# UIG Task Force

# 13.2.7: Advanced Machine Learning

# 13.2.7.a

# Machine Learning Prediction using LDZ Input Energy as model input value

# **Findings Summary:** Adding the LDZ input data as an input into the Neural Network improves the model performance and reduces predicted UIG volatility

- Hypothesis:
  - The total amount of gas input to the LDZ system for a given day may prove to be useful information for the Machine Learning model to learn from, and therefore better predict daily Non-Daily Metered usage
- Analysis method:
  - Extract LDZ input gas information from historical Gemini data for the years 2006-2017
  - Alter the ML model to allow LDZ gas as an input feature
  - Retrain the model with the additional LDZ input data
  - Test and compare models (with LDZ input gas vs. without LDZ input gas)
- Summary of Results:
  - It appears that LDZ Input does correlate with UIG, and so including it as a variable improves NDM energy prediction, consequently reducing UIG volatility at allocation
  - The improvement is more significant for some LDZs (e.g. SC vs. EM)
  - However, a bigger net benefit is seen by training the ML model on consumption data that covers a longer period (i.e. the 2006 – 2015 trained version of the Neural Net Model) compared with including LDZ input data in a smaller training dataset covering a shorter period.

# A note on models

- We have used different models to produce results. These models can vary according to the type of machine learning algorithm, configuration, the input data presented, and the duration of the training data.
- The following summary of the model is presented with each set of results:

| Model Summary | |
|---|---|
| **Model** | Functional NN |
| **Segment** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard set*, LDZ input |
| **Training** | GY 2011-2015 |
| **Testing** | GY 2016 |

The name of the model; XGBoost, Sequential NN, Functional NN
Sequential NN – the original neural net model from the previous phase of work
Functional NN – a new model developed in this phase

Whether different versions of the model were trained on each LDZ or each EUC separately or if these were bunched together

Which inputs we used in the model

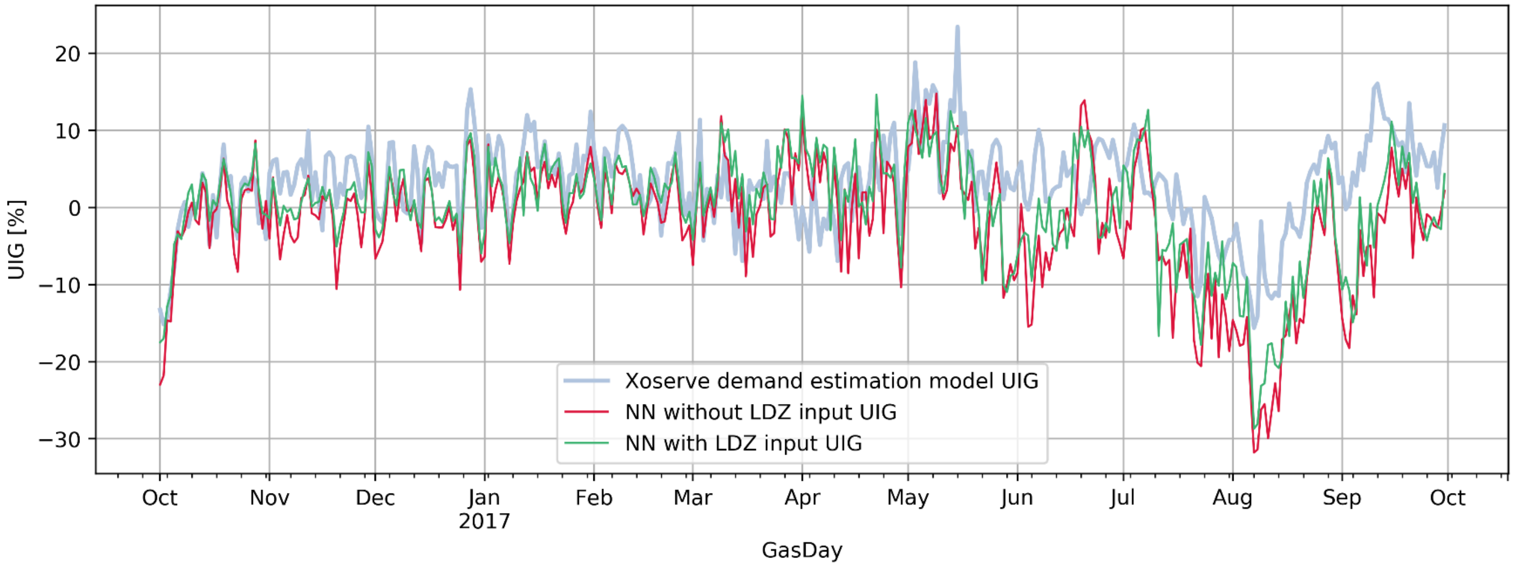
***Standard set:** Calculated AQ, temperature (a subset of), temperature previous day (a subset of), humidity (a subset of), solar irradiance, wind speed (a subset of), precipitation (a subset of), day of week, holiday, CWV

# Machine Learning with LDZ Input – UIG Results for EA LDZ

| Model | Absolute Mean UIG (Base UIG GWhr) | UIG Standard Deviation (UIG Volatility GWhr) |
|---|---|---|
| Xoserve model | 4.58 | 7.26 |
| NN without LDZ input | 1.16 | 6.13 |
| NN with LDZ input | 0.16 | 6.53 |

| Difference in UIG – compared to baseline ML Model (%) |
|---|
| **-0.9%** |

| Model Summary | |
|---|---|
| **Model** | Sequential Neural Network Model |
| **Model Details** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard Input Dataset \| LDZ Input |
| **Training Data** | Gas Year 2006-2017 (excluding. 2016) |
| **Testing Data** | Gas Year 2016 |

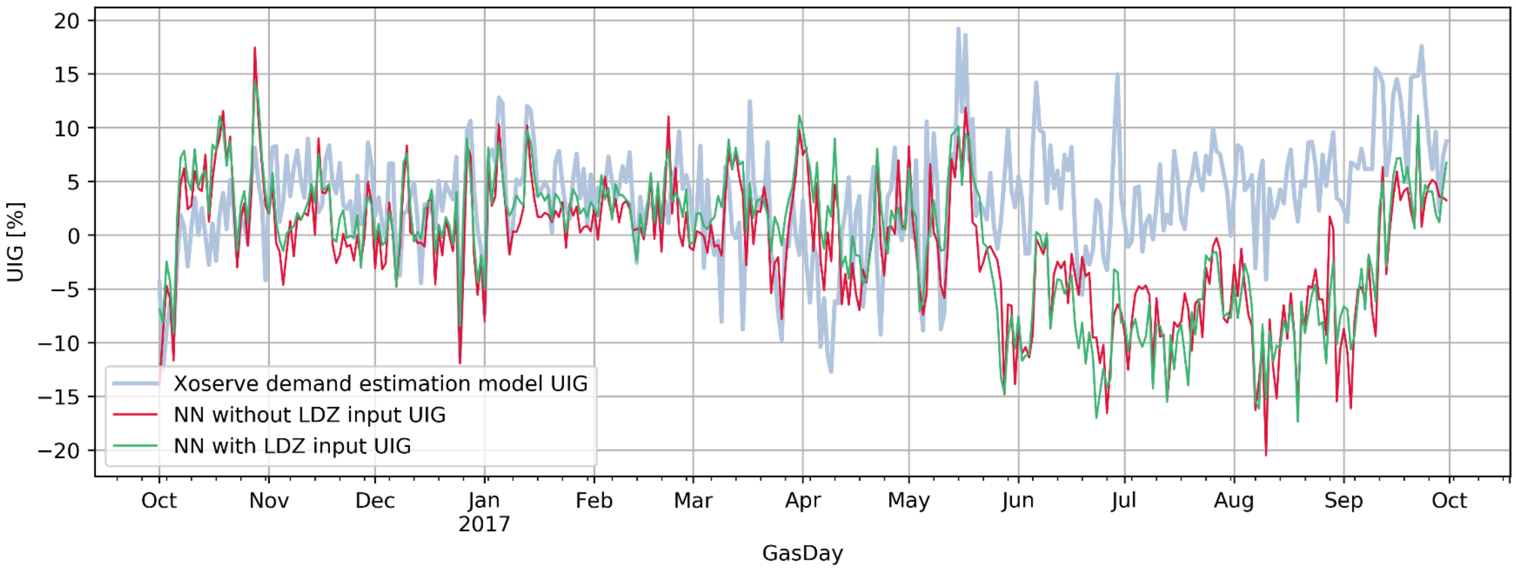


# Machine Learning with LDZ Input – UIG Results for EM LDZ

| Model | Absolute Mean UIG (Base UIG GWhr) | UIG Standard Deviation (UIG Volatility GWhr) |
|---|---|---|
| Xoserve model | 5.36 | 8.55 |
| NN without LDZ input | 0.6 | 7.87 |
| NN with LDZ input | 2.24 | 7.64 |

