UIG Taskforce

13.1.4: Influence of Geographical Factors on Demand Estimation

Summary of Findings

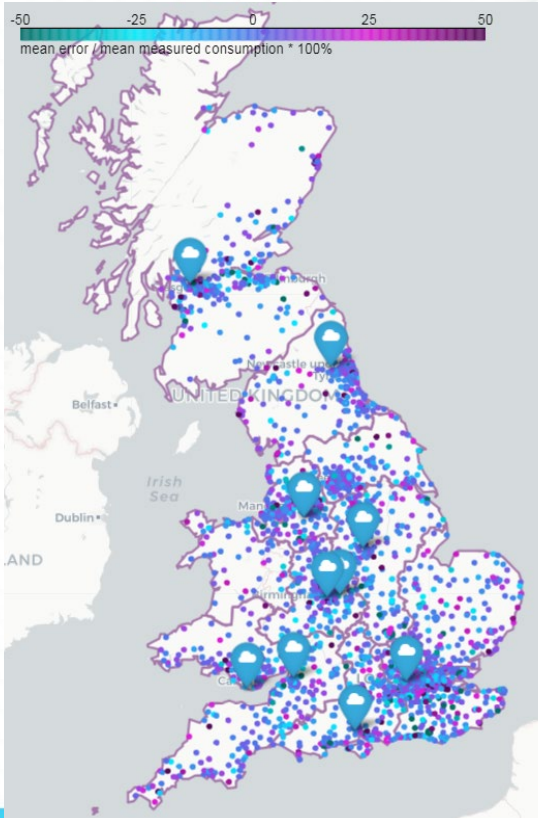
Findings Status	[Closed]
UIG Impact Peak Volatility %	N/A
UIG Impact Annual Average %	N/A
Confidence in Percentages	N/A

Area & Ref #	Accuracy of NDM Algorithm (Including EUC Definitions) - Influence of Geographical Factors on Demand Estimation (Ref #13.1.4)
UIG Hypothesis	There may be geographical trends on the sample data difference between the sample data set meter point level allocation and the actual measured daily usage (effectively the UIG for the Sample Sites) that can be easily visualised. Temperature is known to be a possible contributor to volatility: this analysis is expected to illustrate UIG is correlated to the distance from the LDZs weather station.
Data Tree References	WCF, CWV, ALP, DAF & AQ.

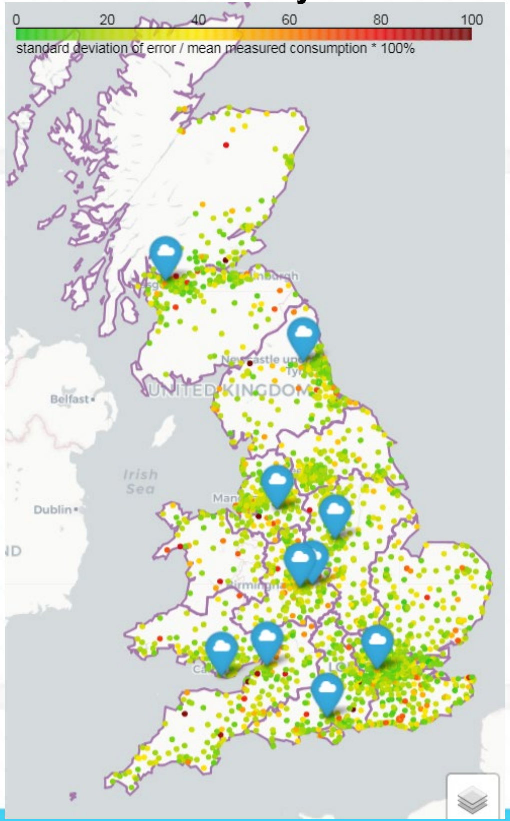
Findings	Approach to analysis
Geographically, the sample set does seem to be largely representative, though dense urban regions are slightly overrepresented. There is no obvious geographic pattern in either the base UIG or Sample Set UIG volatility. Regions where the model performs well are interspersed by clusters of sites where it does not perform as well. Typically, where the model performs well in one season it will also perform well throughout the year. Conversely the sites which do not perform well demonstrate consistent levels of variance throughout the year. This suggests that there are some regions which are well modelled and some which could be improved, but those improvements could be independent of geographic factors. This means that variance in the NDM Allocation model which can cause UIG does not appear to have a strong regional component, and sites where the modelled usage varies significantly from allocated usage could have other features (building type, building age, social demographic, microclimates etc.) which may influence demand but are not clear from a geographic cluster. The results suggest that there is no correlation between regions with higher levels of UIG and distance from weather stations, or obvious correlation to other geographical parameters (e.g., high altitude areas, distance to coast etc.) and that the NDM Sample data has appropriate regional coverage to capture demand patterns across the country.	Calculate the difference between the sample data set meter point level allocation and the actual usage. This was carried out for the gas year 2016 to 2017. We visually inspected the plots/maps to see if there are regions where the model performs particularly well or poorly. Check if regions consistently perform well or poorly over time.

