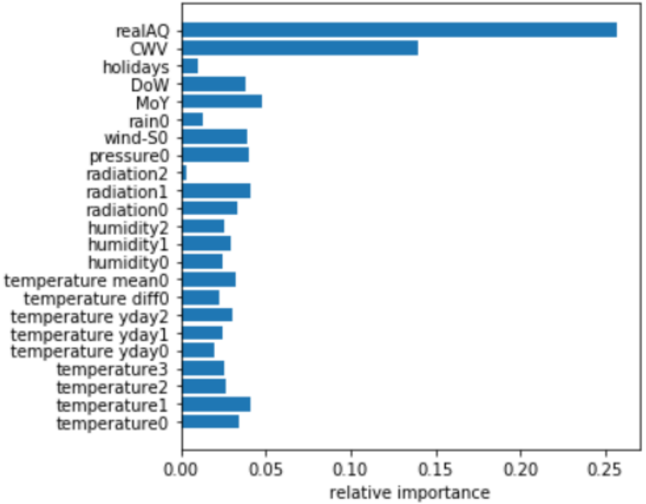
Supporting evidence 3: Feature importance for XGBoost

- A useful feature of decision trees is an output showing how often the inputs are used in the model which tells us how important they are to the accuracy of the model, and therefore we can estimate how useful they are as predictors.
- AQ is clearly the most important input as would be expected given the AQ is the calculated annual usage level for a site.
- Inputs temperature0, temperature1, temperature2, temperature3 are mean average temperatures for 05:00-10:00, 11:00-16:00, 17:00-22:00 and 23:00-04:00. temperature yday0, temperature yday1 and temperature yday2 are the mean average temperatures from the previous day for 05:00-12:00, 13:00-20:00 and 21:00-04:00.
- The mean average temperature over the past 3 days is surprisingly important for model A.
- Comparing model A with model B, the CWV in model B takes over the role of the temperature features, showing that this is a well designed model input and that the Effective Temperature memory in the CWV is important for accurately estimating usage.

model A



model B



Summary of Findings: Testing XGBoost model on EUC02

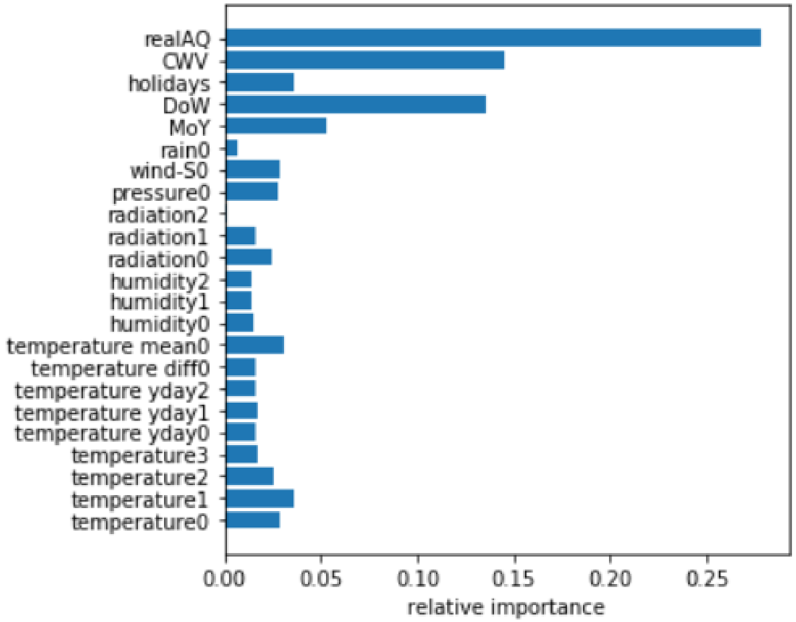
Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	It may be possible to improve the NDM demand estimation model through the use of machine learning (ML). This may be because the dependence of gas usage on weather or AQ is complex and difficult to capture in a 'hand crafted' model. We also have additional weather data available which might help with the prediction. This data can easily be incorporated into a ML model. To test this hypothesis we will train ML models to predict NDM demand for domestic meters in EUC1. Apply the best model for domestic users to predict usage in EUC2 to see if the model can be easily extended to other sites.
Data Tree References	EUC, CWV

Findings Status	**Ongoing**
UIG Impact Peak Volatility %	**Ongoing**
UIG Impact Annual Average %	**Ongoing**
Confidence in Percentages	**Ongoing**

Findings	Approach to analysis
The resulting models generally did not perform as well as the current NDM models. The ML model performed slightly better over the winter but not as well during warmer weather. This may be because gas consumers in EUC2 are inherently more diverse and hard to predict, or because the model weather fit parameters are tailored too closely to EUC1 usage. ***What does this mean?:*** *More work would have to be done to determine if an ML model could produce improved predictions. Any improvements are likely to be similar to those seen in EUC 01. If we move NDM allocation to a ML based prediction, then different models may be required for each EUC.*	The best model that had been obtained for EUC1 was using XGBoost including the CWV. This was applied to EUC2 and error metrics were calculated. LDZs EA, EM and SC were used.

Supporting evidence 1: Results table for ML Model tested on EUC02 data – lower is better.

- **We then used the best XGBoost model parameters for training on EUC1 to train on EUC2 (with some modifications for the higher gas usage per meter in this EUC)**
- **The models produced did not result in better predictions than the current NDM model. This may be because the fit parameters were not carefully tuned for this EUC, or because there are more diverse gas users in this EUC, making usage harder to predict.**
- **As might be expected, day of the week is an important predictor of gas usage in EUC2 (compared to EUC1) as it contains industrial sites.**

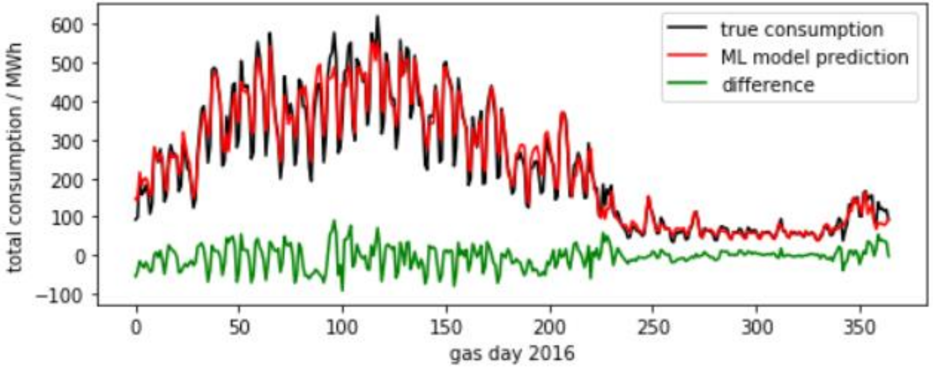


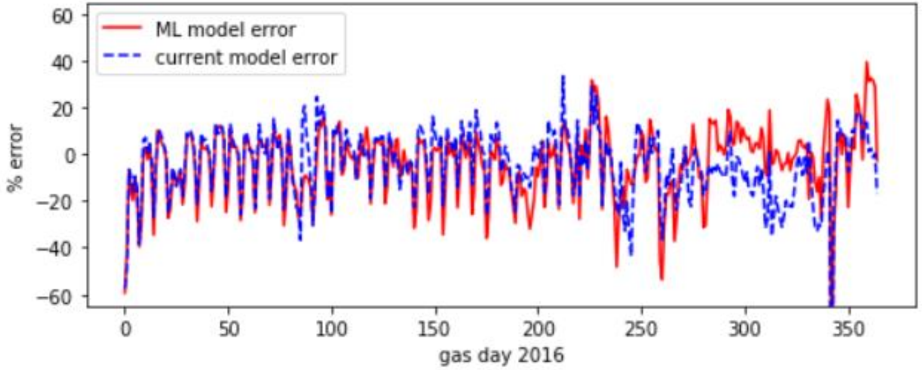
LDZ	Model	Mean daily error over year (kWh)	RMS error on total daily usage (kWh)	Mean day-day total abs error diff (kWh)	StD. on day-day total error (kWh)
EA	current	1,827	10,467	7,442	10,306
	XGB 12-15, w17 B	-853	13,256	10,653	13,229
SC	current	551	5,722	4,030	5,696
	XGB 12-15, w17 B	1,564	7,811	6,599	7,653
EM	current	1,100	29,787	21,788	29,767
	XGB 12-15, w17 B	-4,360	29,024	21,817	28,696

Supporting evidence 2: Plots of EUC 02 ML results (continued on next page)

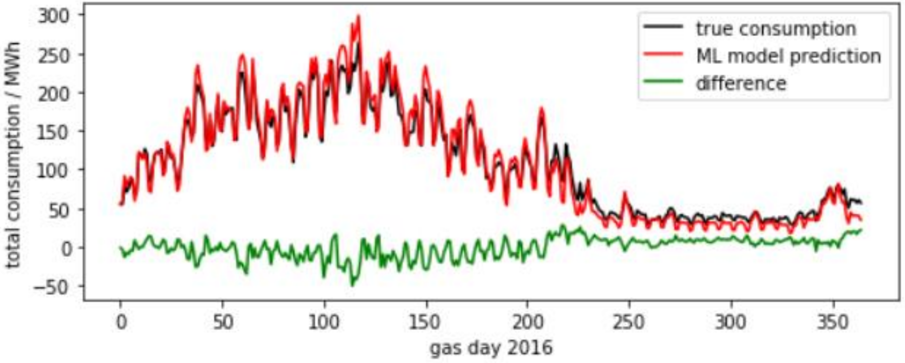
- For the best model (trained on all available data and using CWV as a feature (model B), the results are plotted here for 3 LDZs for the 2016 training year.
- The errors on the ML model and the current NMD model are also plotted.

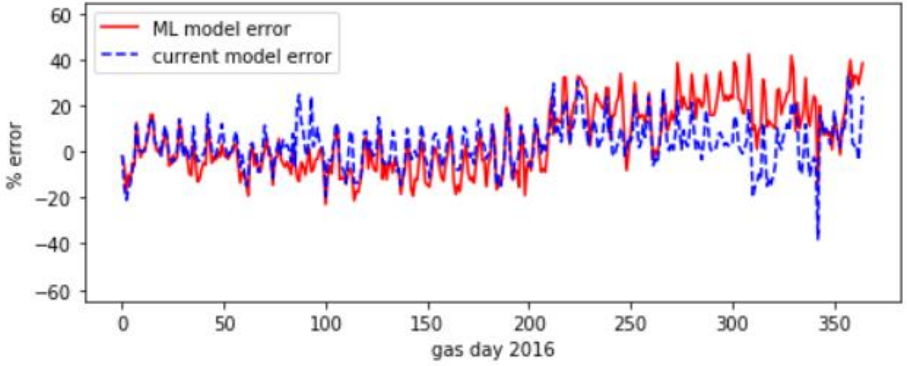
EM, XGB 12-15 including Winter 2017 B



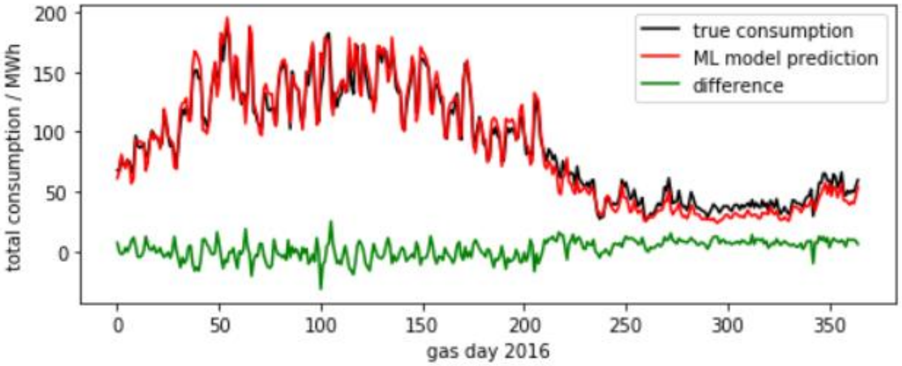


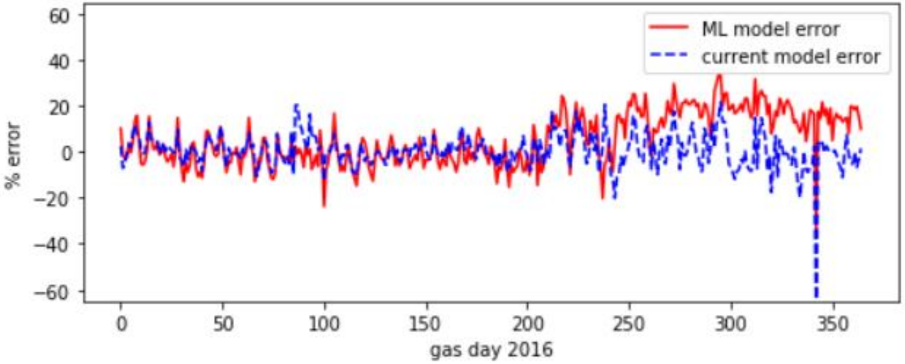
EA, XGB 12-15 Winter 2017 B





SC, XGB 12-15 Winter 2017 B





Summary of Findings: Machine Learning EUC01 with Neural Network

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	It may be possible to improve the NDM demand estimation model through the use of machine learning (ML). This may be because the dependence of gas usage on weather or AQ is complex and difficult to capture in a 'hand crafted' model. We also have additional weather data available which might help with the prediction. This data can easily be incorporated into a ML model. To test this hypothesis we will train ML models to predict NDM demand for domestic meters in EUC1. After initially obtaining modestly improved results using XGBoost for machine learning, it was determined that it would be worthwhile to test using a different ML ensemble. Neural Networks (NN) are scalable, all purpose algorithms that have been successfully applied to a wide variety of time-series prediction problems, such as predicting Gas usage over time. Therefore ML using an NN was tested. It was also thought that the NN may interpolate and extrapolate better than XGB, and so produce better outputs in extreme weather scenarios.
Data Tree References	CWV, Weather

Findings Status	**Ongoing**
UIG Impact Peak Volatility %	**15-20% but can make it worse**
UIG Impact Annual Average %	**10-25%**
Confidence in Percentages	**N/A**

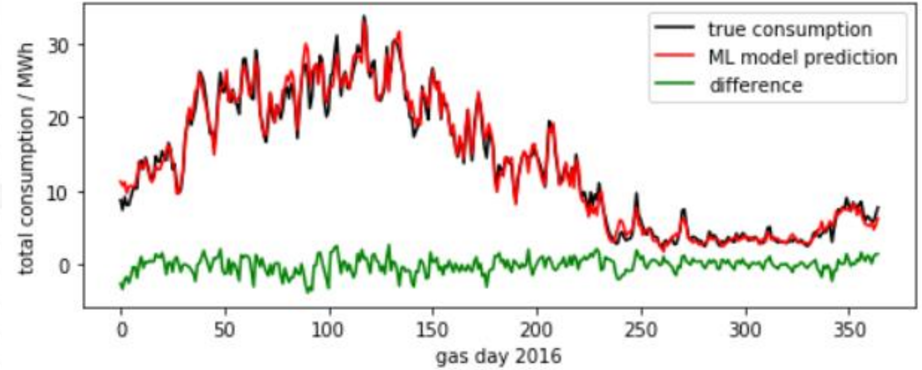
Findings	Approach to analysis
The resulting models performed similarly to the current best XGB model on the test year. However models for all three of the test LDZs exhibited occurrences of large percentage error (but relatively small absolute error) around gas day 240. It is possible that further improvements could be made to this model by using more advanced regularisation techniques in the fit, or engineering of the inputs (processing of the raw data). The addition of more historic weather data is likely to be beneficial given the seasonal difference observed in the performance of the model. ***What does this mean:*** *Neural networks might still offer future performance improvements vs. XGBoost. Either more engineering/pre-processing of the inputs to the NN would be required, or more advanced regularisation techniques for the fit would be required. With the addition of more data it may be the case that the NN would improve.*	Missing weather data was interpolated and flag/indicator parameters were encoded such that they could be fed into Neural Networks. An appropriate width and depth of the Neural Network was determined by varying these parameters and comparing the outputs. Different inputs and regularisation techniques were tested. The best models are plotted here. Additionally, fitting the whole country with a single model was tested.