| Difference in UIG – compared to baseline ML Model (%) |
|---|
| **-0.4%** |

| Model Summary | |
|---|---|
| **Model** | Sequential Neural Network Model |
| **Model Details** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard Input Dataset \| LDZ Input |
| **Training Data** | Gas Year 2006-2017 (excluding. 2016) |
| **Testing Data** | Gas Year 2016 |

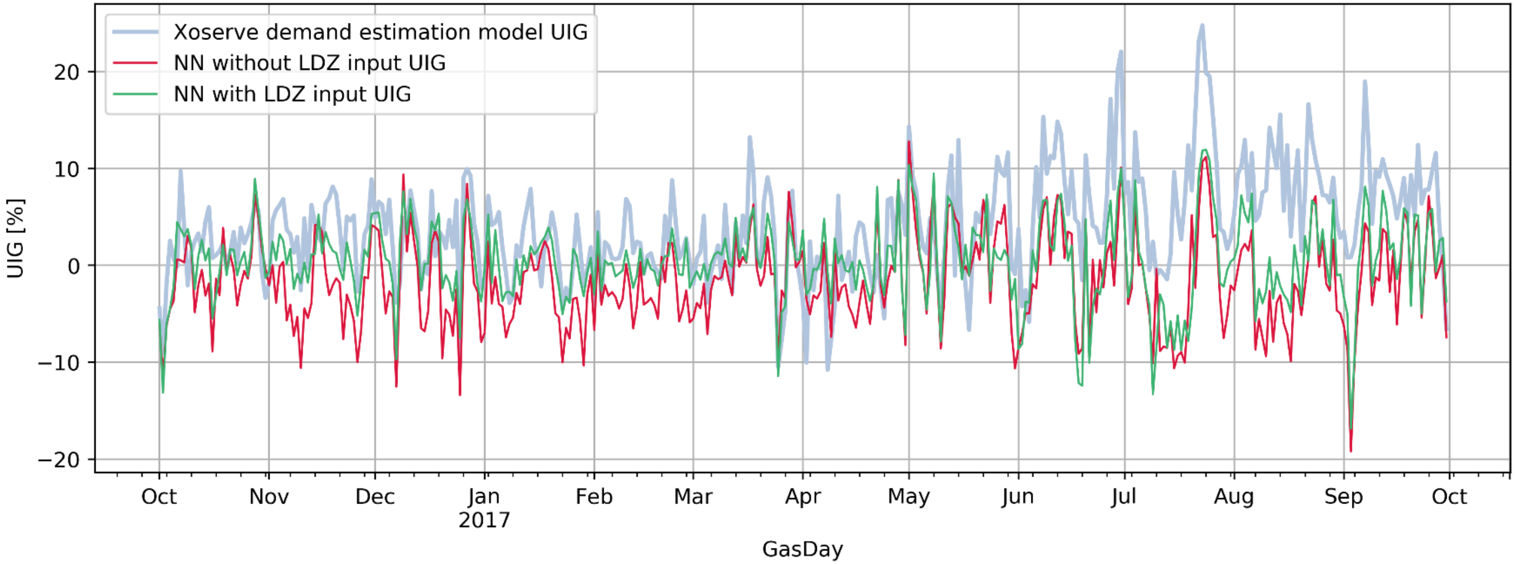


# Machine Learning with LDZ Input – UIG Results for SC LDZ

| Model | Absolute Mean UIG (Base UIG GWhr) | UIG Standard Deviation (UIG Volatility GWhr) |
|---|---|---|
| Xoserve model | 4.56 | 6.18 |
| NN without LDZ input | 1.6 | 6.5 |
| NN with LDZ input | 0.27 | 5.21 |

| Difference in UIG – compared to baseline ML Model (%) |
|---|
| **-2.6%** |

| Model Summary | |
|---|---|
| **Model** | Sequential Neural Network Model |
| **Model Details** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard Input Dataset \| LDZ Input |
| **Training Data** | Gas Year 2006-2017 (excluding. 2016) |
| **Testing Data** | Gas Year 2016 |



**UIG Volatility at Allocation is improved by adding the LDZ Input as an input: although UIG Base gets slightly worse. Adding the LDZ Input improves the UIG volatility by around 1.5%**

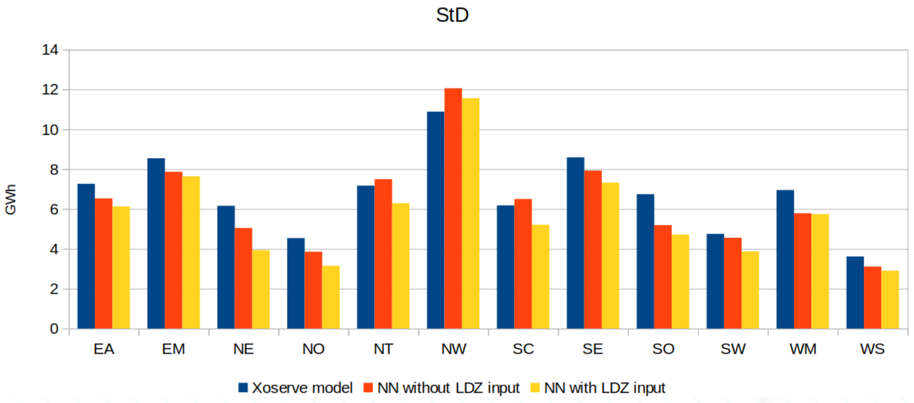
| | Metrics Summary – Averaged across all LDZs | | |
|---|---|---|---|
| | Base UIG: Absolute Mean [GWh] | UIG Volatility: Standard Deviation [GWh] | UIG Volatility: Mean Absolute Difference [GWh] |
| **Xoserve model** | 4.4 | 6.8 | 4.6 |
| **NN without LDZ input** | 1.1 | 6.3 | 4.2 |
| **NN with LDZ input** | 1.7 | 5.7 | 3.5 |

| Model Summary | |
|---|---|
| **Model** | Sequential Neural Network Model |
| **Model Details** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard Input Dataset \| LDZ Input |
| **Training Data** | Gas Year 2006-2017 (excluding. 2016) |
| **Testing Data** | Gas Year 2016 |

# Model Performance Comparison – With vs. Without LDZ Input

| Model Summary | |
|---|---|
| **Model** | Sequential Neural Network Model |
| **Model Details** | 1 model per LDZ for EUC1 only |
| **Inputs** | Standard Input Dataset \| LDZ Input |
| **Training Data** | Gas Year 2006-2017 (excluding. 2016) |
| **Testing Data** | Gas Year 2016 |