Supporting Evidence (1/2) – Screenshot of the map

Sample set 'base' demand error



Sample set demand error 'volatility'

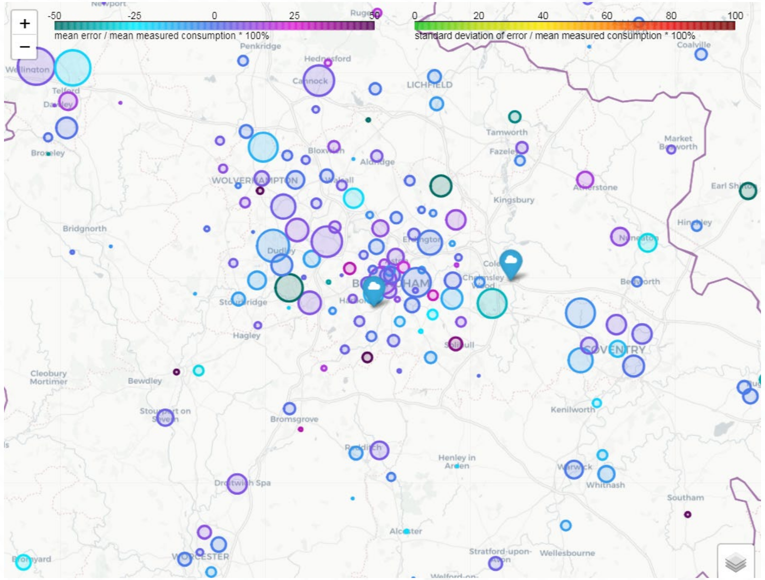


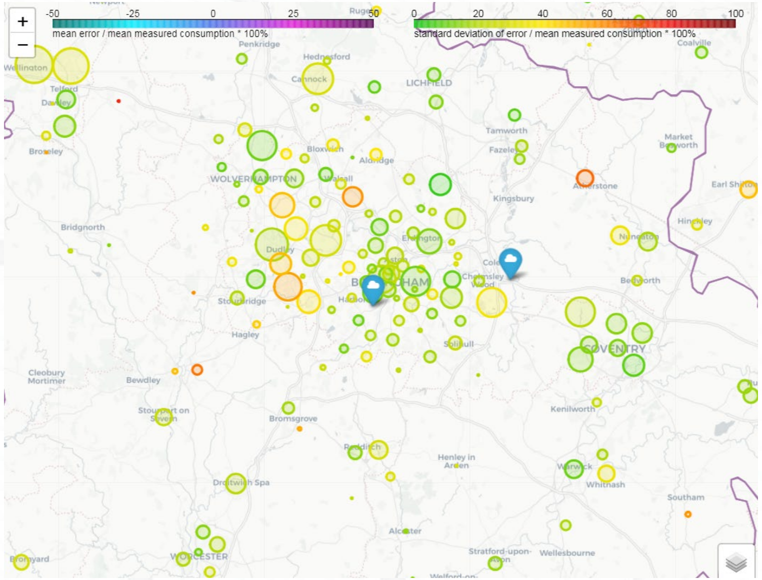
Plots for October to December 2016 shown opposite. Weather stations are indicated by the cloud icons, and approximate LDZ boundaries are shown.