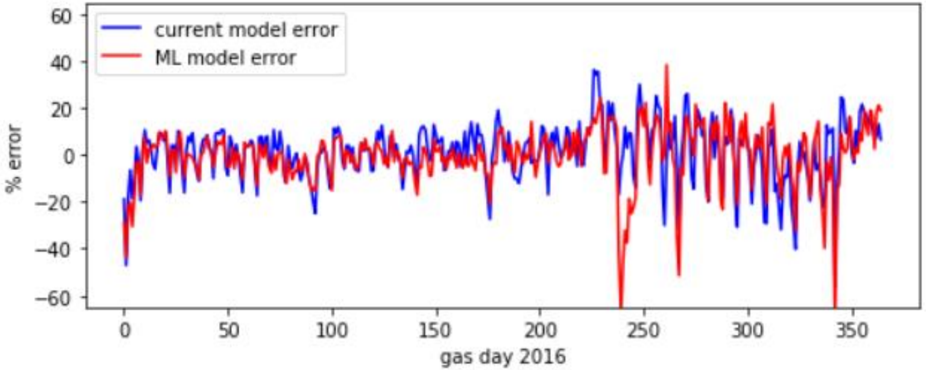
Supporting evidence 1: Plots of ML results. Lower is better.

- For the best model (trained on all available data and using CWV as a feature (model B), the results are plotted here for 3 LDZs (EM, EA & SC) for the 2016 training year.
- The errors on the Neural Network model (ML model) and the current NDM model are also plotted.
- Gas Day 240 shows worse results in the ML Model. This day is late May Bank Holiday (29/05/2017), suggesting that the model does not have enough historic data on the relationship between weather, demand and bank holidays to predict accurately.

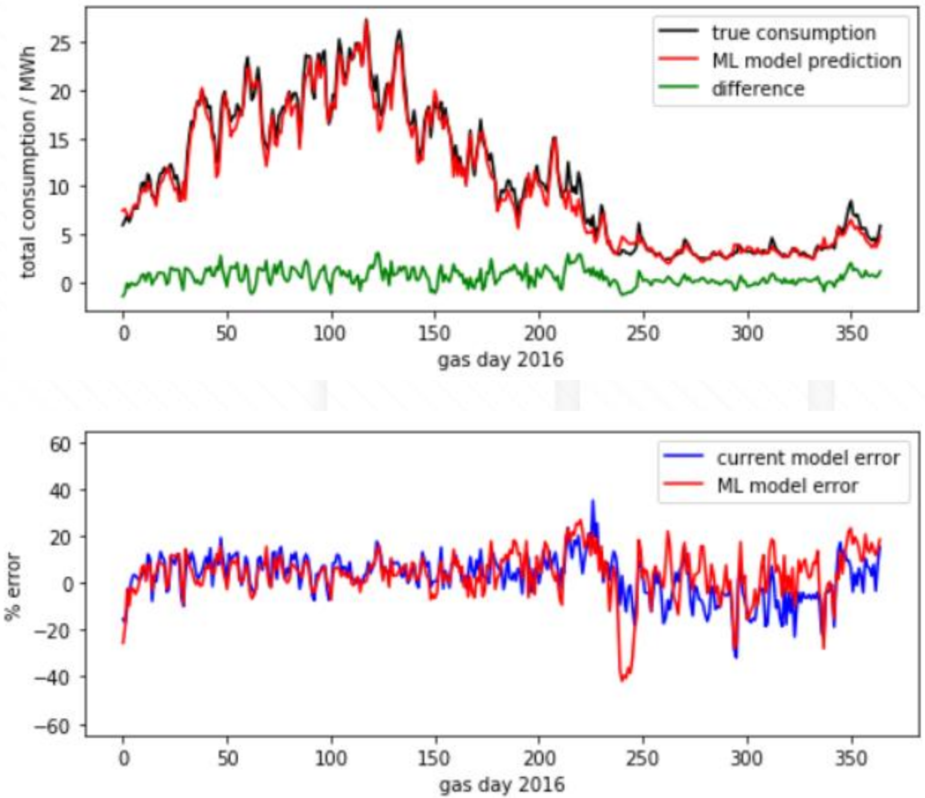
LDZ	Model	Mean daily error (kWh)	RMS error on total daily usage (kWh)	StD. on total daily error (kWh)	Mean (abs) day-to-day change in error (kWh)
EA	current	570	1,086	924	791
	Neural Network 12-15, w17 B	441	938	828	722
	XGB 12-15, w17 B	561	940	754	715
SC	current	283	711	652	537
	Neural Network 12-15, w17 B	163	569	545	431
	XGB 12-15, w17 B	214	590	550	440
EM	current	226	1,378	1,359	1,051
	Neural Network 12-15, w17 B	-56	1,082	1,080	826
	XGB 12-15, w17 B	40	1,004	1,004	754

EM, Neural Network 12-15 Winter 17 B

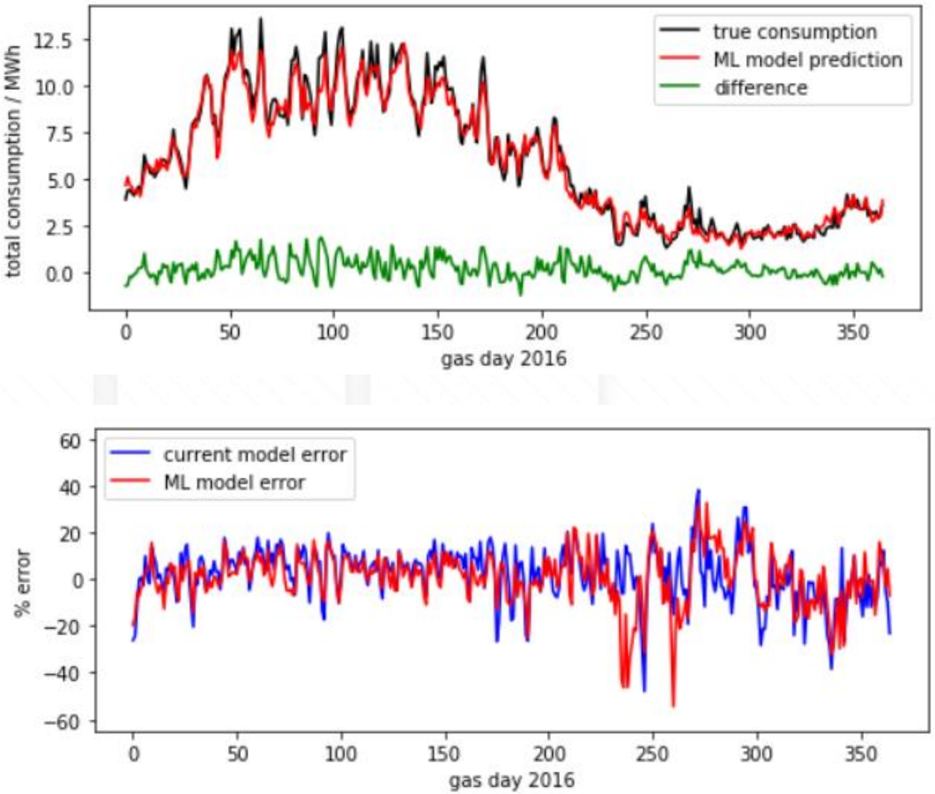




EA, Neural Network 12-15 Winter 17 B



SC, Neural Network, 12-15 Winter 17 B



Summary of Findings: Test XGBoost model on full LDZ EUC01 AQ

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	In previous analysis, we were able to show that a ML model for NDM consumption in EUC1 meters applied to LDZs EA, EM and SC resulted in an improved prediction for gas consumption of the sample set meters. However, from these results it was not possible to be sure that the ML improvement would be seen when the ML model was applied to predict consumption of all EUC1 meters in an LDZ, because many of the meters in the training years were the same as those in the test year. Nor could it be determined how this improvement would scale for the prediction the net consumption of a couple of million meters versus the few hundred in the sample set test year. The aim of this analysis is to answer these questions by using ML models previously derived to predict the usage of all of the NDM meters in the LDZs. Four error metrics were calculated for each LDZ prediction. The ‘mean daily error’ is the cumulative error over a year divided by the number of days in a year, and is a measure of the bias of a model. The standard deviation of the daily error, and the mean absolute day-to-day change in error are measures of volatility. The RMS error is a measure of the typical prediction error that would be expected from a model in any given day (which could be positive or negative).
Data Tree References	Energy, EUC, annual quantity

Findings Status	
UIG Impact Peak Volatility %	**20%**
UIG Impact Annual Average %	**25%**
Confidence in Percentages	**+/-5% standard deviation (varies by LDZ)**

Findings	Approach to analysis
Decision tree XGBoost ML models were generated for 12 LDZs (a model for WN could not be generated due to the limited number of sample Meter Points). All of the predictions made using the ML model for EUC01 outperform the current NDM model except for one error metric in SC. The mean daily error in the EUC01-ML model is significantly lower than that of the current NDM model, by the order 50%, which means that this model would result in a smaller UIG accumulation over time. The RMS error is reduced by ~25%, which can be interpreted as a 25% reduction in prediction error (volatility) each day. This is strong evidence that ML can improve the NDM model at least for EUC01.	Daily AQ profiles values were derived from UK-Link AQ data. Results for the daily gas consumption using XGBoost models were then calculated for 5% of the meters in the LDZ. The summed daily usage of this 5% of meters was then scaled to the full LDZ using the ratio of total AQs in the 5% sample to the total EUC01 AQ in the LDZ. For the other EUCs, predicted consumption was computed using the current NDM model WAALPs. The DM energy was added. This total predicted consumption for the LDZ was then compared with the ‘true consumption’ (calc input energy, stock change, shrinkage). Error metrics for the current NDM model and the ML EUC01 model were calculated.

Supporting evidence 1:

- Test year is gas year 2016 with system AQs. Trained with pseudo-AQs (one per meter, gas year) calculated from real usage and WAALPs
- Model B: uses AQ, holiday indicators, day of week, month of year, and a set of raw weather inputs including, a temperature gradient feature, temperatures from the previous day and the mean temperature of the past 3 days, and the CWV. The model was trained on
    - Years 2012 to 2015 and winter 2017 (XGB 12-15 w17)
- We considered several different error metrics of base error and volatility error, but the model was trained to minimise the daily usage error on each meter.
- Mean daily error is a measure of base error, RMS error on total daily usage is a combine measure of base and volatility, standard deviation (StD) on total daily error is a measure of the spread of the error values, and therefore a measure of volatility, mean absolute day-to-day change in error is also a measure of volatility.

- The ML algorithm has a seasonal performance profile – it is better than the NDM algorithm in the inter but doesn't perform as well during the summer. This can be clearly observed on the following slides.

Summary

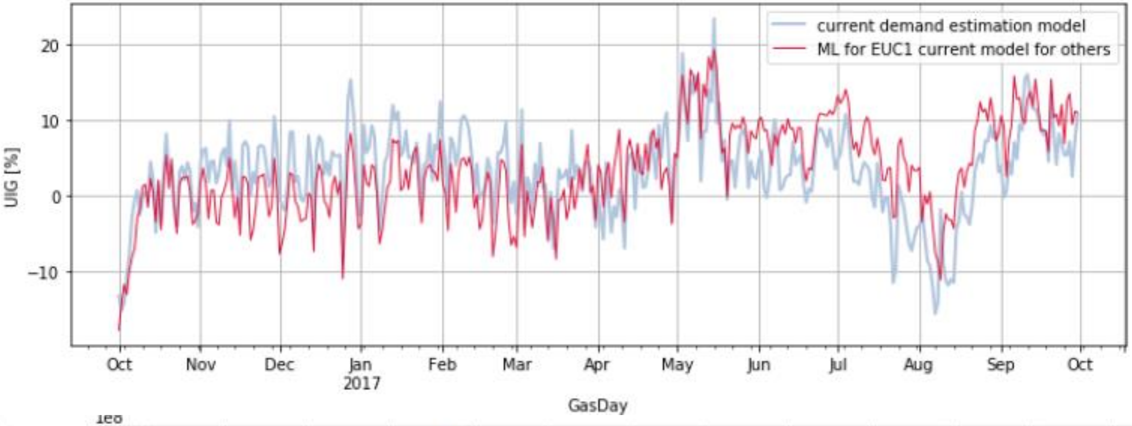
N.B. WN is missing as there was not enough data to fit the ML model

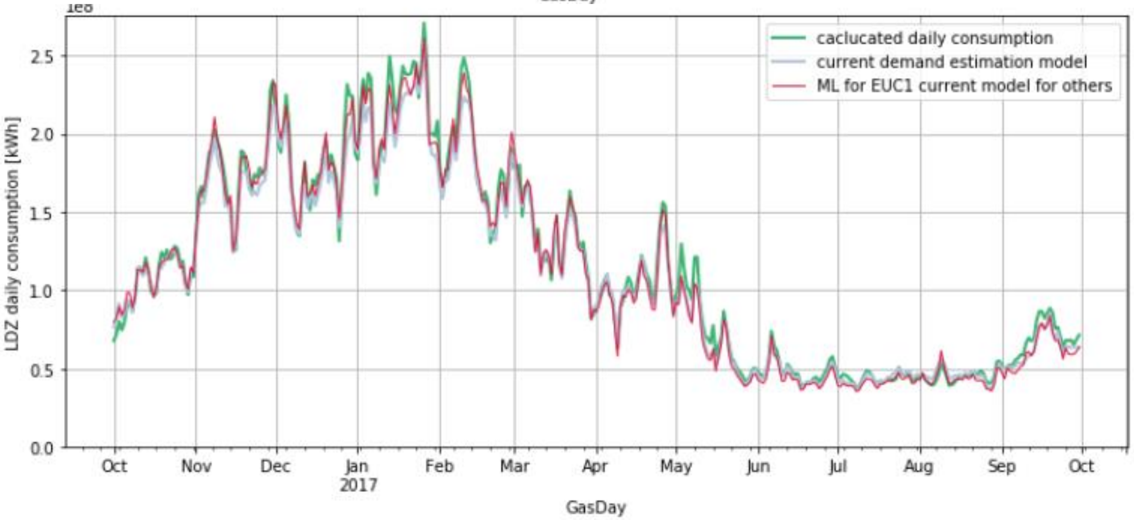
LDZ	Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
EA	current	4.58	8.59	7.26	4.34
	XGB 12-15, w17 B	2.61	6.49	5.94	3.60
	% reduction	43%	24%	18%	17%
EM	current	5.36	10.09	8.55	6.13
	XGB 12-15, w17 B	3.48	7.25	6.36	4.54
	% reduction	35%	28%	26%	26%
NE	current	3.57	7.12	6.16	4.23
	XGB 12-15, w17 B	1.44	5.11	4.90	3.45
	% reduction	60%	28%	21%	18%
NO	current	3.56	5.77	4.54	3.52
	XGB 12-15, w17 B	2.39	4.41	3.71	2.62
	% reduction	33%	24%	18%	25%
NT	current	6.46	9.65	7.17	4.50
	XGB 12-15, w17 B	5.62	8.08	5.81	3.46
	% reduction	13%	16%	19%	23%
NW	current	8.46	13.79	10.89	7.51
	XGB 12-15, w17 B	5.32	10.13	8.62	6.0
	% reduction	37%	27%	21%	20%

LDZ	Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
SC	current	4.56	7.68	6.18	5.07
	XGB 12-15, w17 B	1.29	7.17	7.05	4.28
	% reduction	72%	7%	-14%	16%
SE	current	2.89	9.07	8.59	5.58
	XGB 12-15, w17 B	2.13	7.66	7.36	4.65
	% reduction	26%	16%	14%	17%
SO	current	2.63	7.24	6.75	4.22
	XGB 12-15, w17 B	0.11	5.06	5.06	3.23
	% reduction	95%	30%	25%	22%
SW	current	2.93	5.58	4.75	3.02
	XGB 12-15, w17 B	0.61	4.44	4.40	2.46
	% reduction	79%	20%	7%	18%
WM	current	5.52	8.88	6.95	4.90
	XGB 12-15, w17 B	3.09	6.46	5.67	3.66
	% reduction	44%	27%	18%	25%
WS	current	1.70	4.00	3.62	2.66
	XGB 12-15, w17 B	0.73	2.97	2.88	2.01
	% reduction	57%	26%	20%	24%