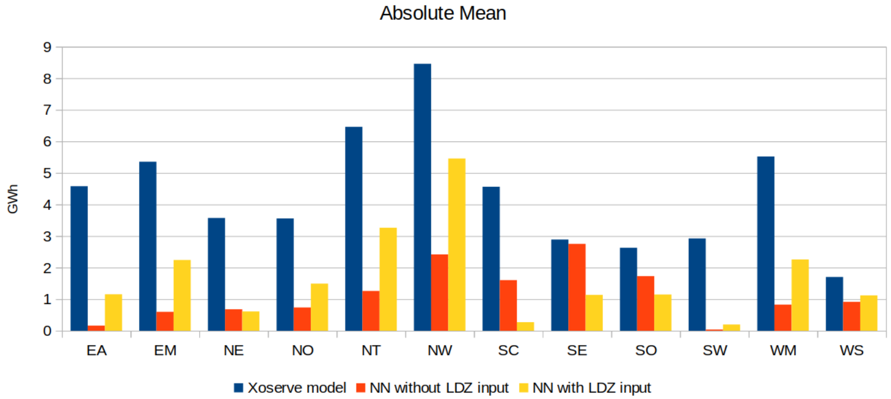
**UIG Volatility (measured as standard deviation):** adding LDZ input data always improves the performance of the neural network. Performance is normally always better than the Xoserve model (exception is NW)

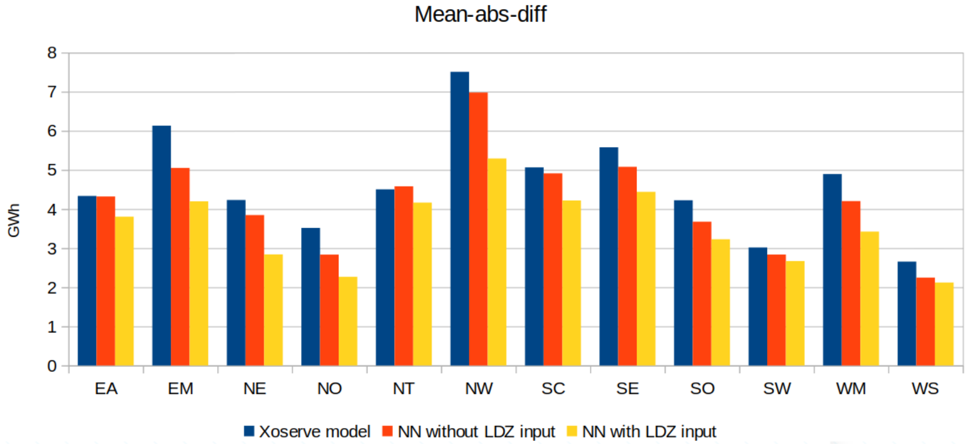


# Model Performance Comparison – With vs. Without LDZ Input

**UIG Base (measured as the mean):** adding LDZ input data doesn't necessarily improve the base UIG (although performance still always better than Xoserve model)

**UIG Volatility (mean absolute difference which is an alternative volatility metric):** adding LDZ input data always improves the performance of the neural network. Performance is always better than the Xoserve model (exception is NW)



# 13.2.7 b

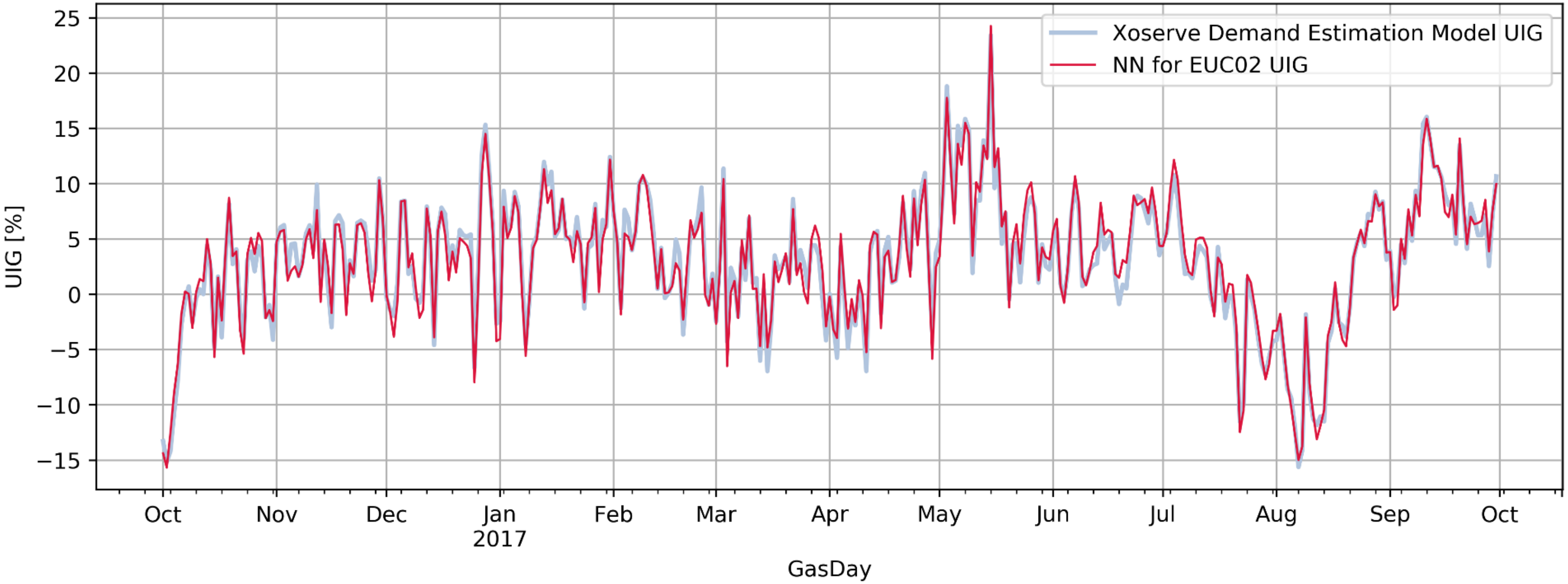
# Improve Modelling for EUC2

# Summary

- Transfer learning stores knowledge gained while solving one problem and applies it to a different but related problem. In this case: transfer learning was used to fit parameters calculated by training on EUC1 to kick-start the fit on EUC2.
- Transfer learning addresses the problem raised by industry stakeholders of neural networks typically requiring a large amount of data to achieve good performance and avoid overfitting. The higher EUCs have significantly fewer data points than the previously investigated EUC1.
- We have applied transfer learning to EUC2, and the subsequent performance reduces the predicted UIG at Allocation compared to the current NDM model. Considering that EUC2 is only a small percentage (~5%) of the total energy, the improvements are larger than expected.

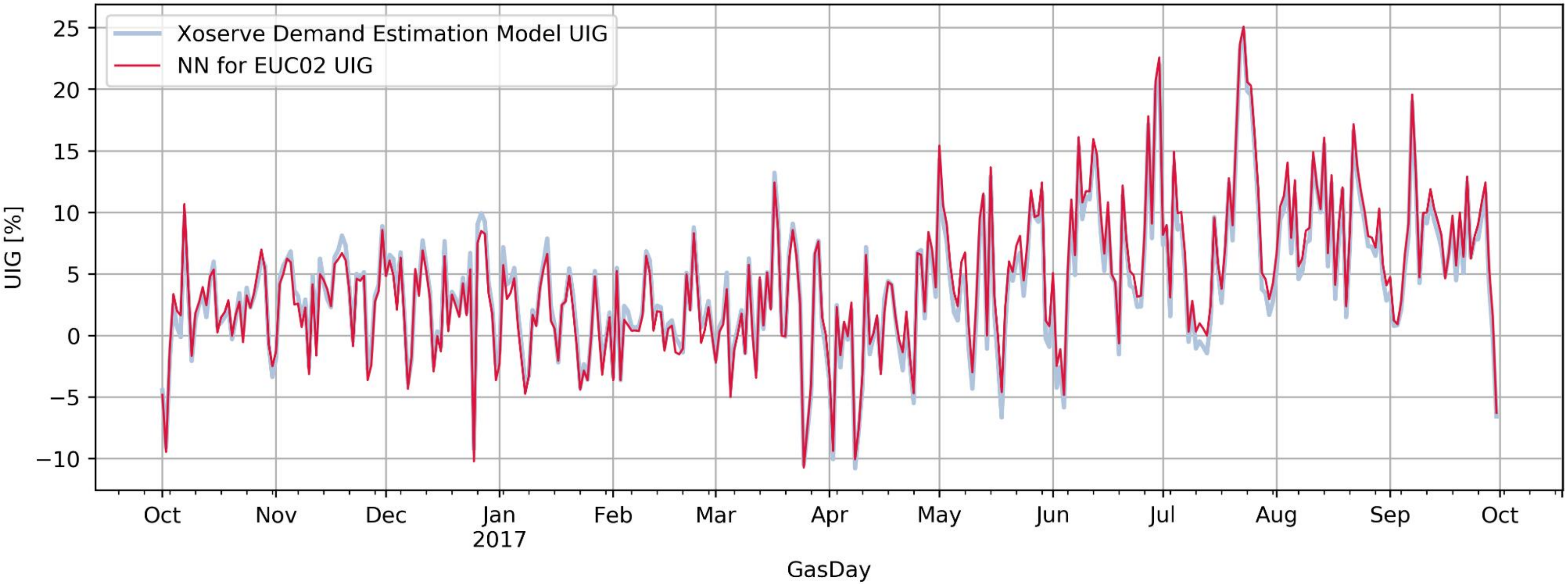
# Results with the sequential NN on EUC 2 for EA LDZ

