DESC TWG

Shipper Data Validation

16th September 2015

Background

- NDM sample data numbers have been decreasing over time

- UNC allows transporters to acquire NDM sample data from third parties (i.e. smart metered data)

- Action DTW0502 was established to allow us to explore this further

- The analysis is required to review the suitability of the additional data and determine whether it would be possible to use it as part of the sample data.

XOserve

respect 〉 commitment 〉 teamwork

Analysis

- Data sets used – Xoserve AMR sample data against British Gas data
- Date range = Apr '14 to Mar '15
- Demand data in gas day
- 01B Domestic (residential) sites used in the analysis
- Aggregated demand by LDZ
- Validation rules had to be relaxed to allow for the British Gas data to pass validation.
  - Existing rule is to reject those MPRNs from the analysis if they have 15 or more days of missing data over the summer. We had to increase this threshold to 40 days as none of the British Gas data would have passed validation
- AMR = 2,835 MPRNs used in the analysis
- British Gas = 2,749 MPRNs used in the analysis

XOserve

respect ⟩ commitment ⟩ teamwork

Tests used to compare data sets

- ## F-test
  - Firstly, an F-test was performed to compare the variability between the two data sets (Xoserve AMR vs British Gas)

- ## T-test
  - A T-test was then carried out to determine if the two sets of data are significantly different from each other

XOserve

respect ❭ commitment ❭ teamwork

Summary of Results

LDZ	F Test Equal Variances	T Test Equal Means	No. of sites AMR	No. of sites BG
EA	✖	✖	261	443
EM	✖	✖	241	329
NE	✖	✖	254	196
NO	✖	✖	221	157
NT	✓	✓	233	244
NW	✖	✖	225	270
SC	✖	✖	224	210
SE	✖	✖	227	89
SO	✖	✖	245	146
SW	✖	✖	234	162
WM	✖	✖	244	370
WS	✖	✖	226	133