EA, XGB 12-15 w17 B

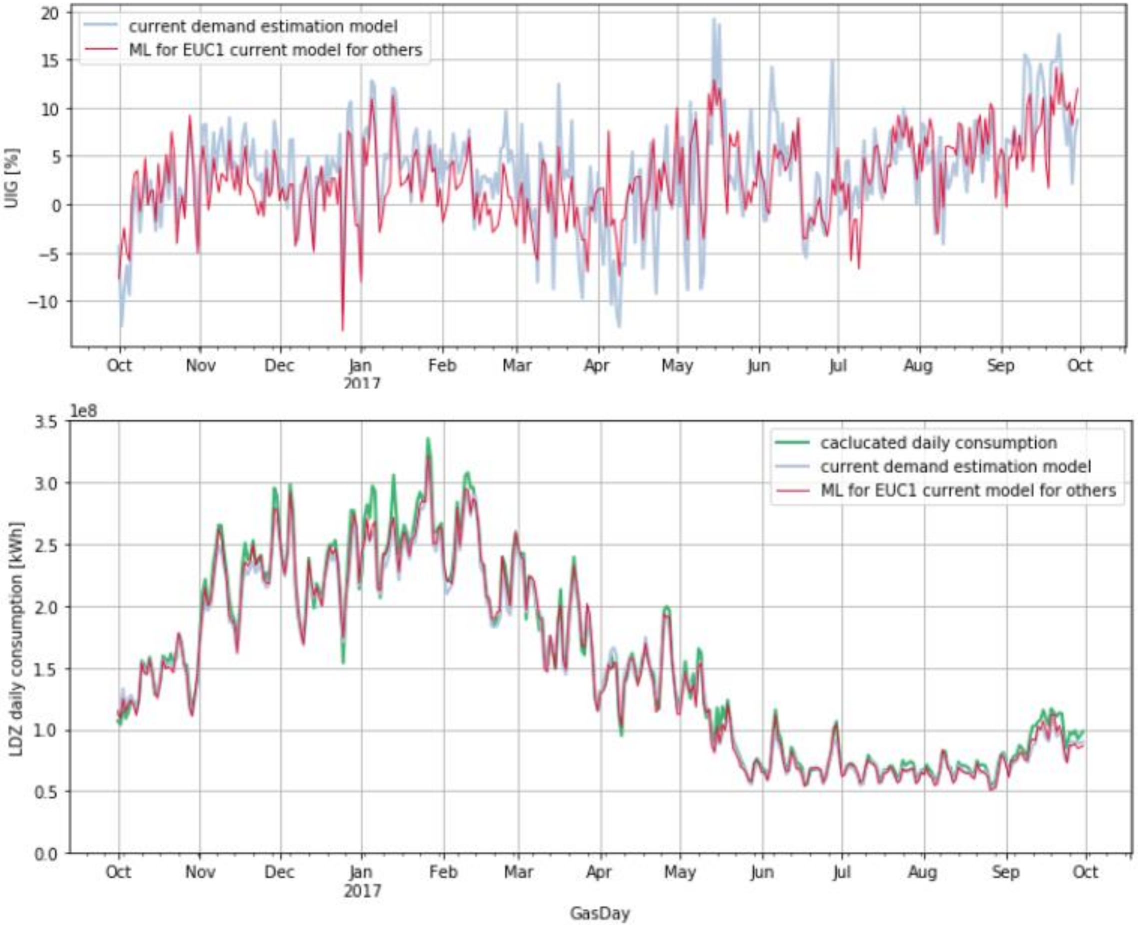
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	4.58	8.59	7.26	4.34
XGB 12-15, w17 B	2.61	6.49	5.94	3.60
% reduction	43%	24%	18%	17%





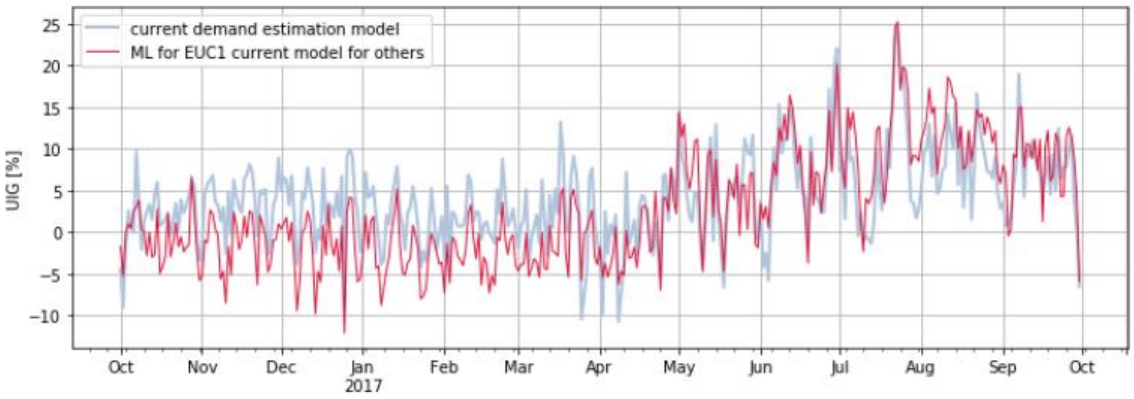
EM, XGB 12-15 w17 B

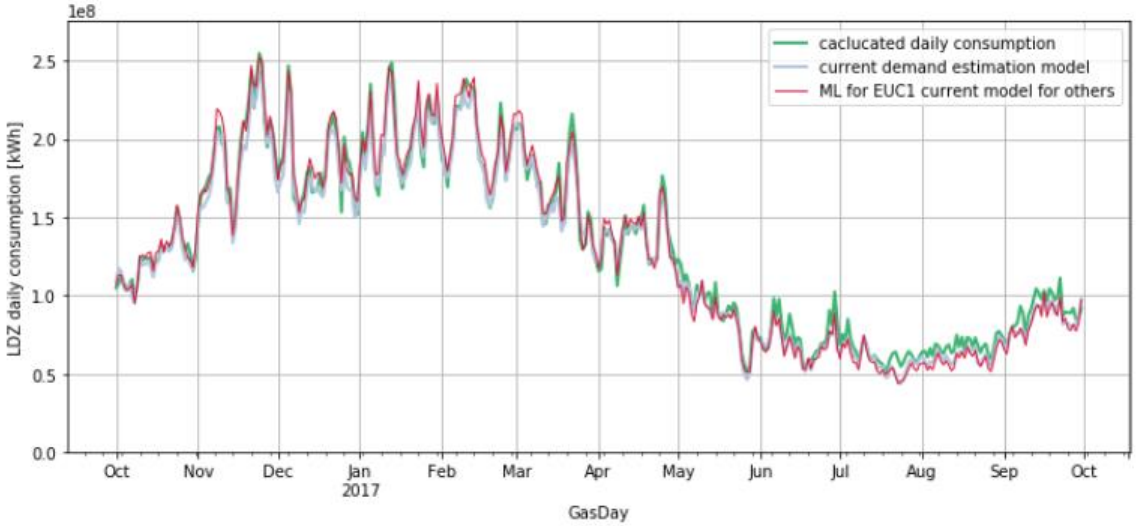
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	5.36	10.09	8.55	6.13
XGB 12-15, w17 B	3.48	7.25	6.36	4.54
% reduction	35%	28%	26%	26%



SC, XGB 12-15 w17 B

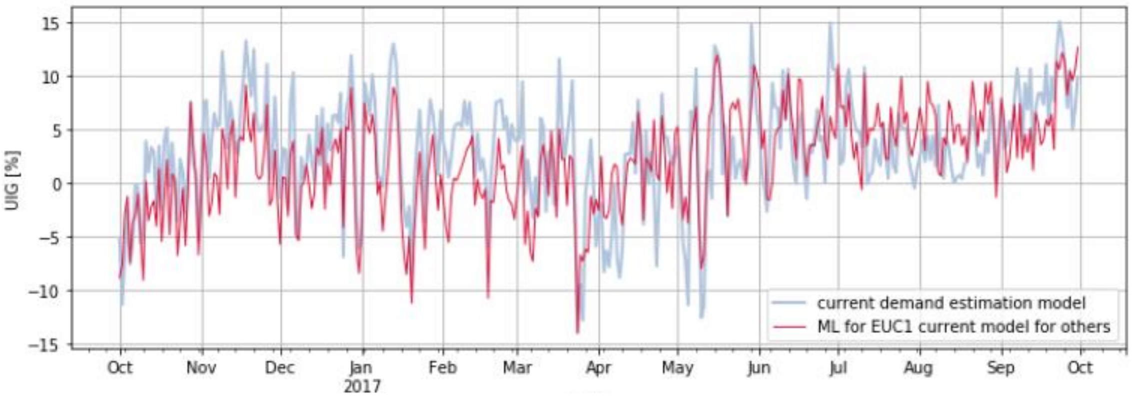
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	4.56	7.68	6.18	5.07
XGB 12-15, w17 B	1.29	7.17	7.05	4.28
% reduction	72%	7%	-14%	16%

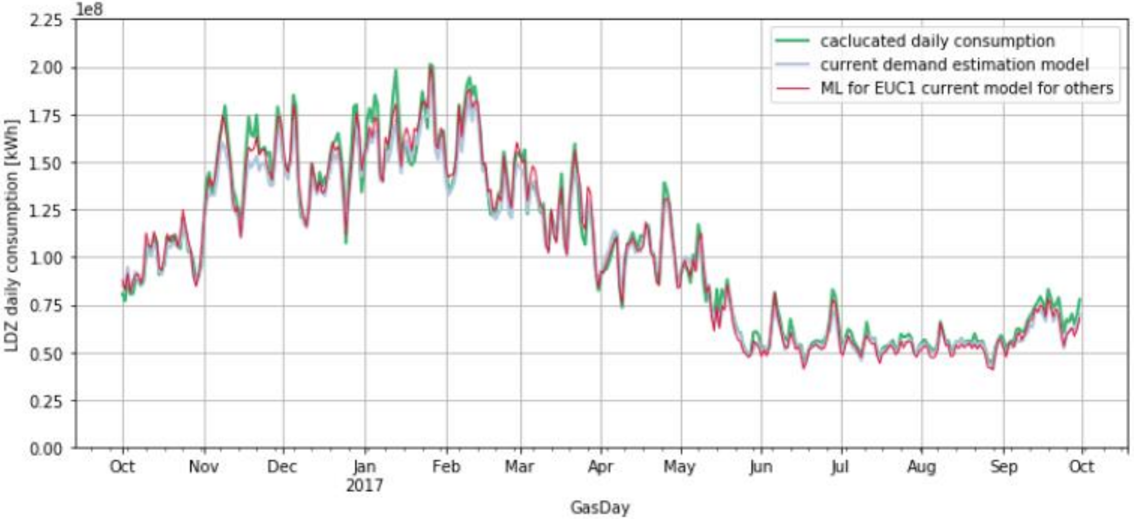




NE, XGB 12-15 w17 B

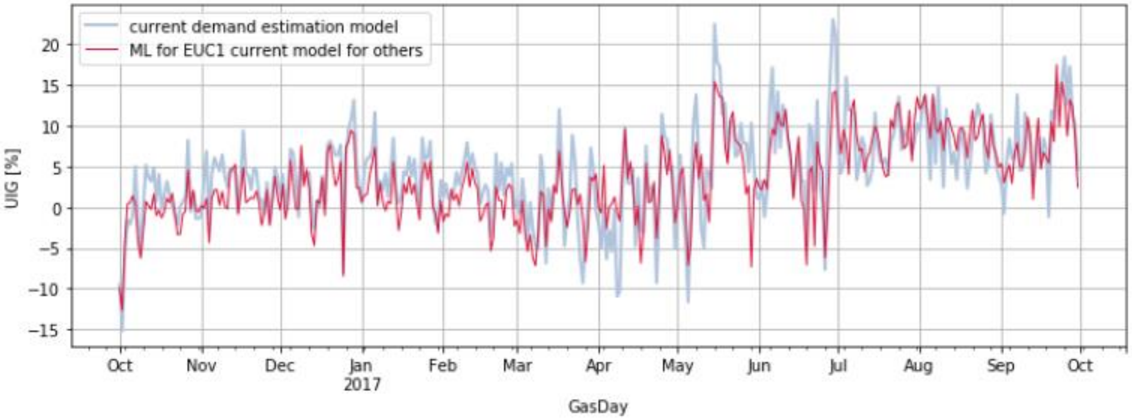
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	3.57	7.12	6.16	4.23
XGB 12-15, w17 B	1.44	5.11	4.90	3.45
% reduction	60%	28%	21%	18%

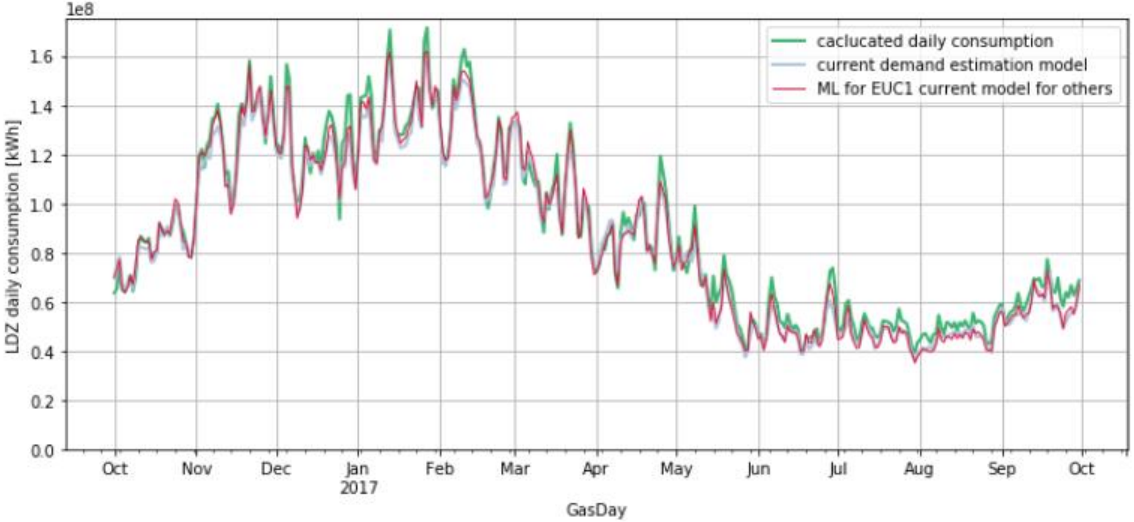




NO, XGB 12-15 w17 B

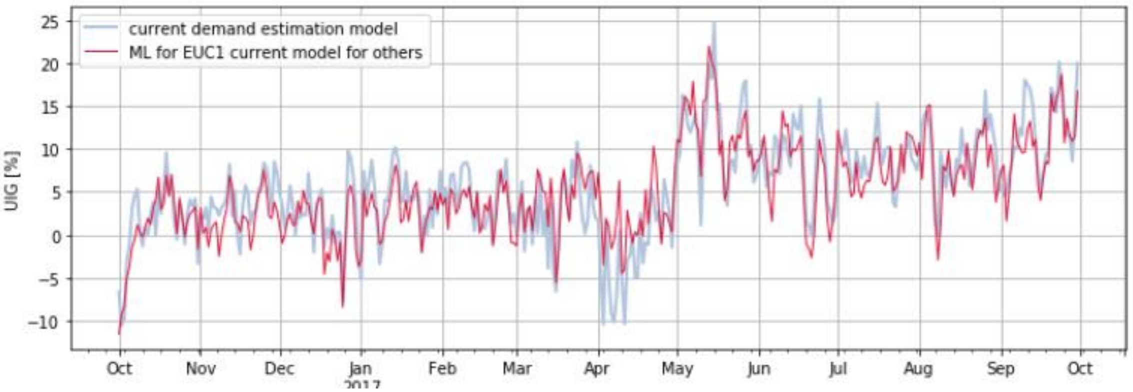
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	3.56	5.77	4.54	3.52
XGB 12-15, w17 B	2.39	4.41	3.71	2.62
% reduction	33%	24%	18%	25%