- Here is the effect of modelling **EUC2 only** (i.e. not including the NN model for EUC1) with the neural net compared to the current NDM model:



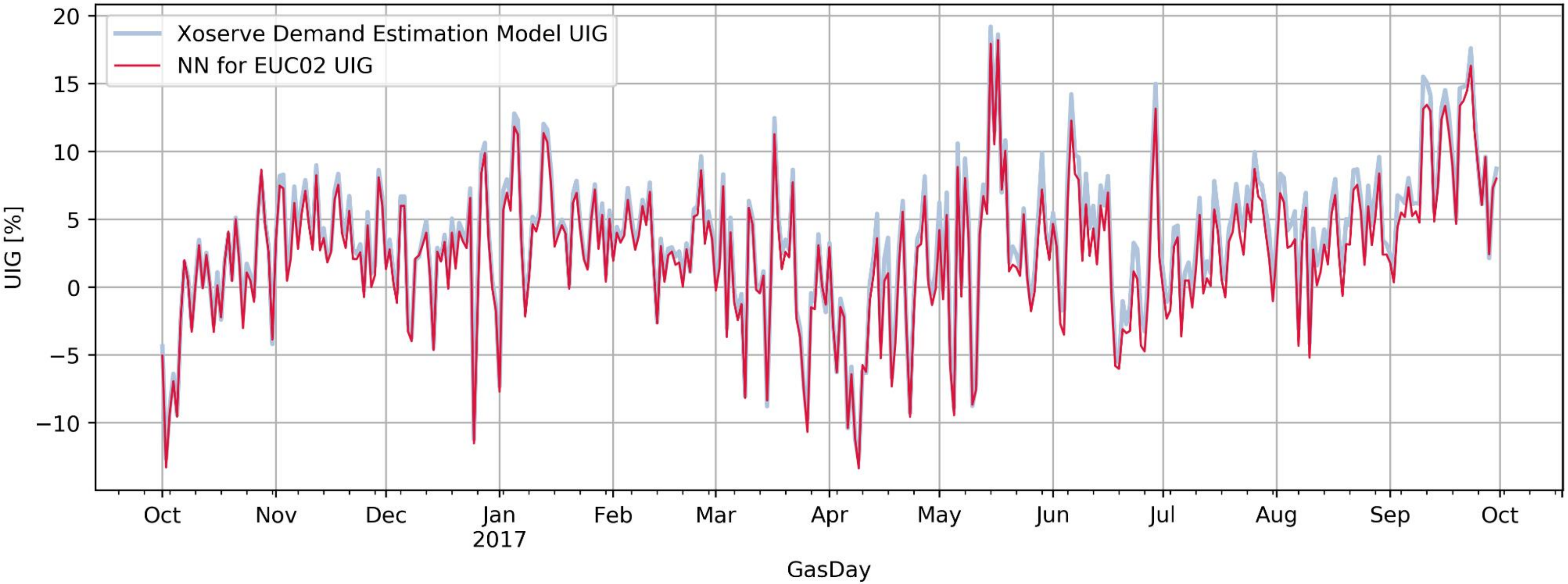
# Results with the sequential NN on EUC 2 for EM LDZ

- Here is the effect of modelling **EUC2 only** (i.e. not including the NN model for EUC1) with the neural net compared to the current NDM model:



# Results with the sequential NN on EUC 2 for SC LDZ

- Here is the effect of modelling **EUC2 only** (i.e. not including the NN model for EUC1) with the neural net compared to the current NDM model:



# Results with the sequential NN on EUC 2

- These are our first results from training the Neural Net on the EUC2 sample set and evaluating it on the whole population for the 2016 gas year. This was trained on gas years 2011 – 2015.

| | Current NDM mean UIG (Base UIG) [GWh] | NN augmented NDM mean UIG [GWh] | % Change in mean | Current NDM StD UIG (UIG Volatility) [GWh] | NN augmented NDM StD UIG [GWh] | % Change in StD |
|---|---|---|---|---|---|---|
| EA | 4.58 | 4.30 | -6.2% | 7.26 | 6.91 | -4.8% |
| EM | 5.36 | 4.23 | -21.0% | 8.55 | 8.15 | -4.7% |
| NE | 3.57 | 3.31 | -7.2% | 6.16 | 6.12 | -0.8% |
| NO | 3.56 | 3.34 | -6.2% | 4.54 | 4.48 | -1.3% |
| NT | 6.46 | 4.92 | -23.9% | 7.17 | 6.66 | -7.1% |
| NW | 8.46 | 7.61 | -10.1% | 10.89 | 10.60 | -2.7% |
| SC | 4.56 | 4.53 | -0.7% | 6.18 | 6.06 | -2.0% |
| SE | 2.89 | 2.38 | -17.7% | 8.59 | 8.54 | -0.6% |
| SO | 2.63 | 2.36 | -10.1% | 6.75 | 6.59 | -2.3% |
| SW | 2.93 | 2.86 | -2.3% | 4.75 | 4.66 | -1.9% |
| WM | 5.52 | 3.92 | -28.9% | 6.95 | 6.77 | -2.6% |
| WS | 1.70 | 1.12 | -34.0% | 3.62 | 3.55 | -1.9% |