LDZ NT (F and T-Test) Results

```
Two sample t-test comparing AMR to British Gas - LDZ NT


                              The TTEST Procedure

                            Variable:  Agg_Demand


          Shipper          N       Mean      Std Dev      Std Err      Minimum       Maximum

            AMR            364     10276.2     7008.3       367.3       2226.5        25297.7
            BG             364     10536.1     7369.3       386.3       2371.5        26290.9
            Diff (1-2)             -259.9      7191.1       533.0

Shipper          Method               Mean        95% CL Mean        Std Dev        95% CL Std Dev

AMR                                  10276.2     9553.8  10998.5      7008.3        6533.5   7558.1
BG                                   10536.1     9776.5  11295.7      7369.3        6870.1   7947.5
Diff (1-2)       Pooled              -259.9     -1306.4   786.6       7191.1        6839.5   7581.1
Diff (1-2)       Satterthwaite       -259.9     -1306.4   786.6

                 Method           Variances        DF      t Value      Pr > |t|

                 Pooled           Equal            726      -0.49        0.6260
                 Satterthwaite    Unequal       724.18      -0.49        0.6260

                            Equality of Variances

                 Method        Num DF      Den DF      F Value     Pr > F

                 Folded F        363         363        1.11       0.3390
```
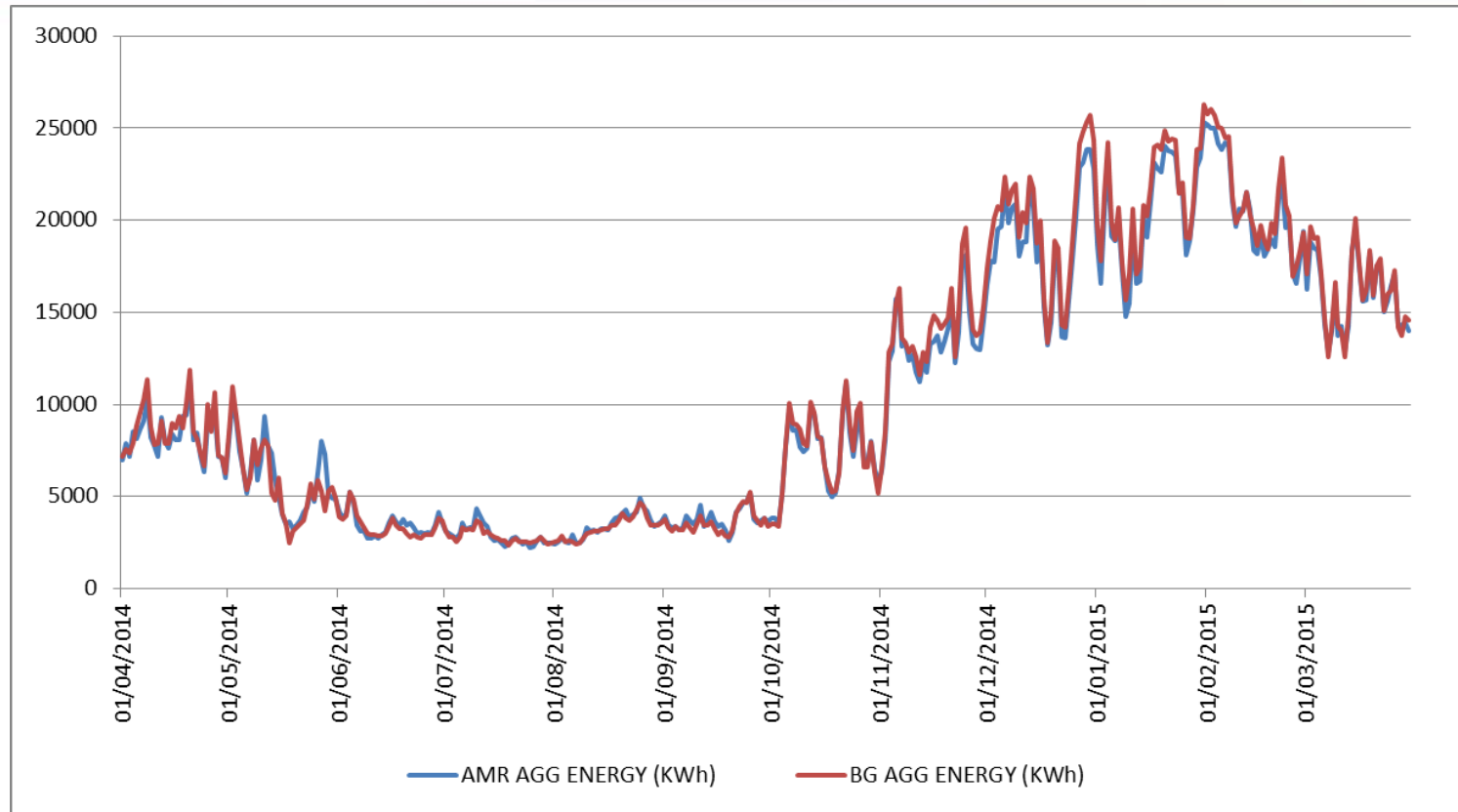
LDZ – NT (Equal variances and means)

LDZ SE (F and T-Test) - Results

```
Two sample t-test comparing AMR to British Gas - LDZ SE

                       The TTEST Procedure

                    Variable:  Agg_Demand
```

Shipper	N	Mean	Std Dev	Std Err	Minimum	Maximum
AMR	364	9713.2	6916.4	362.5	2006.1	25102.5
BG	364	3879.5	2908.9	152.5	714.5	10336.8
Diff (1-2)		5833.7	5305.6	393.3		

Shipper	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
AMR		9713.2	9000.3	10426.1	6916.4	6447.8	7459.0
BG		3879.5	3579.6	4179.3	2908.9	2711.8	3137.1
Diff (1-2)	Pooled	5833.7	5061.7	6605.8	5305.6	5046.2	5593.3
Diff (1-2)	Satterthwaite	5833.7	5061.0	6606.5			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	726	14.83	<.0001
Satterthwaite	Unequal	487.52	14.83	<.0001

```
                    Equality of Variances
```