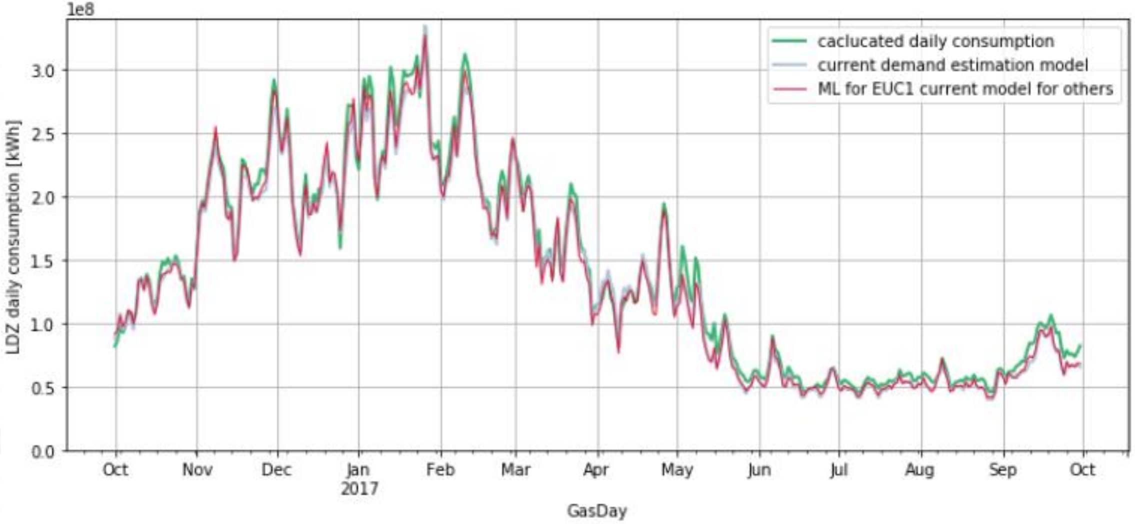




NT, XGB 12-15 w17 B

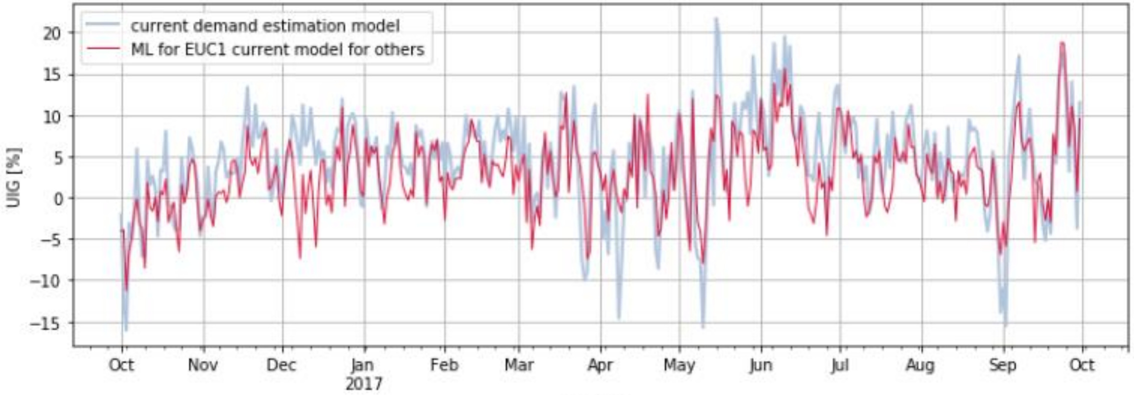
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	6.46	9.65	7.17	4.50
XGB 12-15, w17 B	5.62	8.08	5.81	3.46
% reduction	13%	16%	19%	23%

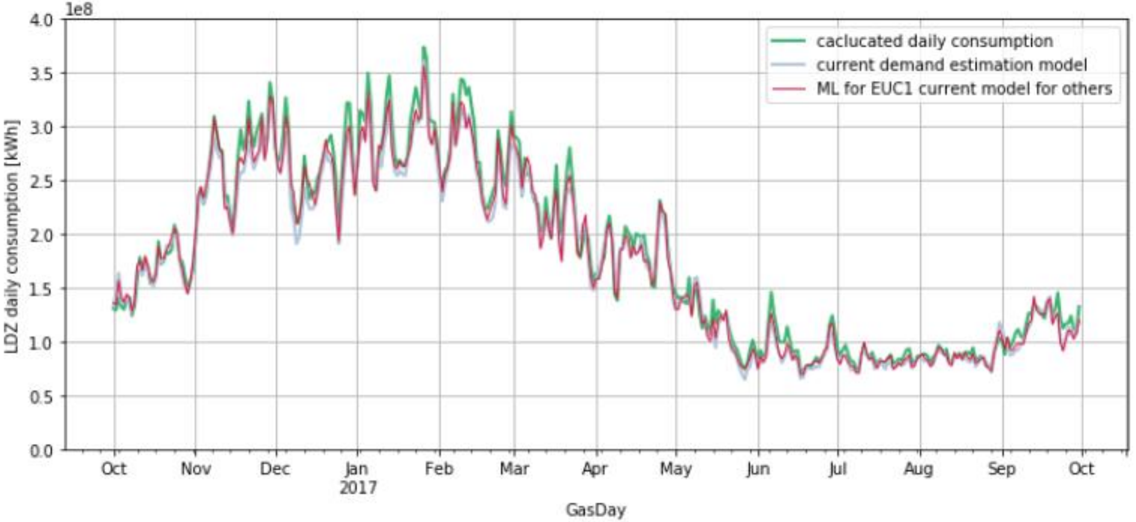




NW, XGB 12-15 w17 B

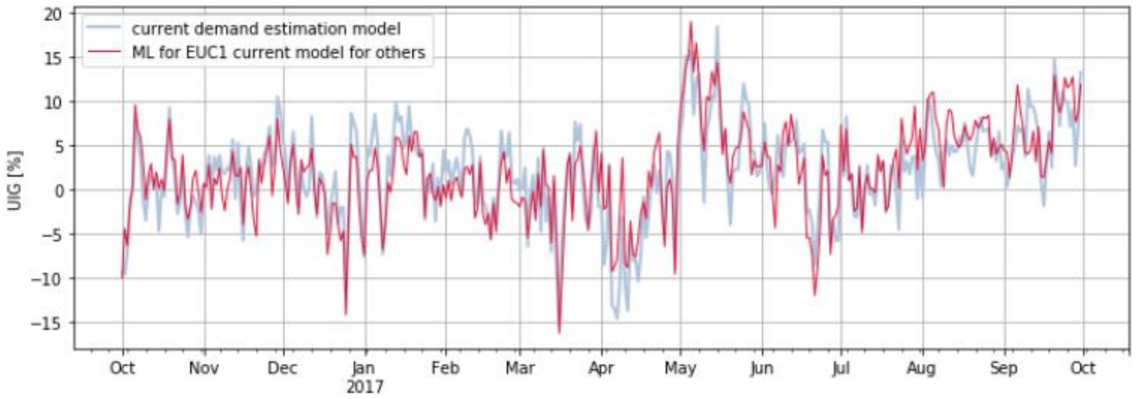
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	8.46	13.79	10.89	7.51
XGB 12-15, w17 B	5.32	10.13	8.62	6.0
% reduction	37%	27%	21%	20%

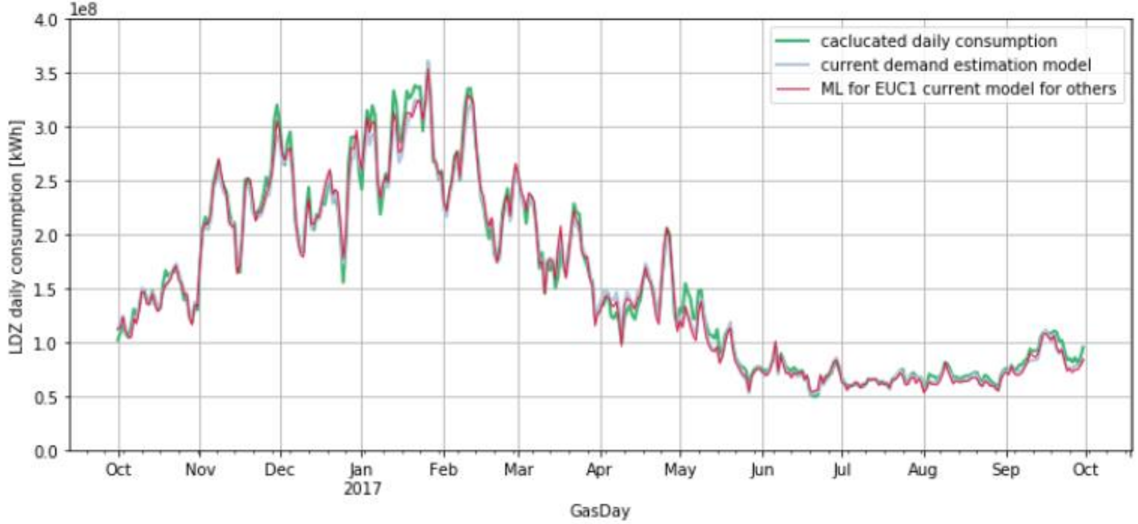




SE, XGB 12-15 w17 B

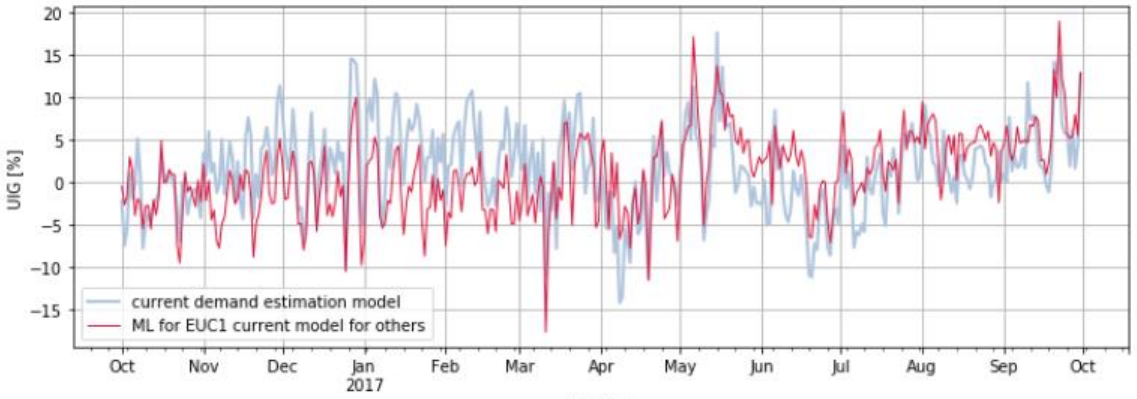
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.89	9.07	8.59	5.58
XGB 12-15, w17 B	2.13	7.66	7.36	4.65
% reduction	26%	16%	14%	17%

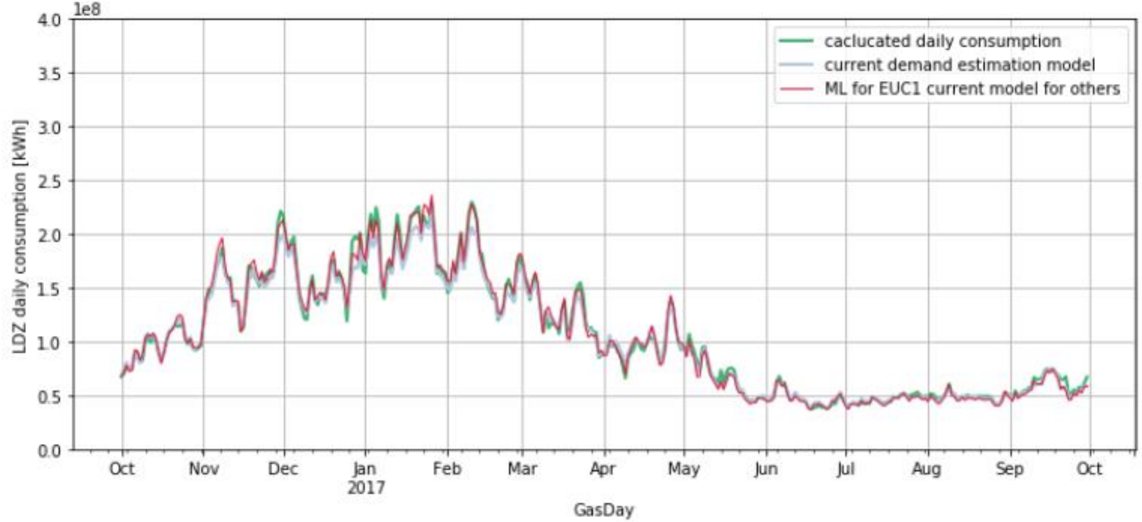




SO, XGB 12-15 w17 B

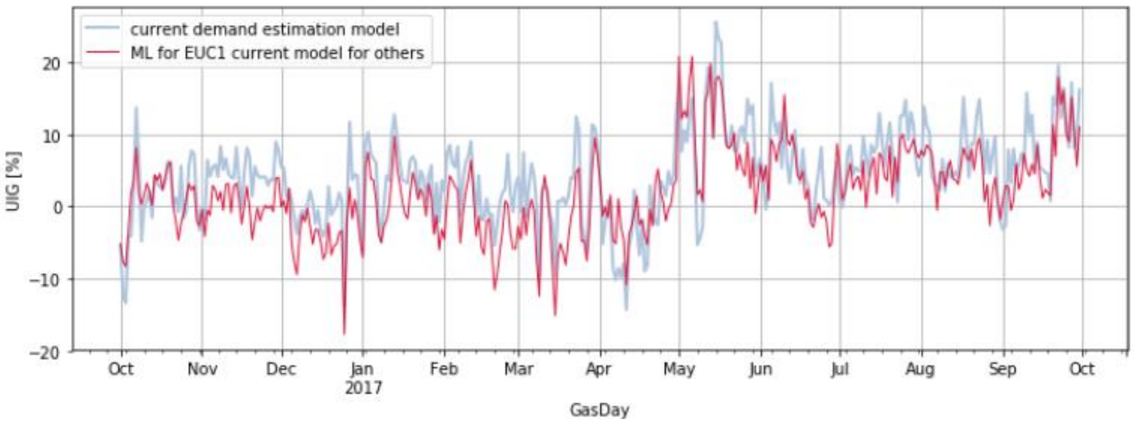
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.63	7.24	6.75	4.22
XGB 12-15, w17 B	0.11	5.06	5.06	3.23
% reduction	95%	30%	25%	22%