# 13.2.7 c

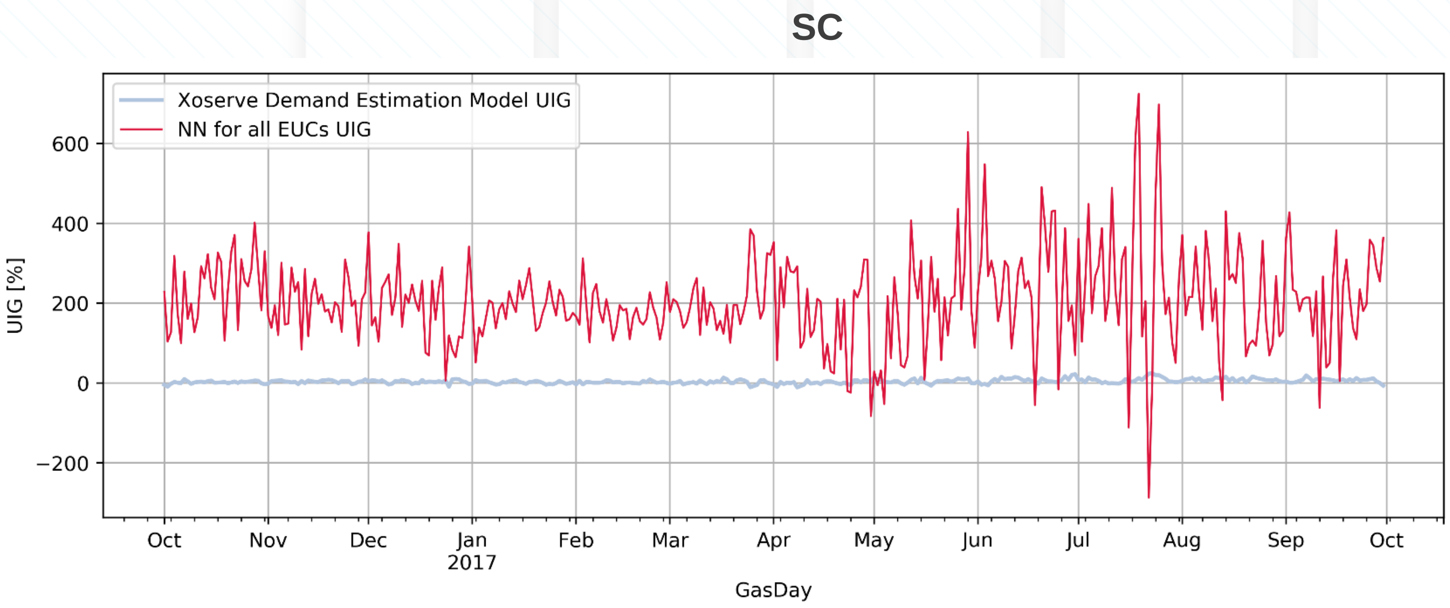
# Improve Modelling for EUCs 3-8

# Summary

- We have taken a two pronged approach to extending machine learning to the entire NDM set.
  - Using the Sequential NN with the option of using transfer learning
  - Building the Functional model and training on groups of LDZs/EUCs simultaneously
- Transfer learning performs well as a separate model for EUC2, but when we split models by EUC and LDZ for the higher EUCs, it appears that there is still not sufficient data to train.
- The functional model addresses the lack of data within a specific LDZ and EUC, with the intention of avoiding the problems we encountered when developing a whole country sequential model in the previous phase of work, where we observed poor scaling with the number of model parameters.

# Transfer Learning Applied to EUCs 3-8

- Transfer learning performs well as a separate model for EUC2, but when we split models by EUC and LDZ for the higher EUCs, it appears that there is still not sufficient data to train.
- On EUC1-9 or EUC 3+, the Sequential Neural Network does not train or perform well (see example below)




| Model Summary | |
|---|---|
| **Model** | Sequential Model |
| **Segment** | 1 model per LDZ for all EUCs |
| **Inputs** | Standard set |
| **Training** | GY 2006-2017 (exc. 2016) |
| **Testing** | GY 2016 |

# Next Step: Develop Functional Model

- Developing a new neural network model (Functional NN) to address following issues with the current Sequential NN (SNN):
  - The SNN was not linear in AQ, this meant it was slow to evaluate and very difficult to back-calculate AQs from meter-reads with this model
  - SNN do not incorporate any prior assumptions about the interaction of the inputs, however many of these can be well estimated or are already understood
  - It has proved challenging to adapt the current SNN to take inputs for multiple LDZs or EUCs in a logical way to cope with the smaller amount of data in the higher numbered EUCs.
- A Functional Model has a hand-coded architecture. The necessity of the hand coded architecture is to add explainability and prior knowledge into the model. This can lead to a significant reduction in the number of fitted parameters, which reduces the susceptibility of a model to overfitting.
- This approach would combine the best parts of hand coded models and neural networks to improve the NDM estimation.
- The data required for input to the Functional NN was pre-processed, covering gas years 2013-2018. The model was trained on years 2013-2015, leaving 2016-2018 for testing.
- The Sequential NN was revised in order to make the model predictions for EUC01 and EUC02 easily combinable (as the results showed it to work well on both, not only EUC01), and now cover the two new test years (2017 and 2018).

# Improve Modelling for EUCs 2+ Conclusions

- It is possible to replace the Xoserve NDM model in its entirety with a neural network whilst reducing UIG (base and volatility).

- The Functional NN was unable to model EUC01 well. Possible causes for this include:
  - The functional NN had to use system EUCs and AQs so less data was used
  - The functional NN included non-domestic meter points (whereas the sequential NN did not)

- Note that comparisons to Xoserve NDM model for GY2017 and GY2018 must consider that Xoserve model includes more recent years to train on than the Neural Network.

**Metrics Comparison for 2016 (averaged over LDZs)**

| Model | Absolute Mean | Root Mean Square Error | Standard Deviation | Mean Absolute Difference |
|---|---|---|---|---|
| Xoserve Model | 3.74 | 7.82 | 6.79 | 4.64 |
| Sequential (EUC 1-2) + Functional (EUC 3-9) | 2.14 | 6.26 | 5.81 | 3.70 |

# Metrics Summary Tables

| Metrics Comparison for 2016 (averaged over LDZs) | | | | |
|---|---|---|---|---|
| Model | Absolute Mean | Root Mean Square Error | Standard Deviation | Mean Absolute Difference |
| Xoserve Model | 3.74 | 7.82 | 6.79 | 4.64 |
| Sequential (EUC 1-2) + Functional (EUC 3-9) | 2.14 | 6.26 | 5.81 | 3.70 |
| Functional Model | 6.97 | 47.60 | 47.01 | 11.41 |

| Metrics Comparison for 2017 (averaged over LDZs) | | | | |
|---|---|---|---|---|
| Model | Absolute Mean | Root Mean Square Error | Standard Deviation | Mean Absolute Difference |
| **Xoserve Model** | 5.57 | 9.96 | 8.17 | 4.72 |
| **Sequential (EUC 1-2) + Functional (EUC 3-9)** | 2.43 | 7.59 | 7.11 | 3.95 |
| **Functional Model** | 12.17 | 59.40 | 58.02 | 12.29 |

| Metrics Comparison for 2018 (averaged over LDZs) | | | | |
|---|---|---|---|---|
| Model | Absolute Mean | Root Mean Square Error | Standard Deviation | Mean Absolute Difference |
| Xoserve Model | 2.51 | 6.70 | 6.16 | 4.28 |
| Sequential (EUC 1-2) + Functional (EUC 3-9) | 1.23 | 6.60 | 6.43 | 4.00 |
| Functional Model | 4.33 | 46.08 | 45.75 | 11.61 |

| Metrics Comparison for 2018 WITHOUT SW | | | | |
|---|---|---|---|---|
| Model | Absolute Mean | Root Mean Square Error | Standard Deviation | Mean Absolute Difference |
| Xoserve Model | 2.51 | 6.70 | 6.16 | 4.28 |
| Sequential (EUC 1-2) + Functional (EUC 3-9) | 0.83 | 5.80 | 5.70 | 4.02 |
| Functional Model | 4.33 | 46.08 | 45.75 | 11.61 |

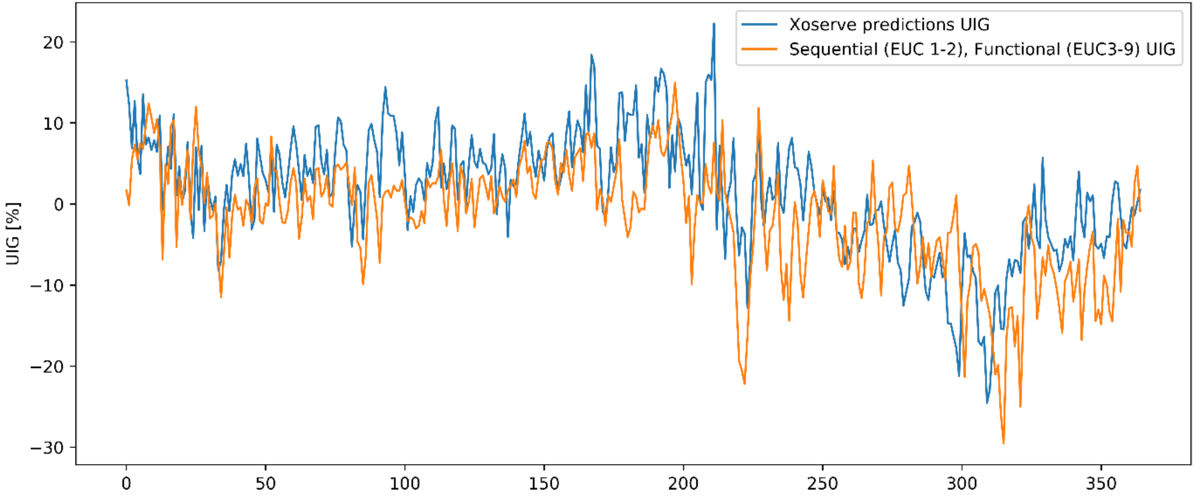
This result skewed by unusual behaviour observed in SW LDZ prediction