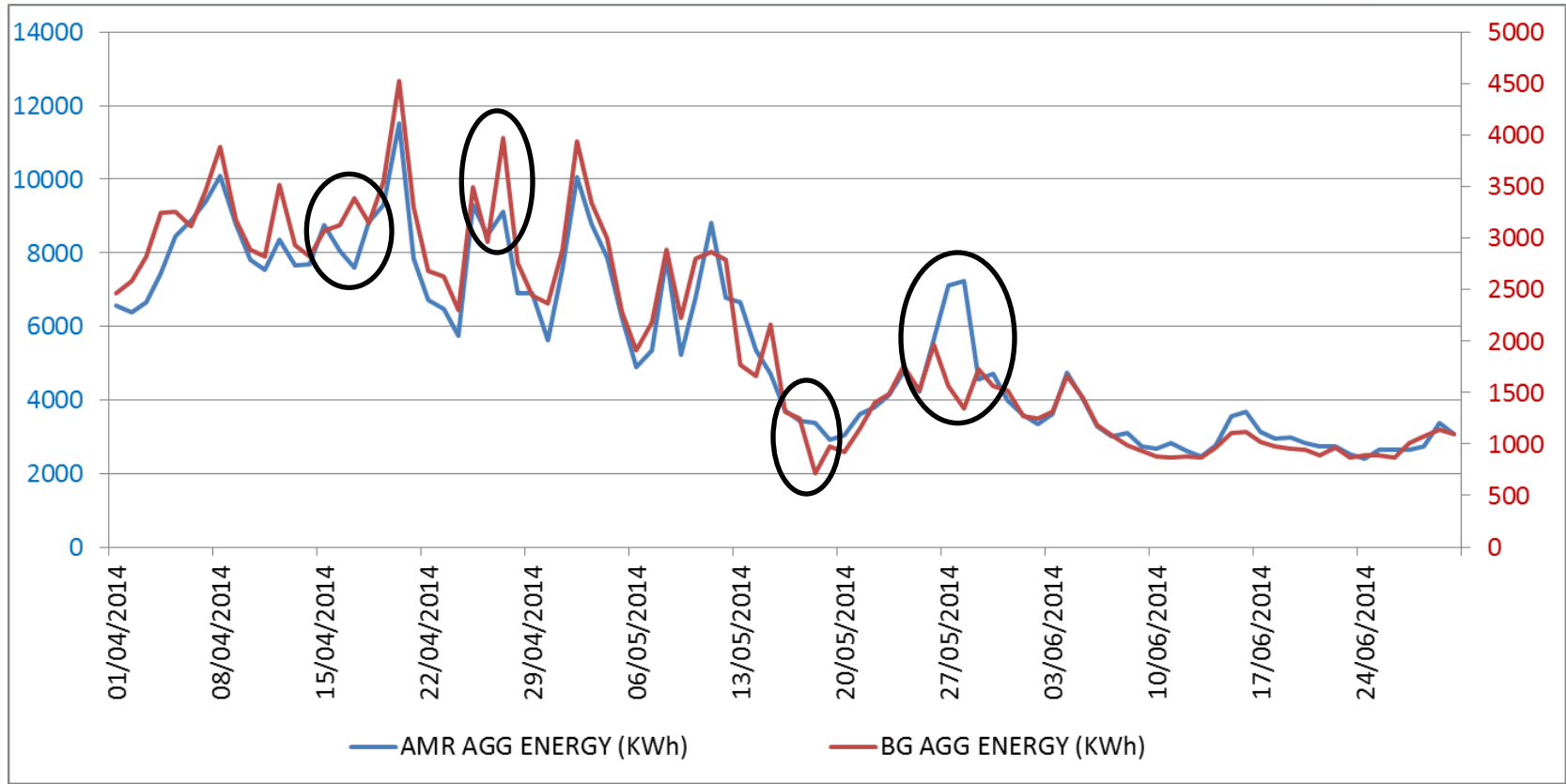| Method	Num DF	Den DF	F Value	Pr > F
Folded F	363	363	5.65	<.0001