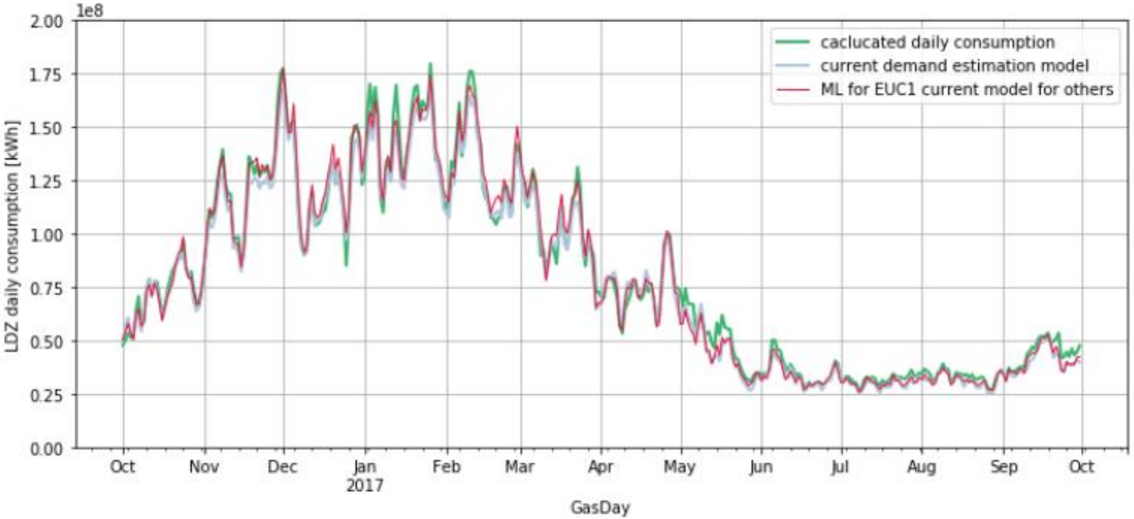




SW, XGB 12-15 w17 B

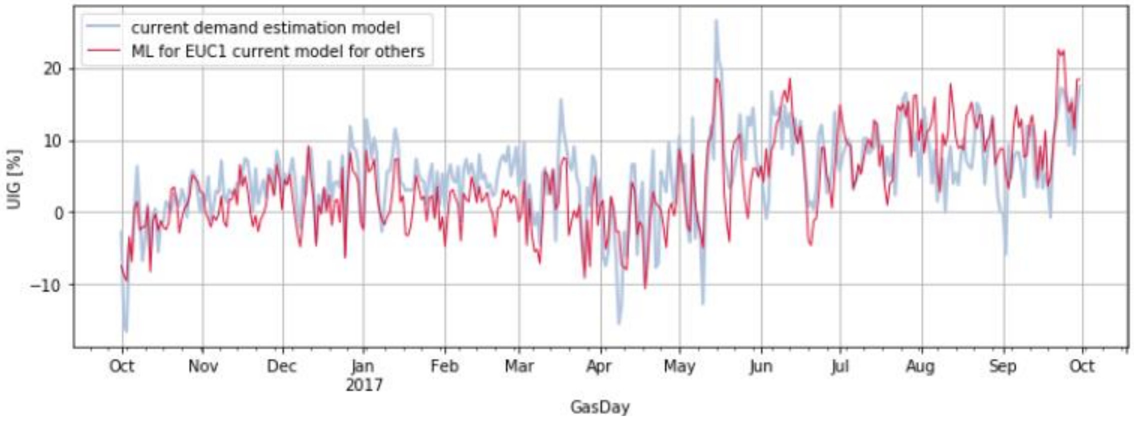
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.93	5.58	4.75	3.02
XGB 12-15, w17 B	0.61	4.44	4.40	2.46
% reduction	79%	20%	7%	18%

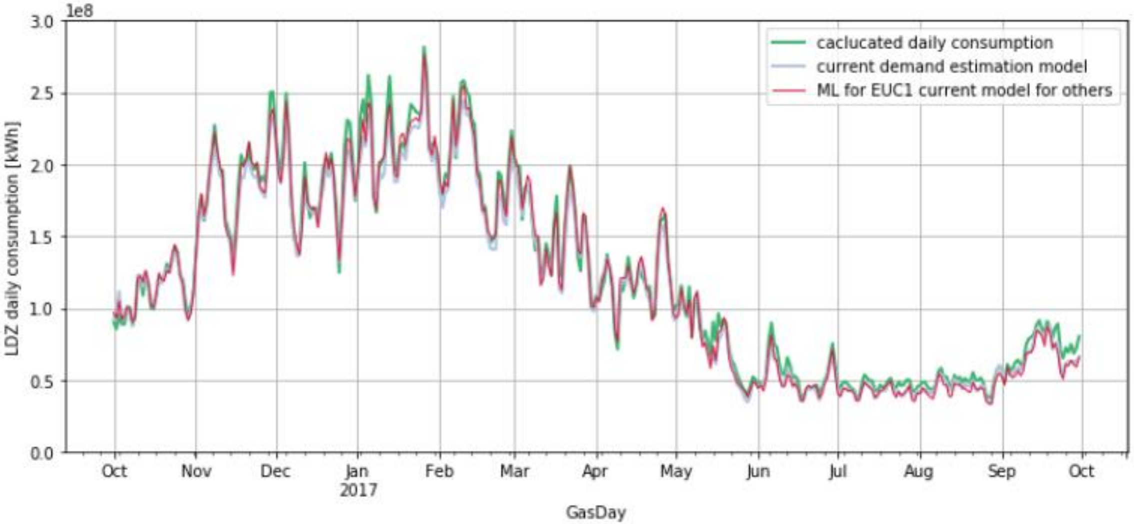




WM, XGB 12-15 w17 B

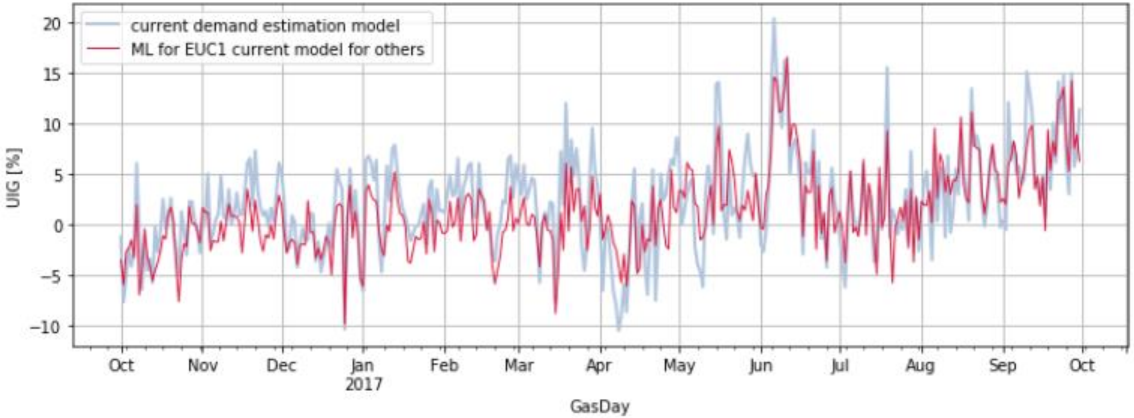
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	5.52	8.88	6.95	4.90
XGB 12-15, w17 B	3.09	6.46	5.67	3.66
% reduction	44%	27%	18%	25%

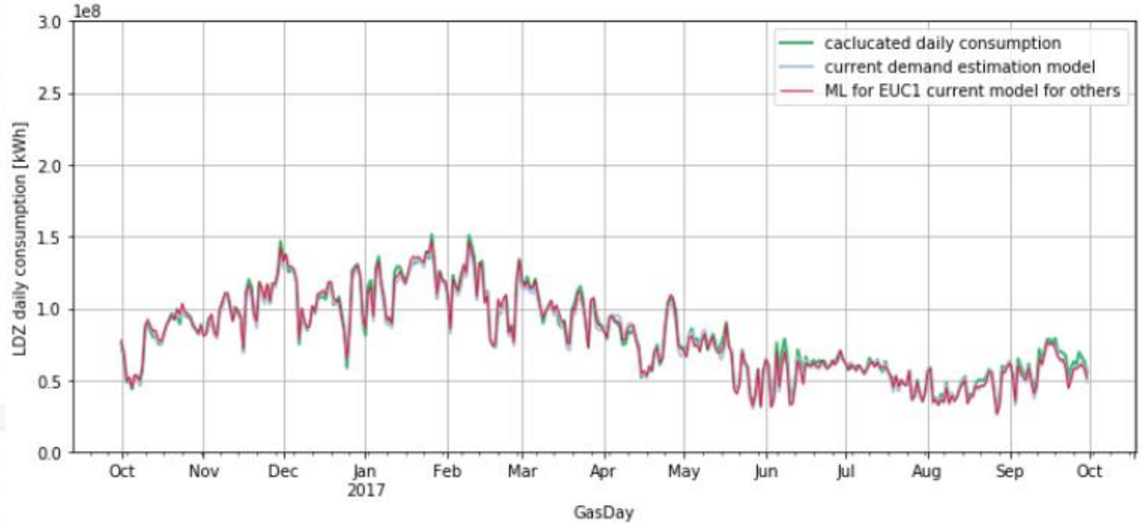




WS, XGB 12-15 w17 B

Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	1.70	4.00	3.62	2.66
XGB 12-15, w17 B	0.73	2.97	2.88	2.01
% reduction	57%	26%	20%	24%





Summary of Findings: Enhance Neural Network Model with additional Historic Data and test on full LDZ AQ

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis	In the previous analysis we used a ML model for EUC1 combined with the current NDM model for the remaining meters to predict total daily consumption for each LDZ. The ML model was built using XGBoost and trained on gas years 2012-2015 and winter 2017 (12-15, w17). Predictions were made for gas year 2016. This model showed improved performance over the current NDM algorithm, and reduced prediction errors by around 25% on average. Previously, we had trained a neural network (NN) as well as XGBoost on the same data. The performance of the neural network was similar to XGBoost. Typically, the performance of neural networks increases more than decision trees when more training data is provided. We therefore retrained the neural network with more historic weather data and sample set AQ data, from gas years 2006-2011 for EUC01. The trained models were evaluated on the whole LDZ using an augmented model with the current NDM algorithm being used for the other EUCs. Performance metrics were calculated and UIG plots produced.
Data Tree References	EUC, Energy, Annual Quantity, Weather

Findings Status	
UIG Impact Peak Volatility %	**20% reduction**
UIG Impact Annual Average %	**70% reduction**
Confidence in Percentages	**+/-5% standard deviation (varies by LDZ)**

Findings	Approach to analysis
NN ML models were generated for 12 LDZs (a model for WN could not be generated due to the limited number of sample meter points). All of the predictions made using the ML model for EUC01 out perform the current NDM model except for one error metric in SC. The mean daily error in the EUC01-ML model is significantly lower than that of the current NDM model, by the order 70%, which means that this model would result in a smaller base level UIG accumulation over time. The RMS error is reduced by ~30%, which can be interpreted as a 25% reduction in prediction error (volatility) each day. The error volatility of these model is similar or slightly increased over the XGBoost models trained in the previous analysis. The additional data appears to improve the performance of the model (no direct comparison between previous NDM model tested on sample set only).	Daily AQ profiles values were derived from UK-Link AQ data. Results for the daily gas consumption using NN models were then calculated for 5% of the meters in the LDZ. The summed daily usage of this 5% of meters was then scaled to the full LDZ using the ratio of total AQs in the 5% sample to the total AQ of EUC01 meters in the LDZ. For the other EUCs, predicted consumption was computed using the current NDM model WAALPs. The DM energy was added. This total predicted consumption for the LDZ was then compared with the ‘true consumption’ (calc input energy, stock change, shrinkage). Error metrics for the current NDM model and the ML EUC1 model were calculated.

Supporting evidence 1: Results table for analysis

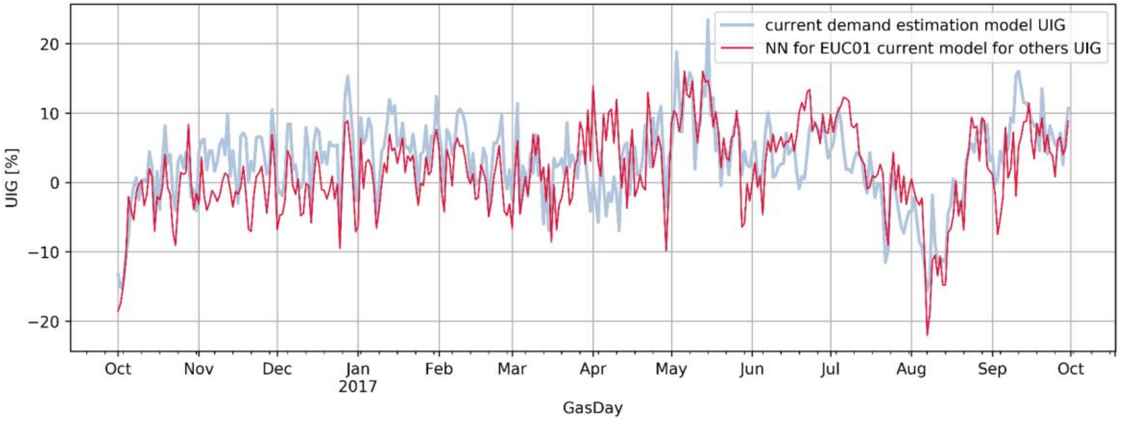
- Test year is gas year 2016 with UK Link AQs. Trained with pseudo-AQs (one per meter, gas year) calculated from real usage and WAALPs
- Model B: uses AQ, holiday indicators, day of week, month of year, and a set of raw weather inputs including, a temperature gradient feature, temperatures from the previous day and the mean temperature of the past 3 days, and the CWV. The model was trained on Years 2006 to 2015 and winter 2017 (NN 06-15 w17)
- We considered several different error metrics of base error and volatility error, but the model was trained to minimise the daily usage error on each meter.
- Mean daily error is a measure of base error, RMS error on total daily usage is a combine measure of base and volatility, standard deviation (StD) on total daily error is a measure of the spread of the error values, and therefore a measure of volatility, mean absolute day-to-day change in error is also a measure of volatility.
- The uppermost table compares the results of the neural network trained for ECU1 model, augmented with the current NDM model, with the results of the current NDM model only.
- The lower table compares the results of two identical neural network models, trained with different amounts of historic sample meter data, where the model NN 12-15, w17 B was trained with sample meter data dating from gas years 2012 to 2015 and the winter of 2017.

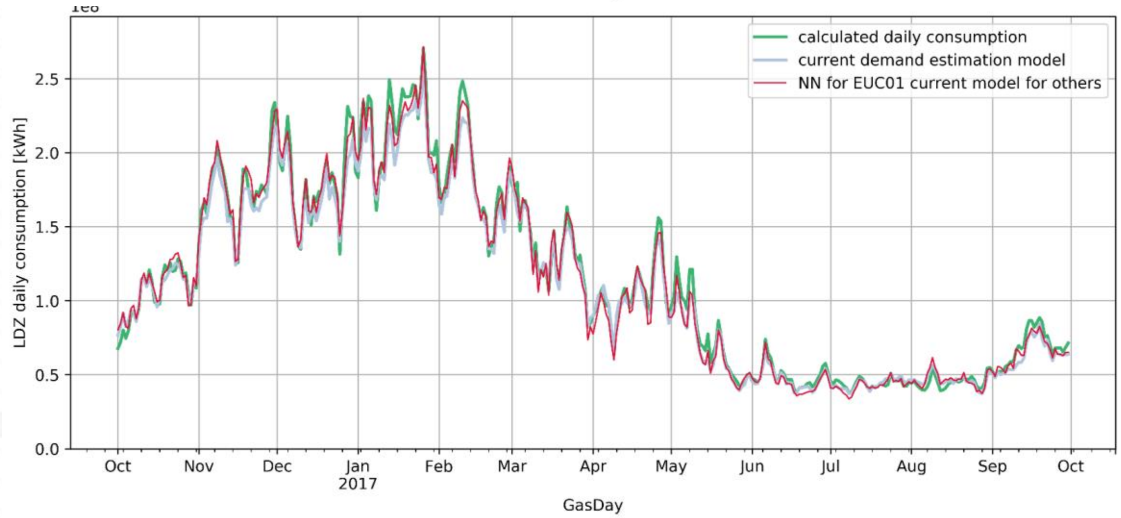
LDZ	Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
EA	current	4.58	8.59	7.26	4.34
	NN 06-15, w17 B	1.61	6.19	5.98	3.93
	% reduction	65%	28%	18%	9%
SC	current	4.56	7.68	6.18	5.07
	NN 06-15, w17 B	0.82	6.52	6.47	4.13
	% reduction	82%	15%	-5%	19%
EM	current	5.36	10.09	8.55	6.13
	NN 06-15, w17 B	0.77	6.20	6.15	4.57
	% reduction	85%	38%	28%	25%

LDZ	Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
EA	NN 12-15, w17 B	2.11	6.30	5.93	3.81
	NN 06-15, w17 B	1.61	6.19	5.98	3.93
SC	NN 12-15, w17 B	1.26	7.04	6.92	4.37
	NN 06-15, w17 B	0.82	6.52	6.47	4.13
EM	NN 12-15, w17 B	4.25	7.90	6.66	4.71
	NN 06-15, w17 B	0.77	6.20	6.15	4.57