# The Functional NN appears work best when modelling EUCs 3-9

## UIG Comparison of ML Models – 2017 UIG Results for EA LDZ

| | Abs Mean (Base UIG) GWh | Standard Deviation (UIG Volatility) GWh |
|---|---|---|
| Xoserve model | **5.12** | **9.28** |
| Sequential (EUC 1-2) + Functional (EUC 3-9) | **0.83** | **6.50** |

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |

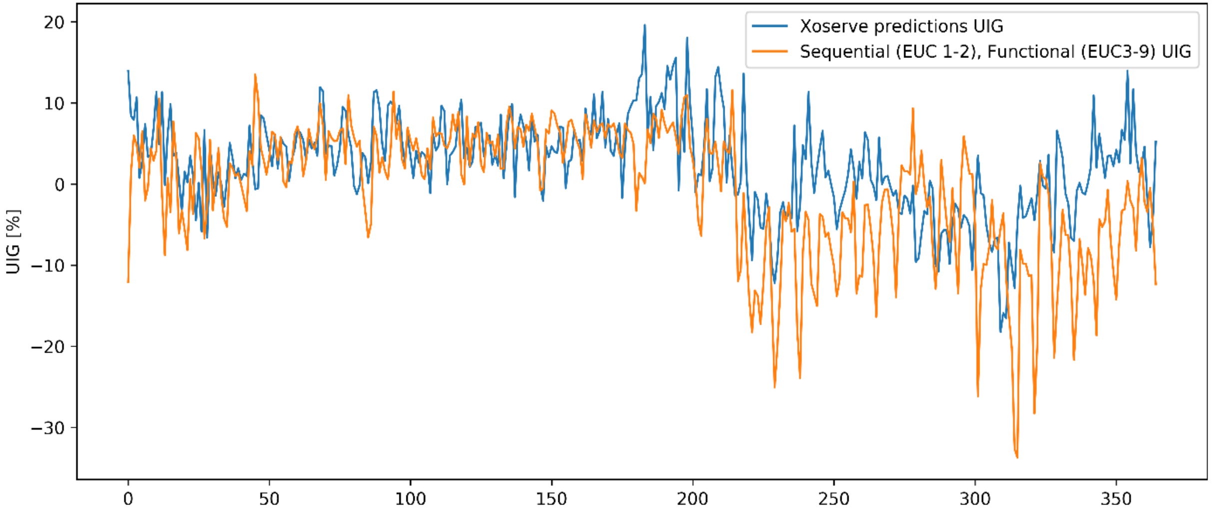


# The Functional NN appears work best when modelling EUCs 3-9

## UIG Comparison of ML Models – 2017 UIG Results for EM LDZ

| | Abs Mean (Base UIG) GWh | Standard Deviation (UIG Volatility) GWh |
|---|---|---|
| **Xoserve model** | **6.85** | **10.45** |
| **Sequential (EUC 1-2) + Functional (EUC 3-9)** | **3.67** | **11.18** |

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY **2013**-2015 |
| **Testing** | GY 2016-2018 |

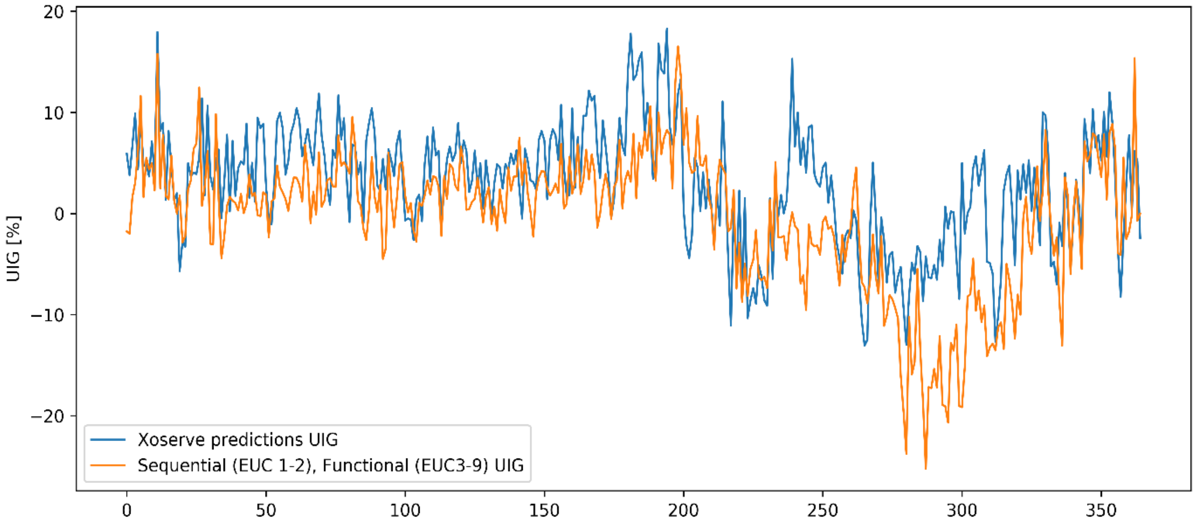


# The Functional NN appears work best when modelling EUCs 3-9

## UIG Comparison of ML Models – 2017 UIG Results for SC LDZ

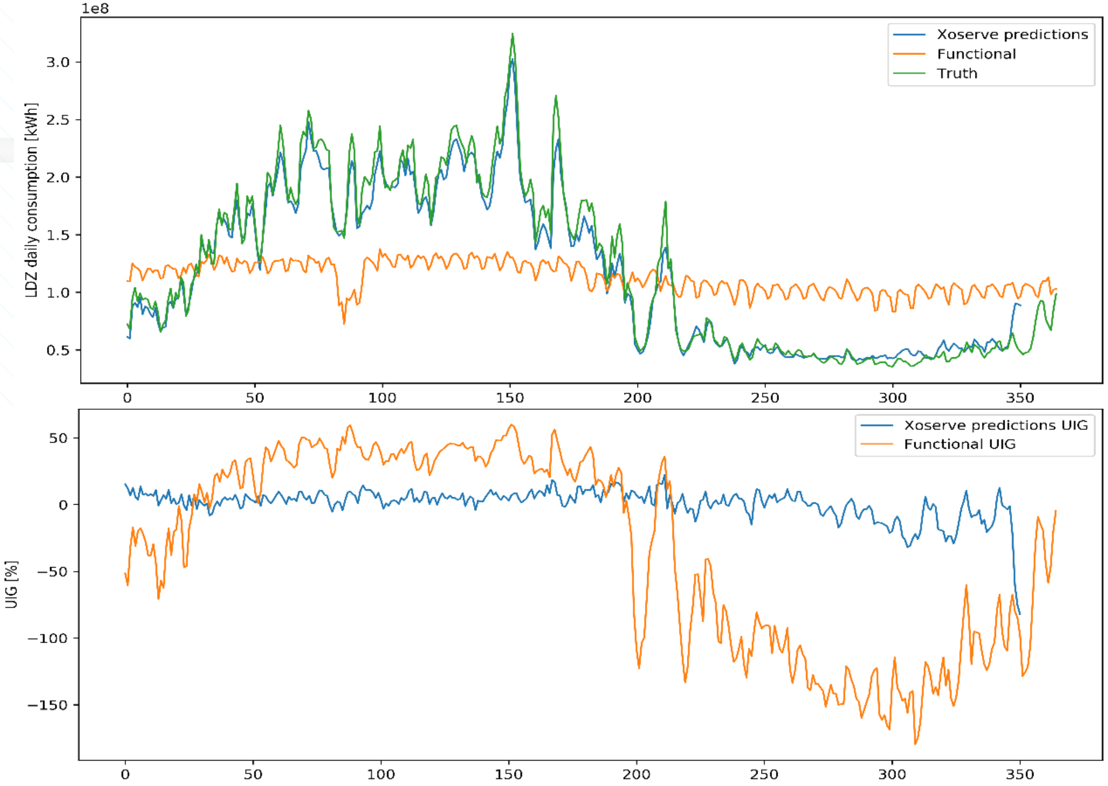
| | Abs Mean (Base UIG) GWh | Standard Deviation (UIG Volatility) GWh |
|---|---|---|
| **Xoserve model** | 6.37 | 9.12 |
| **Sequential (EUC 1-2) + Functional (EUC 3-9)** | 2.13 | 6.71 |