respect ❯ commitment ❯ teamwork
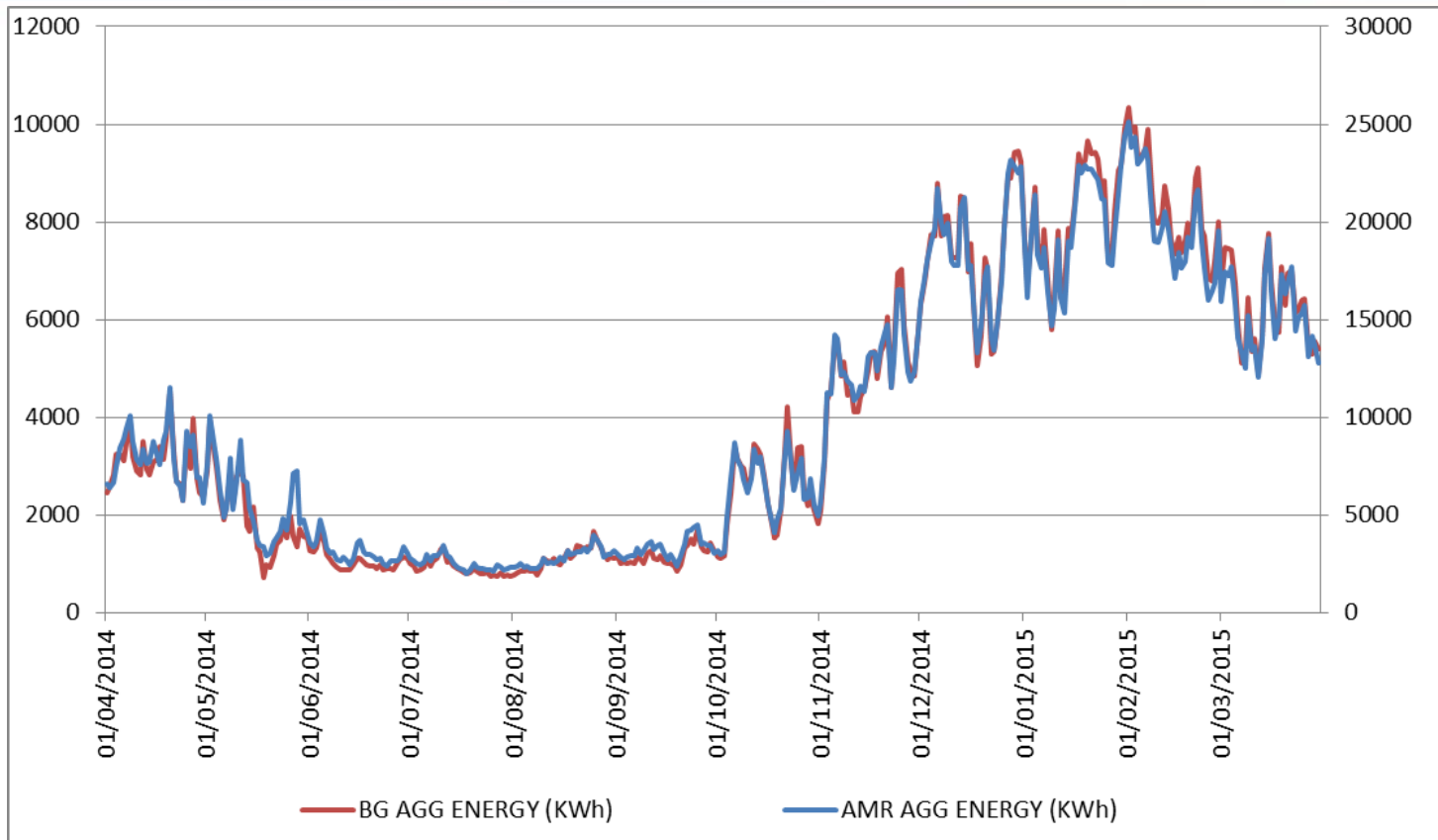
LDZ – SE (April to June)

Highlighted above shows the days where no data was provided in the British Gas Sample. These days had to be infilled, which appears to have affected the shape of the British Gas consumption data. This was applicable on the same days for all other LDZ's also.

LDZ – SE (an example of unequal means and variance)

This graph shows that although the demand levels differ, the overall consumption shape looks very similar. This is the same for all other LDZ's.

Next Steps

- The results from the t-test indicate that the means for the AMR sample data and the British Gas data are statistically different. This could be due to:
    - the difference in sample sizes
    - missing records in the data set provided by British Gas resulted in large amounts of data having to be infilled
    - sites being used in the analysis that would not have passed validation under normal circumstances
- Despite the issues above, when analysing the consumption pattern of the British Gas data in comparison to the AMR data, they appear to be very similar in shape
- Further work to be carried out to compare the weather sensitivity of British Gas data in comparison to the AMR data

- Continue to explore the Eon data

xoserve

respect › commitment › teamwork