EA, NN 06-15 w17 B

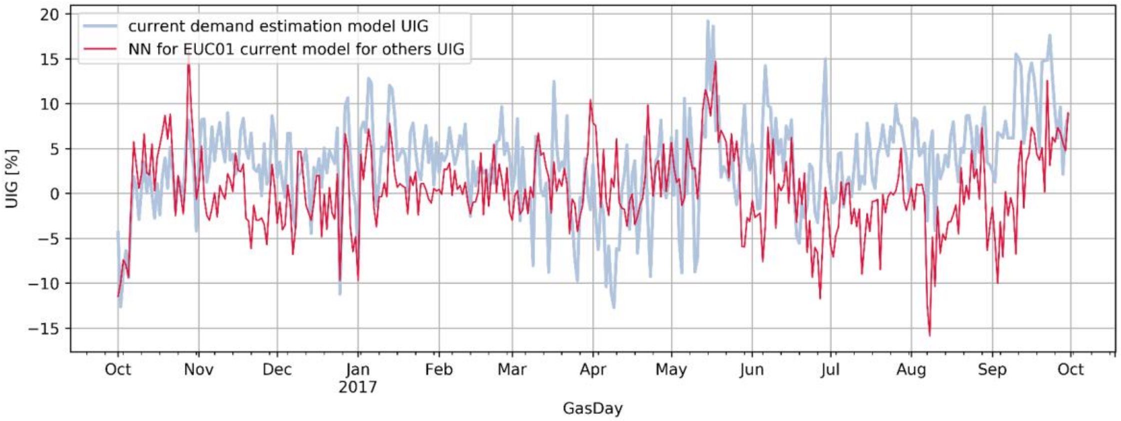
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	4.58	8.59	7.26	4.34
XGB 12-15, w17 B	2.61	6.49	5.94	3.60
NN 12-16, w17 B	2.11	6.30	5.93	3.81
NN 06-16, w17 B	1.61	6.19	5.98	3.93

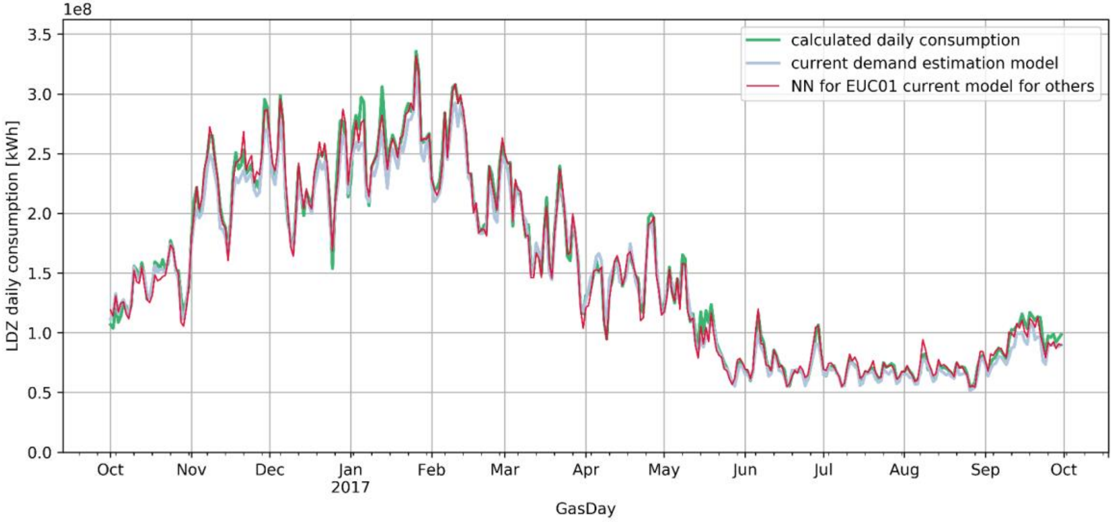




EM, NN 06-15 w17 B

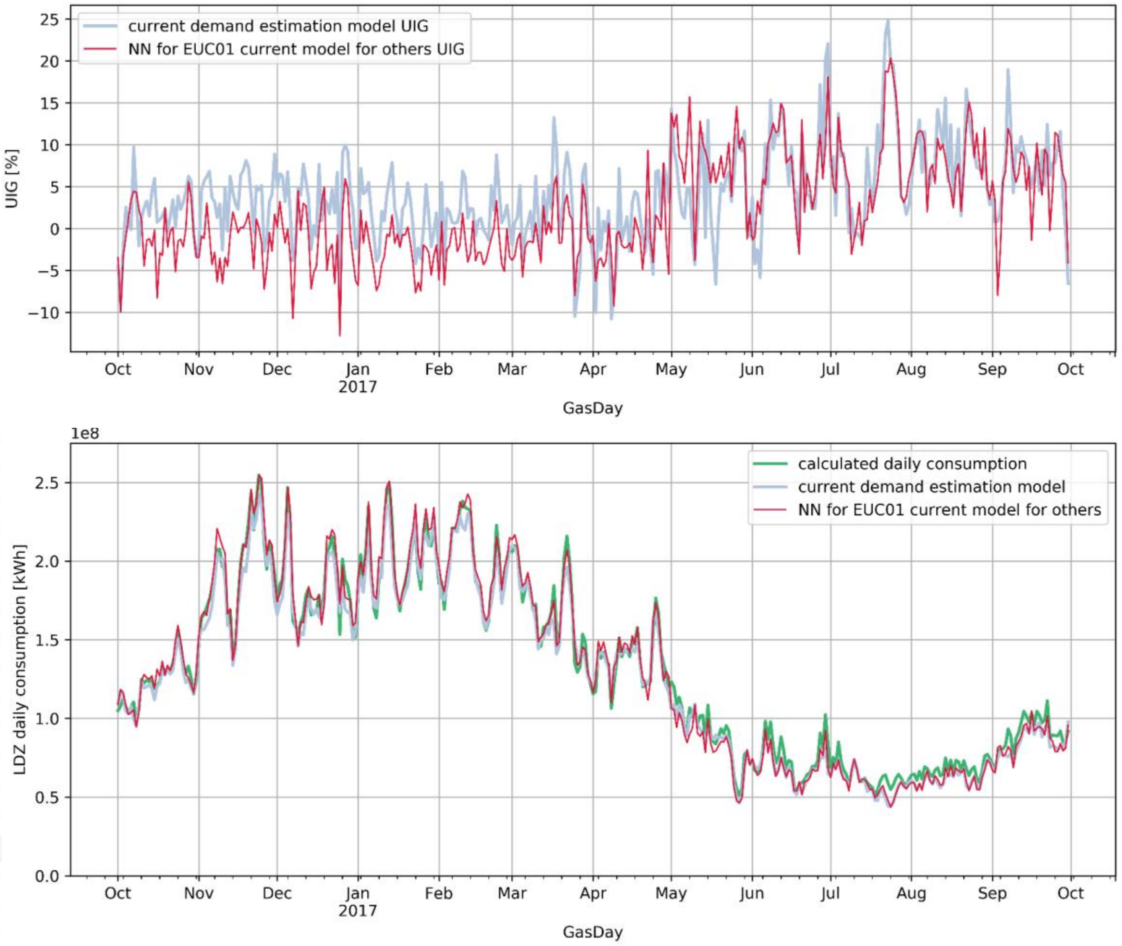
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	5.36	10.09	8.55	6.13
XGB 12-15, w17 B	3.48	7.25	6.36	4.54
NN 12-16, w17 B	4.25	7.90	6.66	4.71
NN 06-16, w17 B	0.77	6.20	6.15	4.57





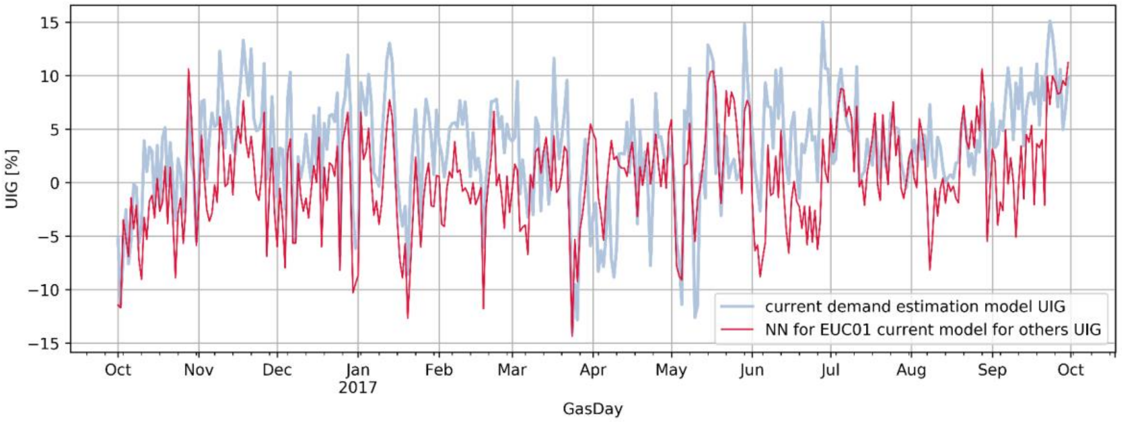
SC, NN 12-15 w17 B

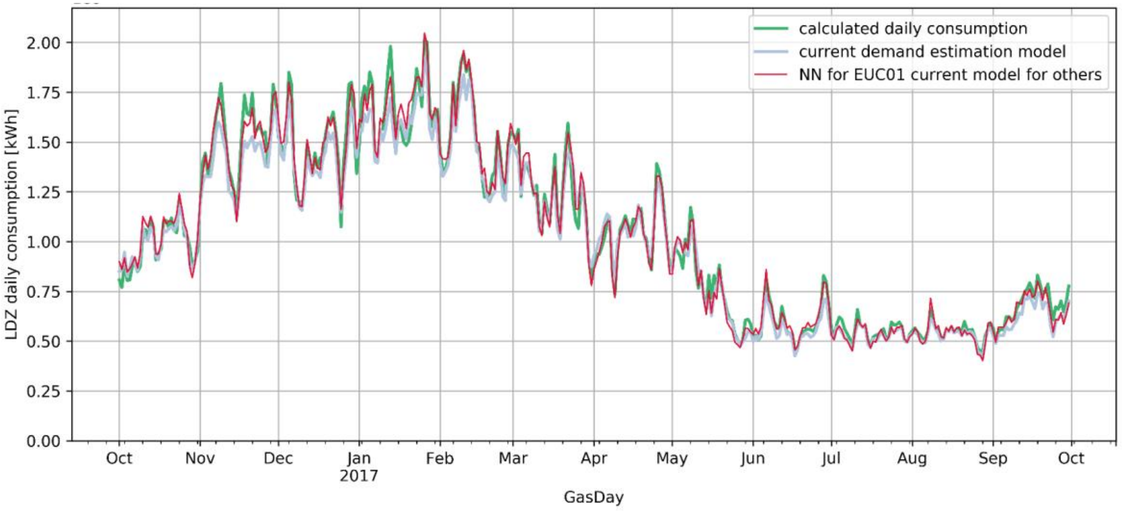
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	4.56	7.68	6.18	5.07
XGB 12-15, w17 B	1.29	7.17	7.05	4.28
NN 12-16, w17 B	1.26	7.04	6.92	4.37
NN 06-16, w17 B	0.82	6.52	6.47	4.13



NE, NN 12-15 w17 B

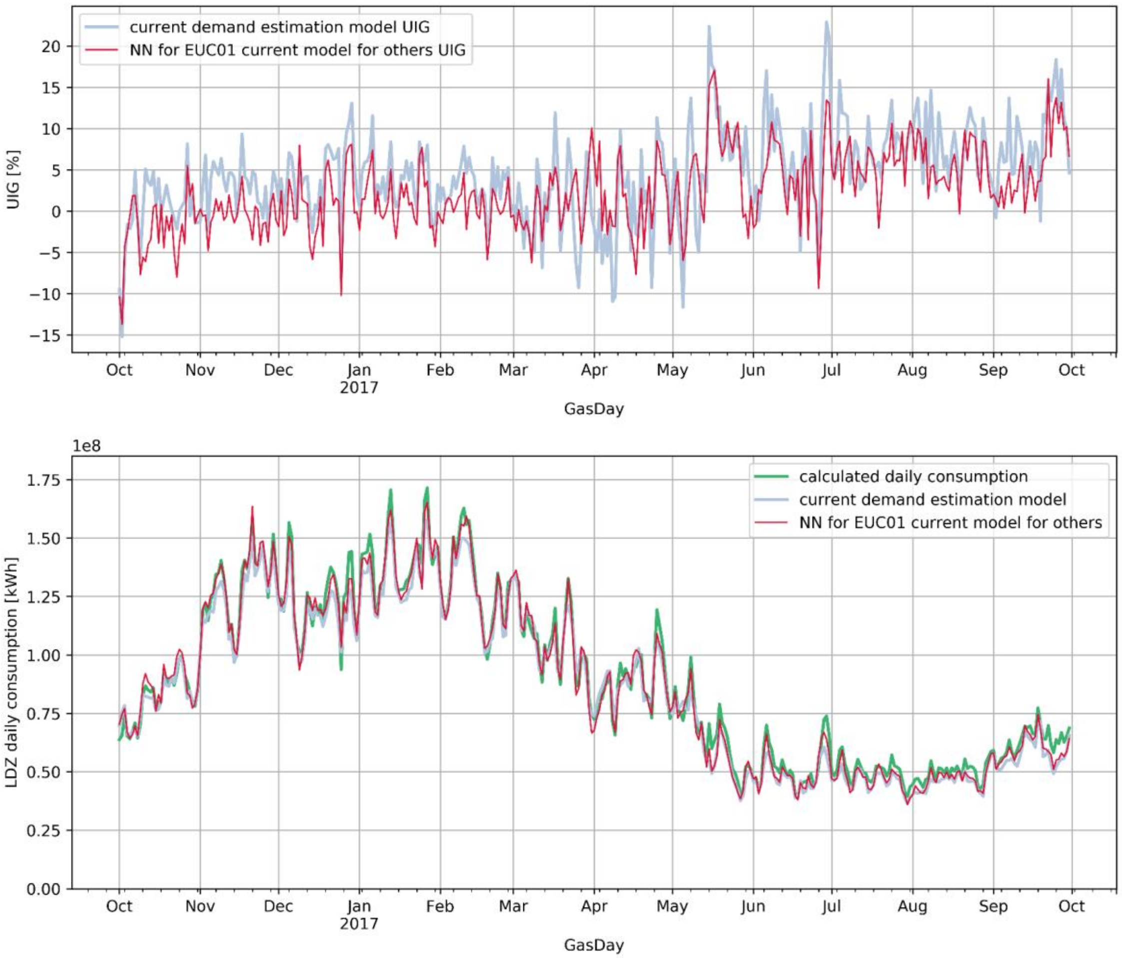
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	3.57	7.12	6.16	4.23
XGB 12-15, w17 B	1.44	5.11	4.90	3.45
NN 12-16, w17 B	0.79	5.19	5.13	3.72
NN 06-16, w17 B	0.08	4.86	4.86	3.55