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |



# The Functional NN was unable to model EUC1 well:
Example of Functional ML Model Results – EA 2017

Comparing to the 3 previous slides, going from using the Functional model for EUC3-9 to all EUCs causes an enormous degradation in performance, largely from the inclusion of EUC1

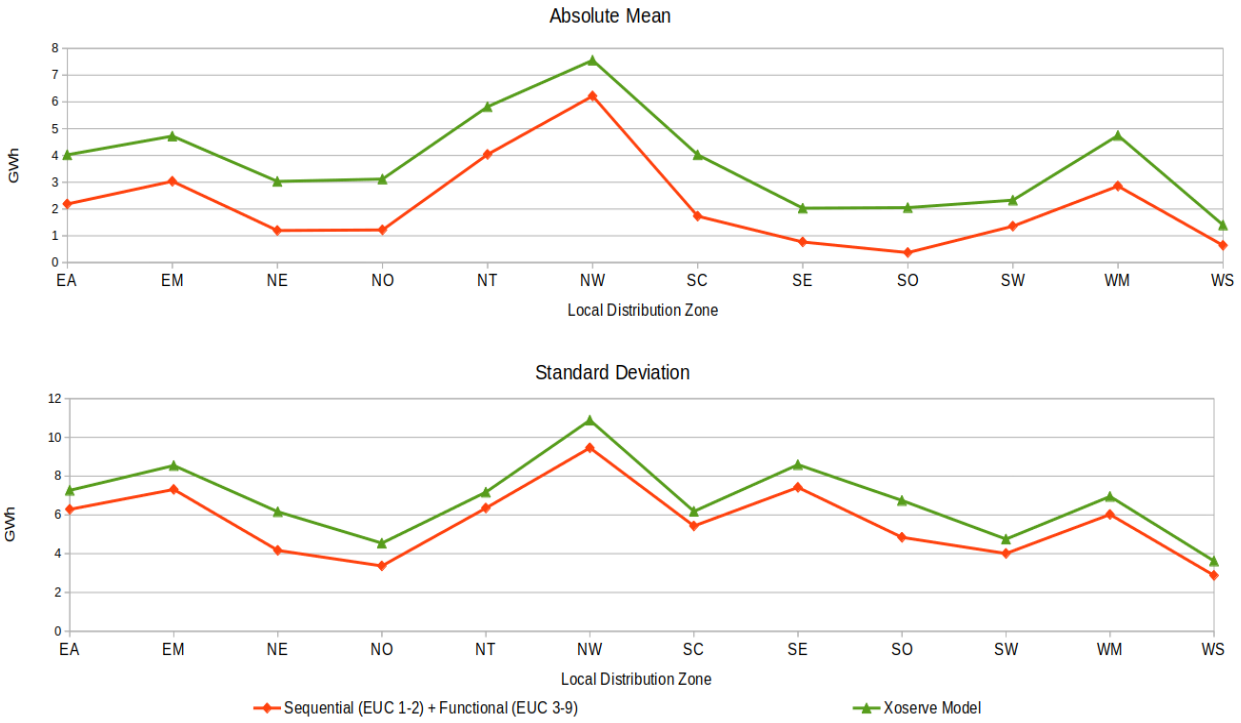



| Models Summary | |
|---|---|
| **Model** | Functional (EUC1-9) |
| **Segment** | 1 model total (all LDZs, all EUCs) |
| **Inputs** | Standard set |
| **Training** | GY 2013-2015 |
| **Testing** | GY 2016-2018 |