Left is the base UIG expressed as average percentage difference between sample set allocation and actual usage. Right is sample set UIG volatility expressed as the standard deviation (i.e. the scale of the variability when comparing daily meter point level allocation against actual usage) as a percentage of average measured consumption .

There is no clear geographic distribution trend for either Base UIG or Volatility. The areas with higher levels of base UIG are not necessarily the most volatile and vice versa, suggesting there may be different local influences for the two components of UIG.

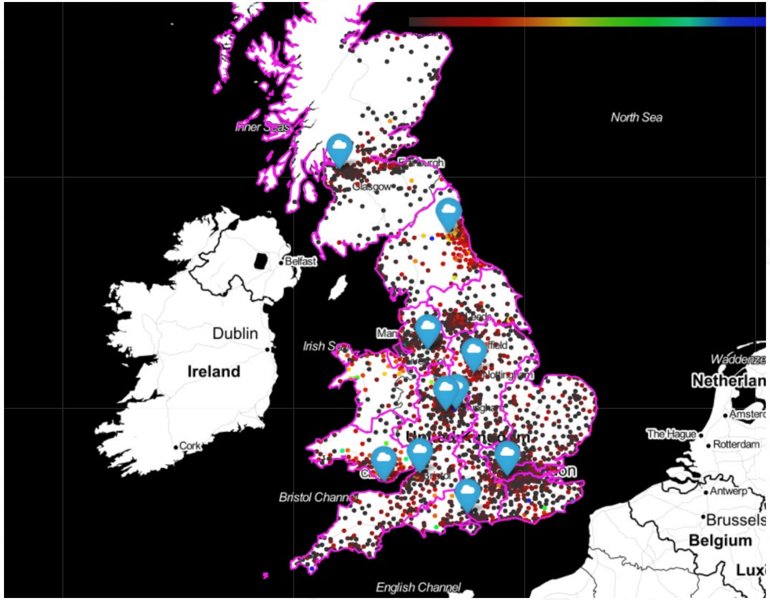
Supporting Evidence (2/2) – Screenshot of the map

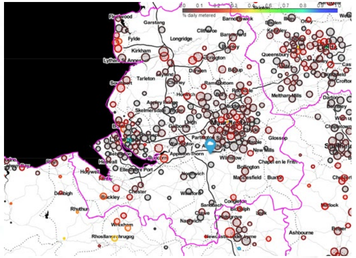


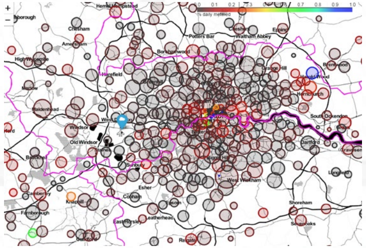


Close up of Birmingham area showing bubbles for each region, where the size of the bubble is the quantity of gas consumed that was measured by the sample meters.

Supporting Evidence (3/3) – Geographical representation in the sample set







The figures on this slide show the geographical distribution of meters in the sample set. Each point represents a UK post code, with its size representing the total number of meters in that postcode and its colour representing the fraction of meters in that postcode which are part of the sample. As on the previous slides, they have been created from the full AQ dataset for the UK in the 2016/17 gas year.

The main figure, of the whole UK, show that the geographical distribution of sample meters is largely uniform, although the Northeast of England is slightly overrepresented. As modelling is plotted by LDZ, this overrepresentation is unlikely to affect demand prediction accuracy.

The two bottom figures show zoom-ins of Liverpool and Manchester (*bottom left*) and London (*bottom right*). Both demonstrate that towns and cities tend to be overrepresented towards there centres in comparison to the outskirts. As usage profiles are likely to be different between these two regions, this overrepresentation could lead to modelling inaccuracies, but these inaccuracies are expected to be relatively insignificant and no further investigation is recommended at this stage.