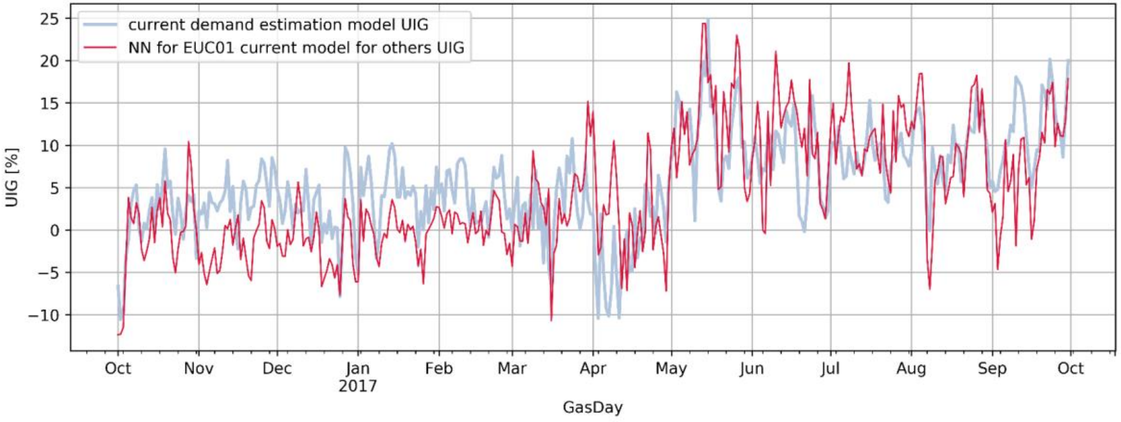
NO, NN 12-15 w17 B

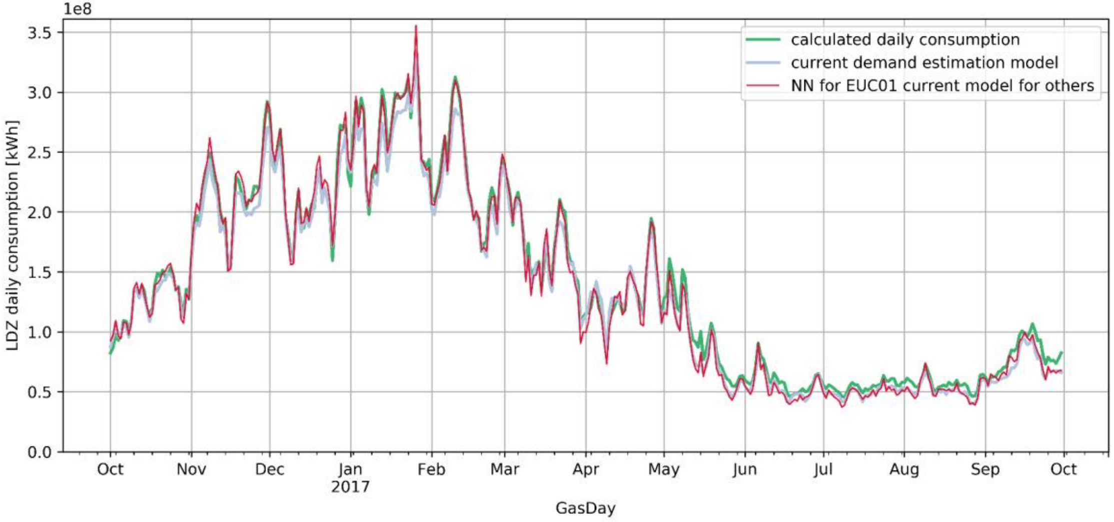
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	3.56	5.77	4.54	3.52
XGB 12-15, w17 B	2.39	4.41	3.71	2.62
NN 12-16, w17 B	2.46	4.50	3.76	2.69
NN 06-16, w17 B	1.54	3.98	3.67	2.65



NT, NN 12-15 w17 B

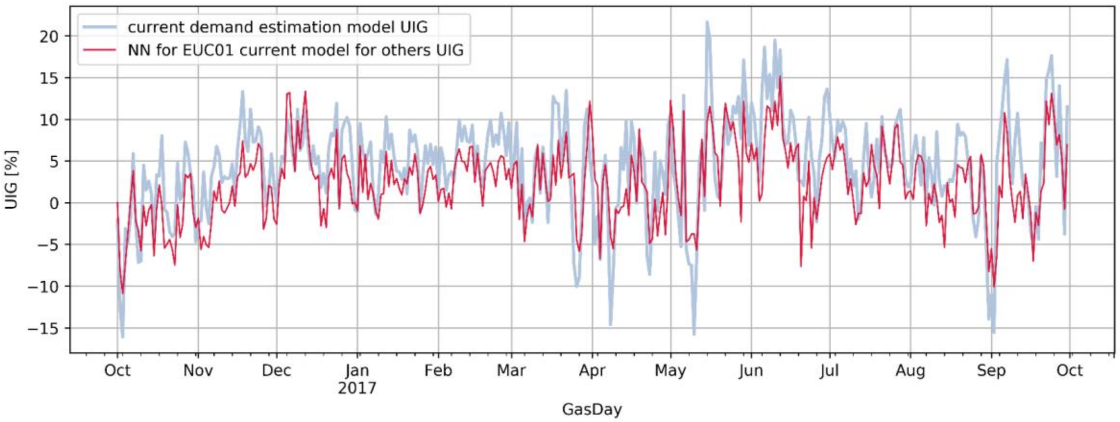
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	6.46	9.65	7.17	4.50
XGB 12-15, w17 B	5.62	8.08	5.81	3.46
NN 12-16, w17 B	4.43	7.67	6.26	3.67
NN 06-16, w17 B	2.83	7.44	6.88	3.82

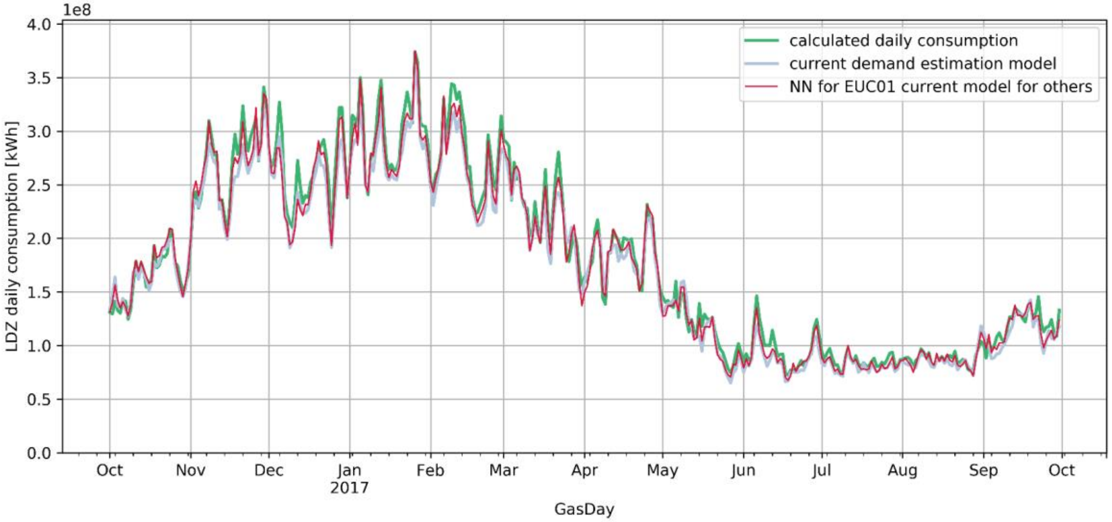




NW, NN 12-15 w17 B

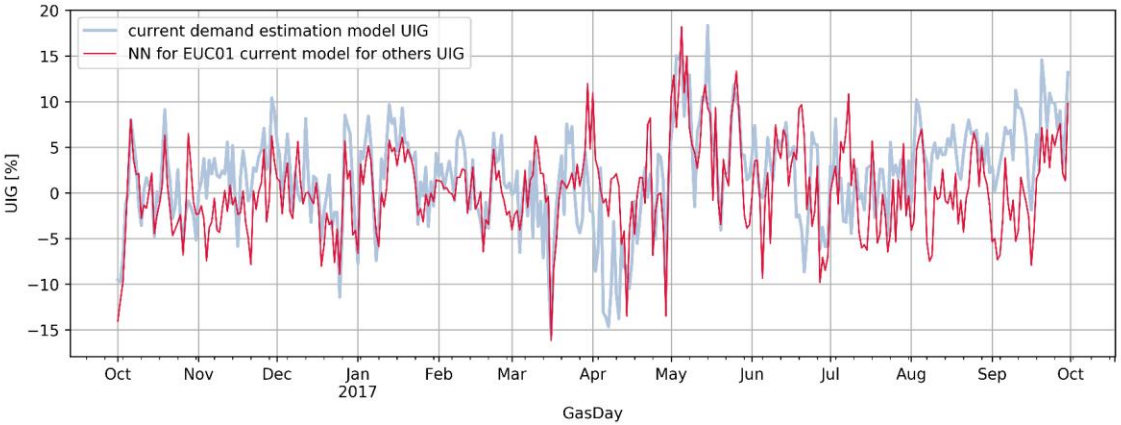
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	8.46	13.79	10.89	7.51
XGB 12-15, w17 B	5.32	10.13	8.62	6.0
NN 12-16, w17 B	4.50	9.19	8.01	5.50
NN 06-16, w17 B	4.28	9.37	8.33	5.63

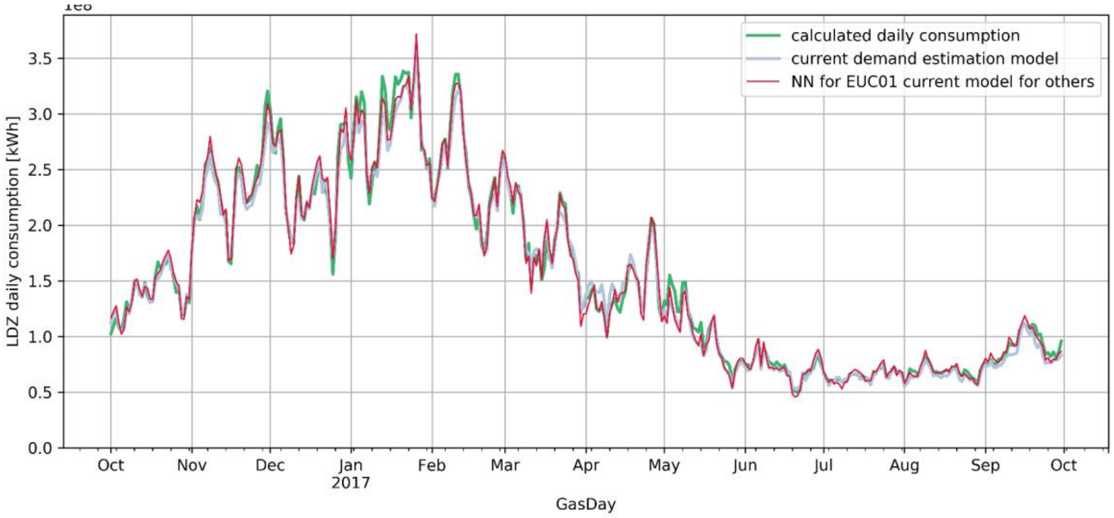




SE, NN 12-15 w17 B

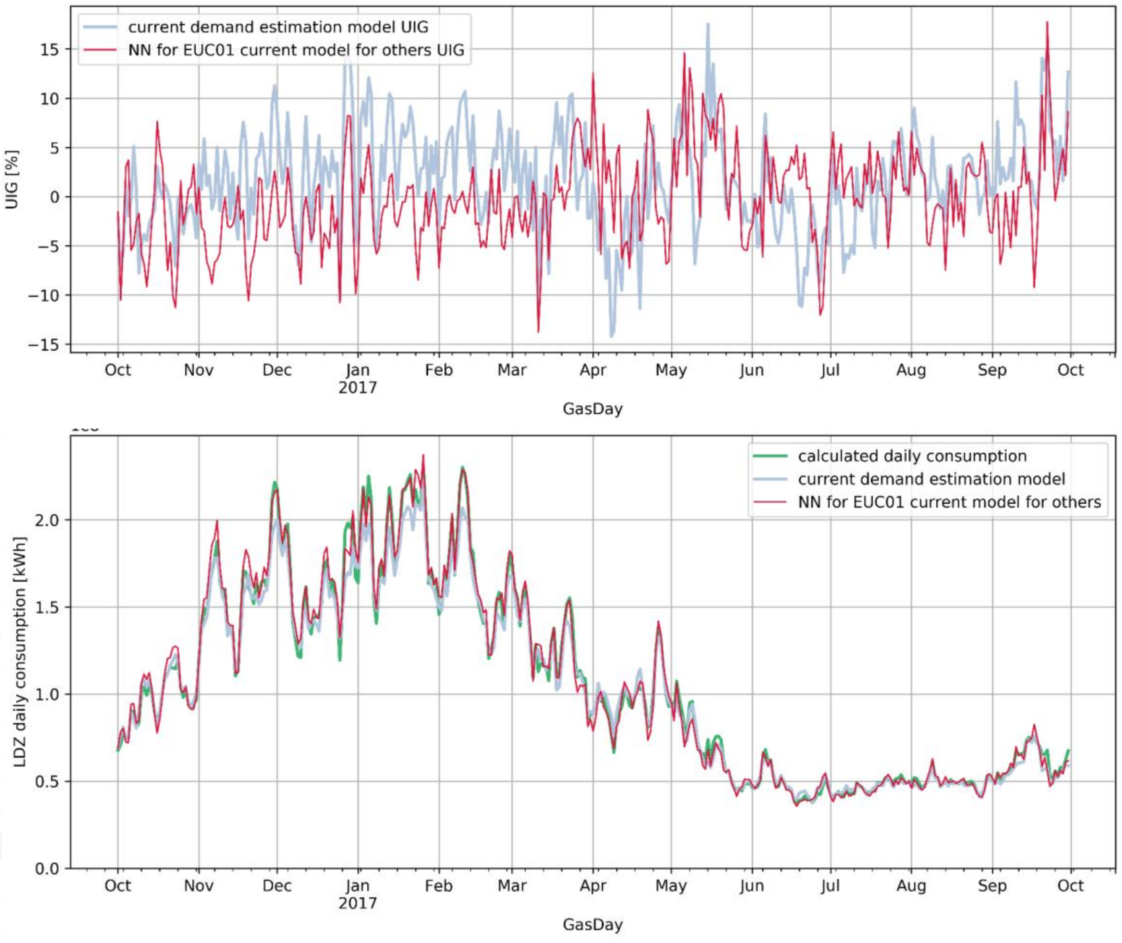
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.89	9.07	8.59	5.58
XGB 12-15, w17 B	2.13	7.66	7.36	4.65
NN 12-16, w17 B	1.29	7.44	7.33	4.63
NN 06-16, w17 B	0.34	6.97	6.97	4.58





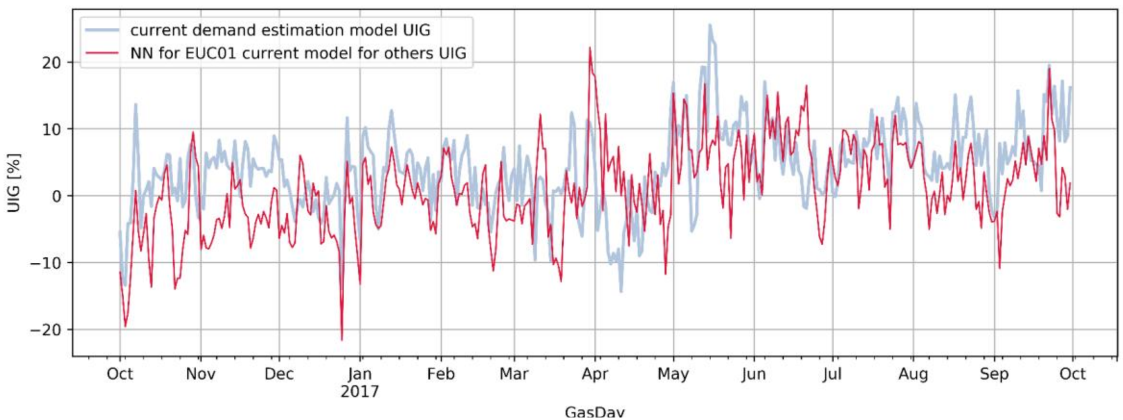
SO, NN 12-15 w17 B

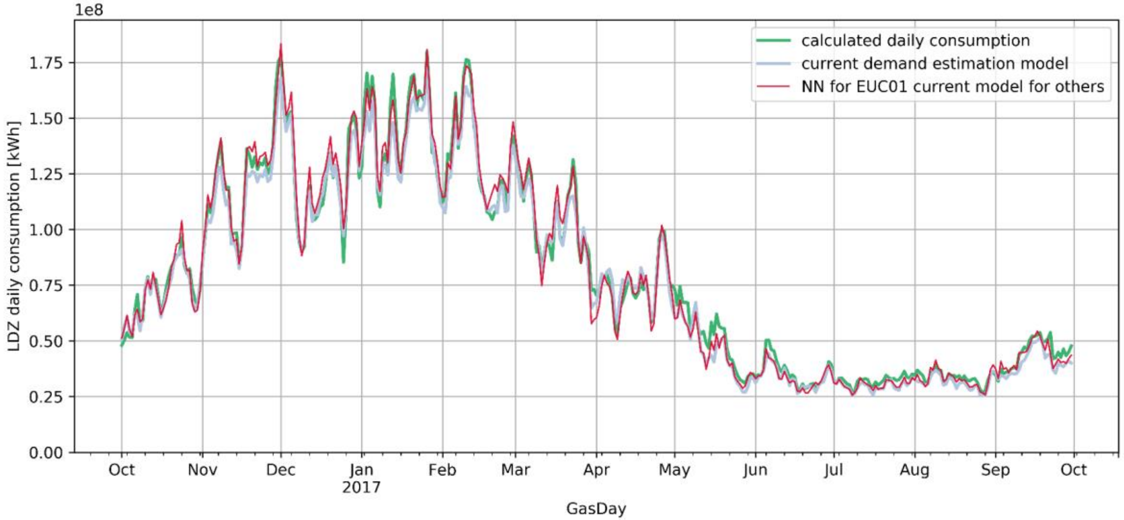
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.63	7.24	6.75	4.22
XGB 12-15, w17 B	0.11	5.06	5.06	3.23
NN 12-16, w17 B	0.41	4.69	4.68	3.24
NN 06-16, w17 B	-0.87	5.13	5.06	3.24