# Sequential/Functional combined NN outperforms the existing NDM model

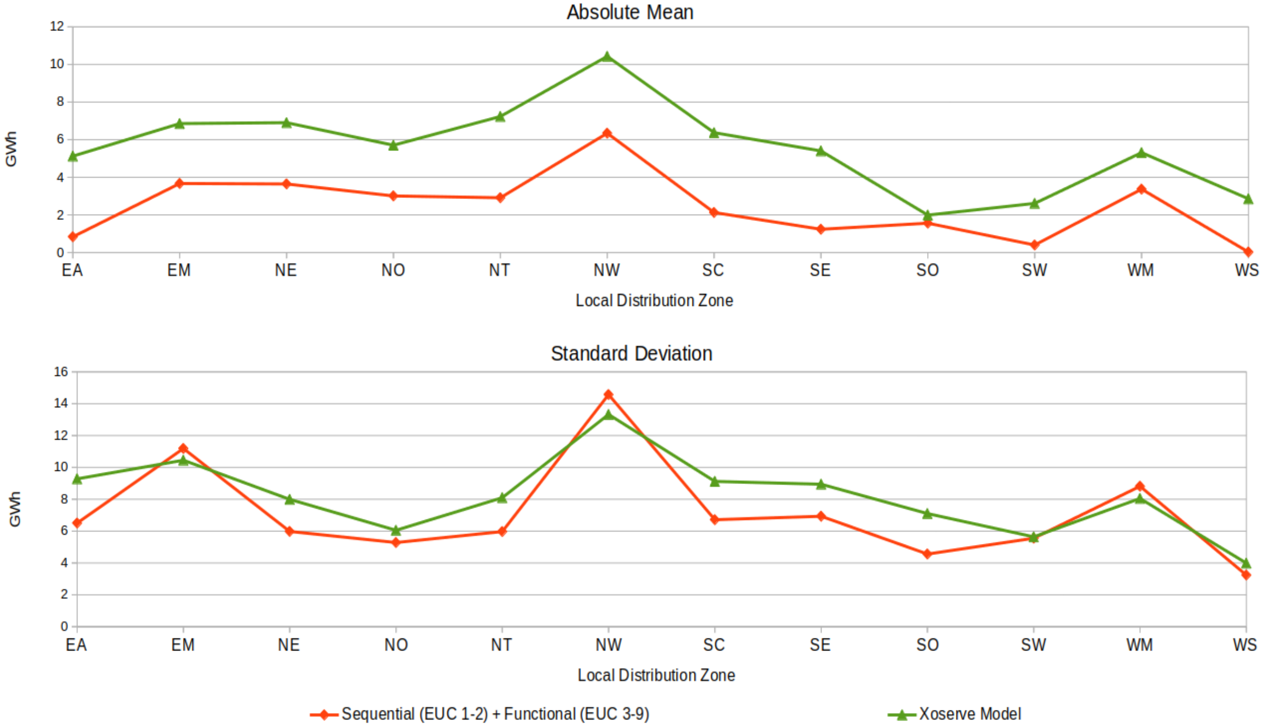
Metrics Comparison of ML Models – 2016

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |



# Sequential/Functional combined NN outperforms the existing NDM model

## Metrics Comparison of ML Models – 2017



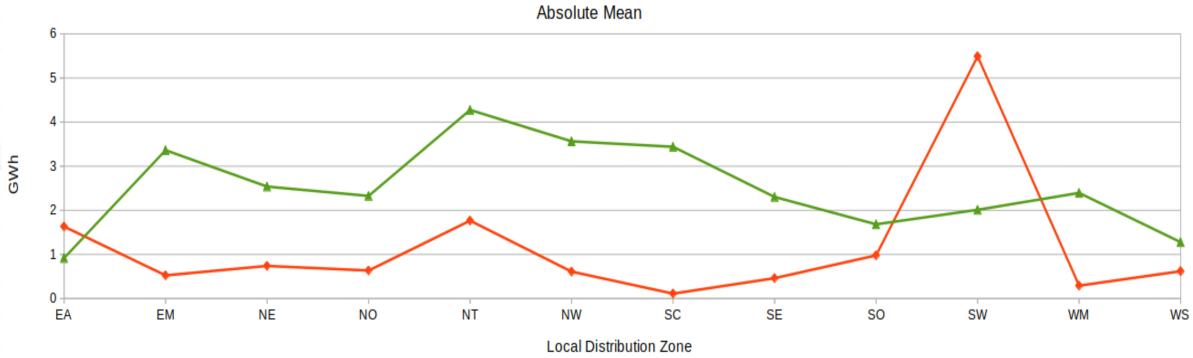

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |

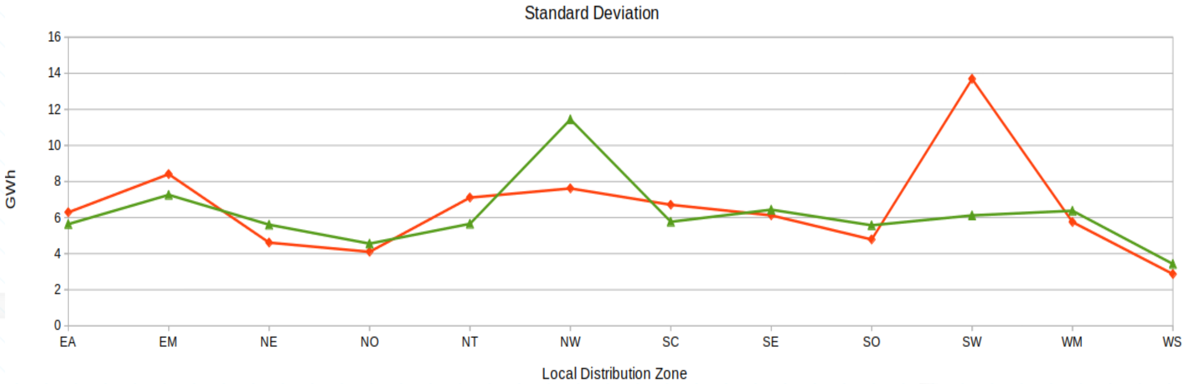
# Sequential/Functional combined NN outperforms the existing NDM model

## Metrics Comparison of ML Models – 2018

! Further investigation required to diagnose the unexpected performance of SW LDZ in 2018.

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |



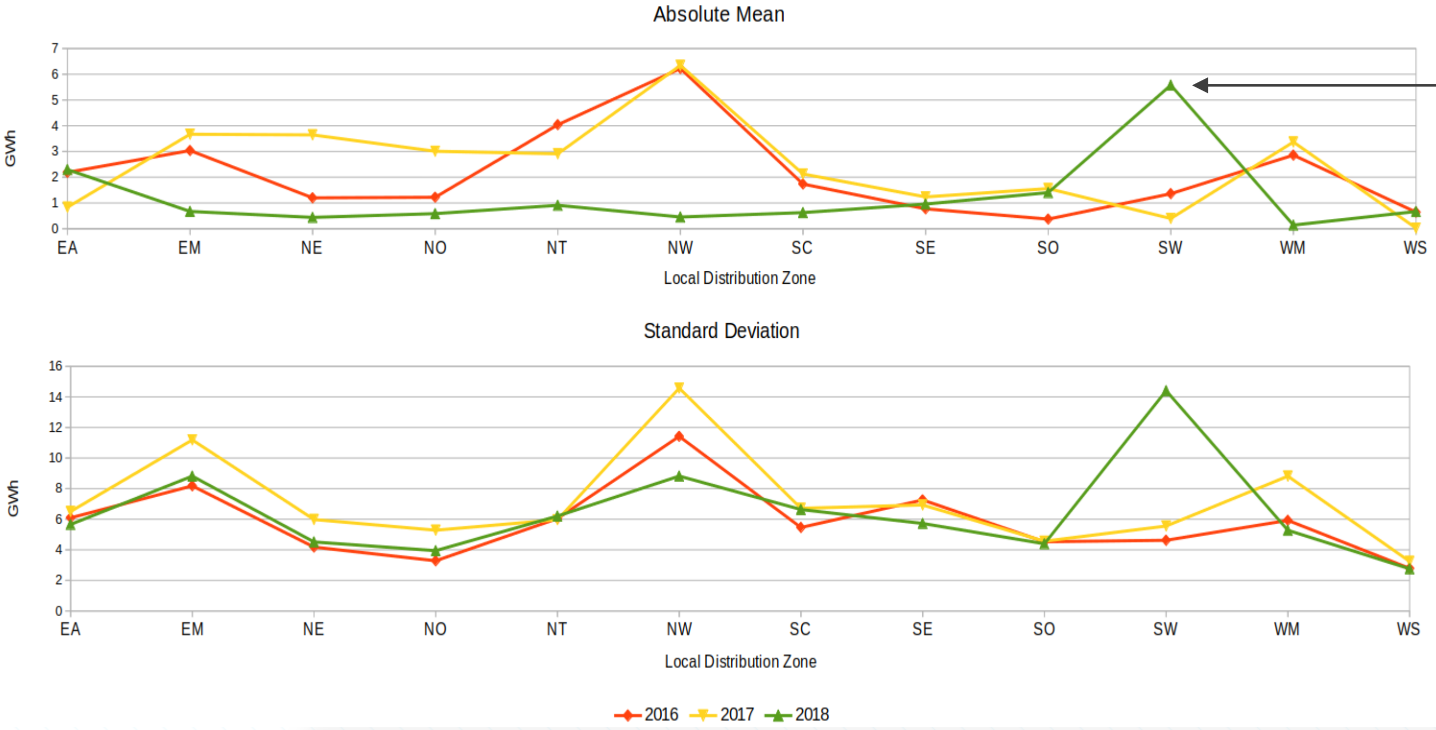
# 13.2.7 d

# Rerunning Neural Network to predict more recent Gas Years

# Rerunning Neural Network to predict more recent Gas Years

- Task:
  - Test the Sequential Neural Network Model on Gas years 2016-2018 with newly available data
- Motivation:
  - This task will confirm that the Sequential NN is general enough to predict well on any year.
    - It can help show that the model has not overfit to the 2016 test set, which can happen if hyperparameters (e.g. number of layers/neurons per layer, etc.) are tuned in order to yield the best results on a particular test set.
    - If the data in 2016 differs to what might usually be expected, and is therefore not representative of other years, the results previously reported may not reflect the performance seen when predicting for other years. This task can give assurance regarding the robustness of the Neural Net when tested on new unseen data.
- Conclusions:
  - GY2016 and GY2017 have equivalent performance
  - GY2018 is slightly worse on average: however, there is an anomalous SW LDZ result that impacts this. Excluding the S LDZ, 2018 also shows improvement over the existing model.
  - Also: note that the Xoserve NDM algorithm has the benefit of being retrained in GY2018 with more recent data compared to the Neural Network.
  - This demonstrates that the Neural Network has consistently better performance than the existing algorithm (SW LDZ excepted)

# Metrics Comparison of Sequential NN on each Test Year




Note that the performance of SW is much worse in 2018.

| Models Summary | |
|---|---|
| **Model** | Sequential (EUC1-2) + Functional (EUC3-9) |
| **Segment** | Sequential – 1 model per LDZ per EUC (1-2)<br>Functional – 1 model total (all LDZs, EUC3-9) |
| **Inputs** | Standard set |
| **Training** | Sequential – GY 2006-2015<br>Functional – GY 2013-2015 |
| **Testing** | GY 2016-2018 |