SW, NN 12-15 w17 B

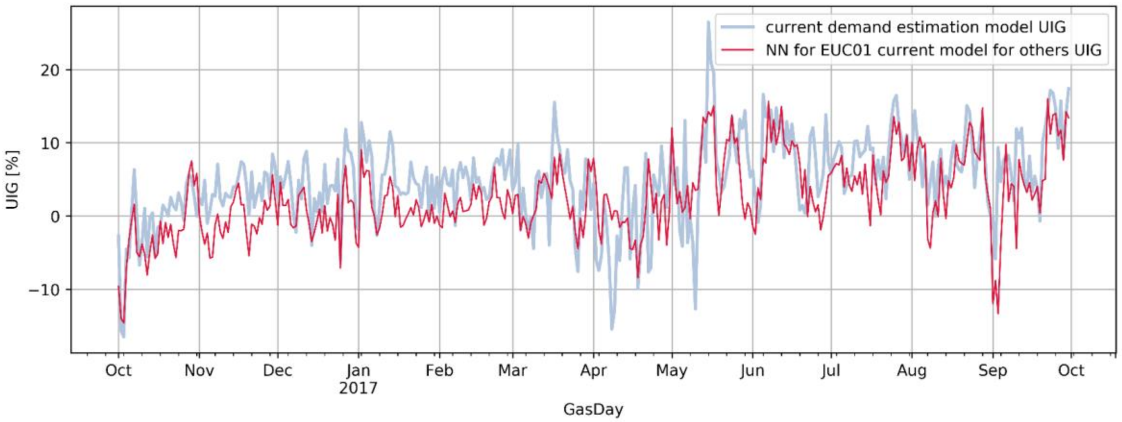
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	2.93	5.58	4.75	3.02
XGB 12-15, w17 B	0.61	4.44	4.40	2.46
NN 12-16, w17 B	1.84	8.39	8.19	5.01
NN 06-16, w17 B	0.05	4.50	4.50	2.52

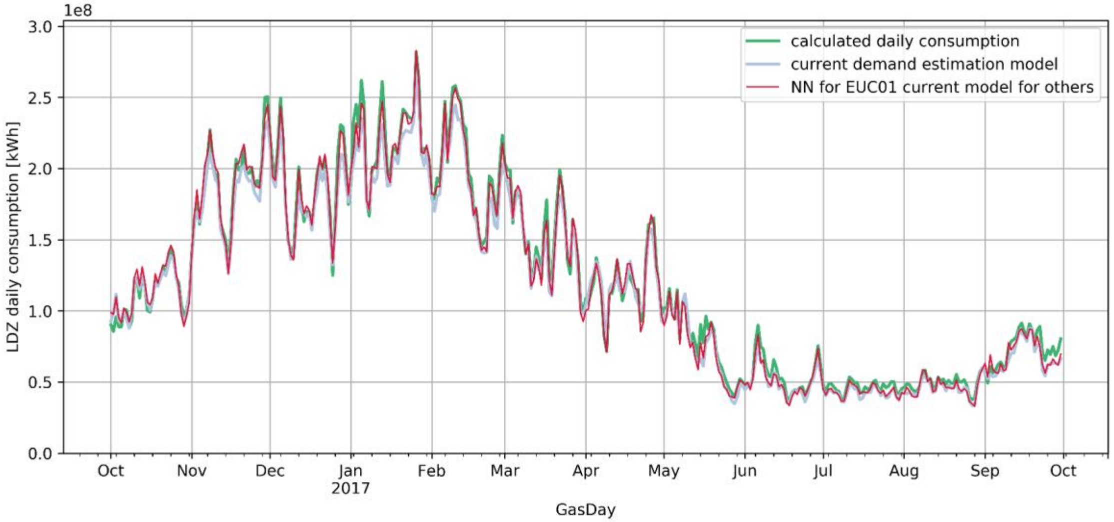




WM, NN 12-15 w17 B

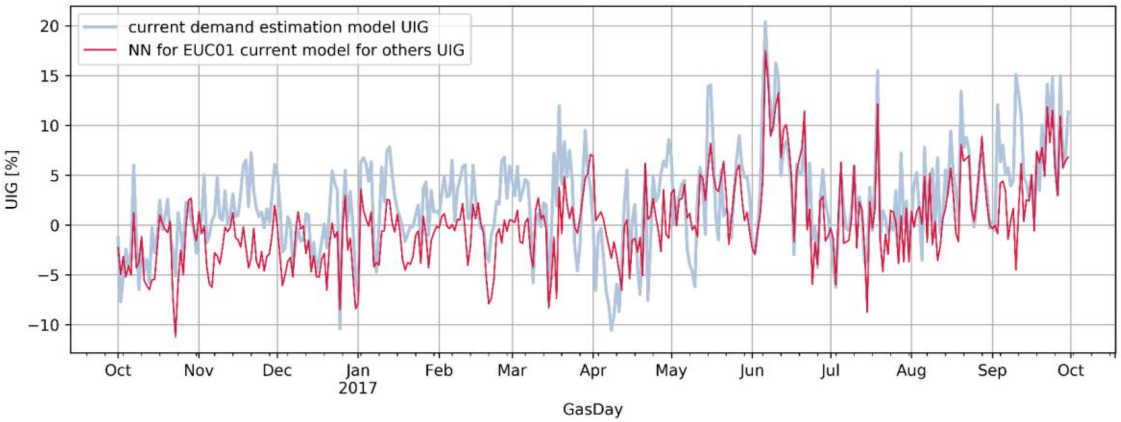
Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	5.52	8.88	6.95	4.90
XGB 12-15, w17 B	3.09	6.46	5.67	3.66
NN 12-16, w17 B	5.87	16.08	14.97	9.63
NN 06-16, w17 B	1.95	5.35	4.98	3.25

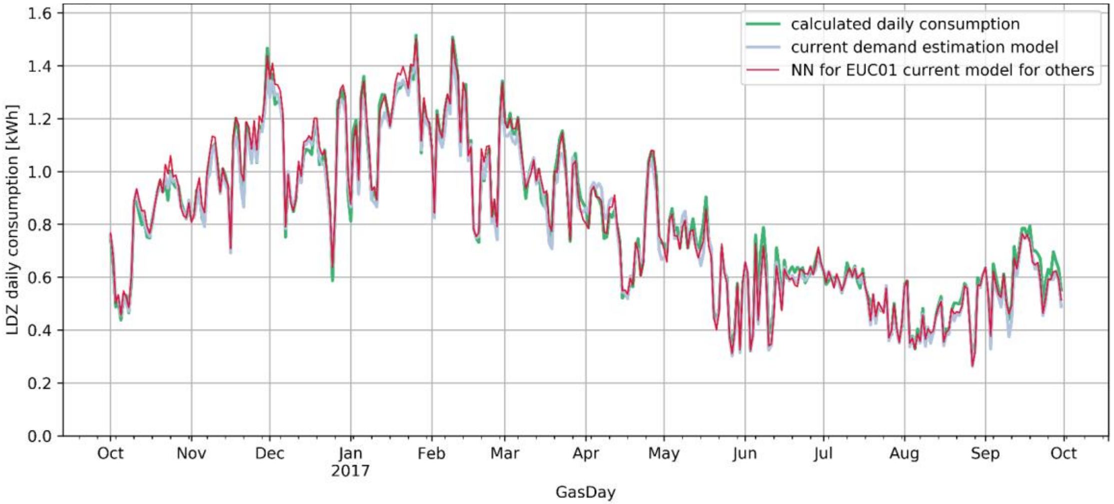




WS, NN 12-15 w17 B

Model	Mean daily error (GWh)	RMS error on total daily usage (GWh)	StD. on total daily error (GWh)	Mean (abs) day-to-day change in error (GWh)
current	1.70	4.00	3.62	2.66
XGB 12-15, w17 B	0.73	2.97	2.88	2.01
NN 12-16, w17 B	1.07	8.47	8.40	5.60
NN 06-16, w17 B	-0.14	3.14	3.14	2.00





Summary of Findings: Trial an EUC01 Neural Network model built at mainland UK level (no LDZ split)

Area & Ref #	Accuracy of NDM Algorithm - Use of Weather Data - Complex machine learning (Ref #13.2.6)
UIG Hypothesis Specific Item	In the previous analysis we used a neural network model for EUC1 combined with the current NDM model for the remaining meters to predict total daily consumption for each LDZ. The model was trained on sample data from gas years 2006-2015 and winter 2017 (06-15, w17). Predictions were made for gas year 2016. This model showed improved performance over the current NDM algorithm, and reduced prediction errors by around 30% on average. Neural networks (and other machine learning techniques) offer the ability to create highly nonlinear models using input data which is difficult to interpret using hand crafted models. In order to do this they typically require large amounts of training data. In this analysis we built a single neural network model for the whole country (with no splits by LDZ). The inputs to the model were raw weather data from ~12 weather stations around the country, day of the week and holiday indicators, and features which were specific to the location of each meter that were thought to be relevant to the gas consumption of an individual residential property. These were location (latitude, longitude), elevation, and population density. To offset the large amount of inputs (and hence model parameters), the model was trained on sample meter data from the whole country rather than LDZs individually. This approach offers an alternative to individual LDZ models, and may be able to capture the gas-usage pattern of individual households more accurately, and hence return a reduced UIG.
Data Tree References	EUC, Energy, Annual Quantity, Weather

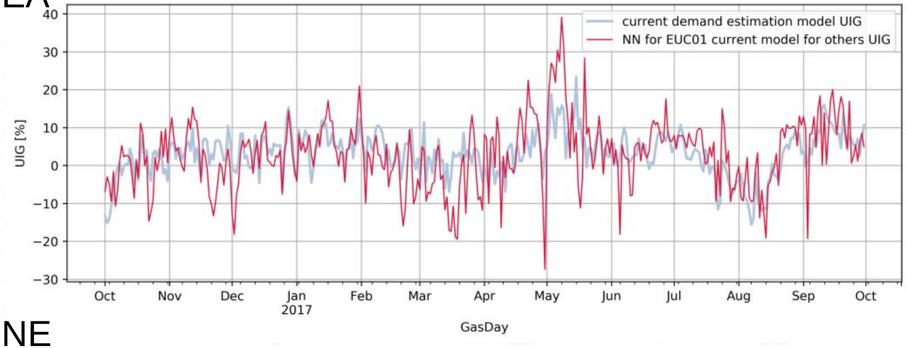
Findings Status	
UIG Impact Peak Volatility %	**None demon-strated**
UIG Impact Annual Average %	**None demon-strated**
Confidence in Percentages	**N/A**

Findings	**Approach to analysis**
Several neural network models were tested where the inputs were raw weather data from 12 weather stations. The structure of the neural networks chosen were general for the this type of regression problem, and not specifically engineered to using intuition or a significant number of trial and improvement steps. None of the trained models produced better results (when used to model EUC1, augmented with the current NDM model for other EUCs) than the current NDM model. Further work could be done on the whole country neural network model structure, and feature engineering of the inputs, to reduce the number of input parameters. However it is likely to be more fruitful to pursue generating similar models to those in the previous analysis on the remaining EUCs, as these have already showed significant promise.	The rough location of each meter was determined from the outside codes (first bit of the postcode) in the Key Data. For each outside code location an elevation was obtained using an online mapping API service. For each outer code region a population density estimate was made from 2011 Census data. This data as well as the weather data and calendar data was used to train a NN. Test results for the model were calculated for each LDZ as was done for the previous analysis. This total predicted consumption for the LDZ was then compared with the 'true consumption' (calc input energy, stock change, shrinkage). Error metrics for the current NDM model and the ML EUC01 model were calculated.

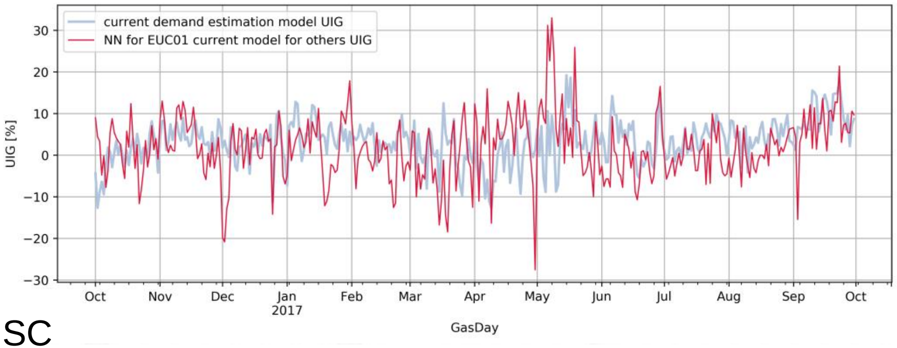
Supporting evidence 1: UIG plots for Mainland UK model

The initial NN results for the whole country model have not performed as well as the current NDM model. It is possible this model could be improved with significantly more work on the neural net architecture, and feature engineering of input parameters. The additional inputs used in this analysis (latitude, longitude, elevation, population density) could be used in the LDZ models similar to those generated in the previous analysis, and/or the individual LDZ models could be trained simultaneously to reduce manual effort. This is likely to be a more productive avenue for investigation than doing a lot of work on whole country model.

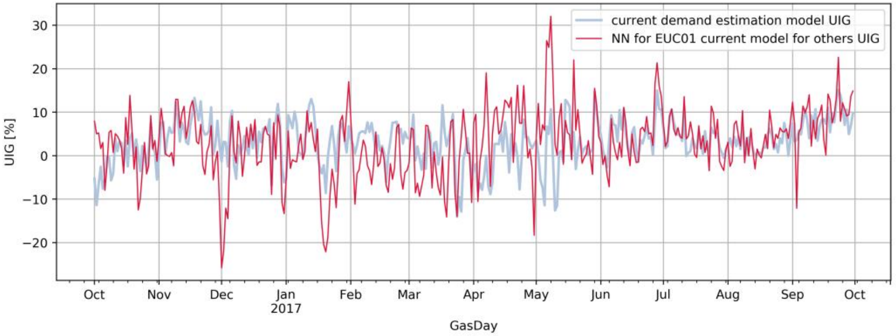
EA



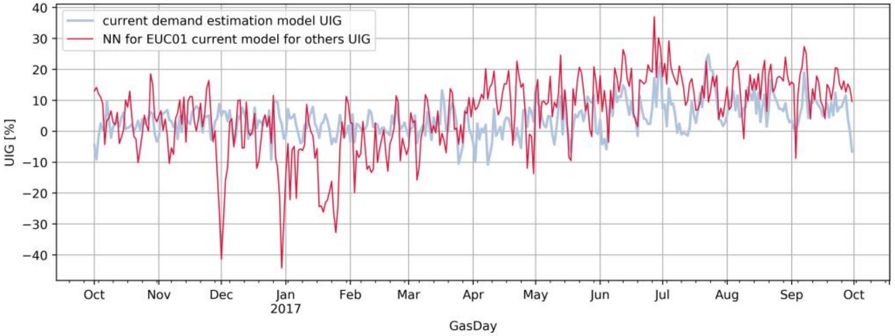
EM



NE



SC



xoserve