



DESC 13TH Feb 2018

Action: DESC1201 – review of sample sizes

Objective

- At the previous DESC meeting, an action (DESC1201) was raised for Xoserve to determine statistically what the optimum sample size should be by EUC and LDZ.
- Mod 0644 proposals include new EUC definitions. If these are implemented then new sample size targets will need to be set. This piece of analysis takes into consideration the possibility of these new EUC bands.

Background

- Current sample sizes have shown a declining trend over recent years. In 2017, Xoserve managed sample numbers have fallen below target in 11 LDZs.
- Smart meter rollout appears to have made a significant impact to the decline. In gas year 16/17 there were 281 supply point losses.
- There does not appear to be any statistical method used in setting the target sample numbers previously.

Band 1 and 2

- The table below shows the current sample target sizes for Band 1 (domestic) & 2 (domestic and I&C) sites and compares them to the actual sample sizes for the last 2 years.

LDZ	EUC Band 1 (XMS)			EUC Band 2 (XMS & TMS)		
	Target Sample Size	Actual Sample Size		Target Sample Size	Actual Sample Size	
		Oct-16	Oct-17		Oct-16	Oct-17
SC	245	235	201	140	168	152
NO	240	263	221	135	127	120
NW	250	248	225	165	176	165
NE	250	246	220	140	129	125
EM	255	260	222	145	222	204
WM	250	246	221	150	148	149
WS	255	251	232	115	98	88
EA	250	271	251	150	227	209
NT	250	238	211	220	249	227
SE	245	233	216	160	234	238
SO	250	250	229	120	195	185
SW	250	250	234	135	167	156
WN	0	0	0	10	4	3

- Band 1 has shown a decreasing trend in actual sample sizes over recent years.
- Band 2 appears to meet sample number requirements in 8/13 LDZs.

Note: Band 2 totals include both Xoserve and Transporter managed sample sites. Target sample size allows for fallout (between 10-20%) through the validation process – so is not the number of sites needed for modelling.

Population Size

EUC Band	Population Size												
	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	2,051,653	1,273,638	2,860,996	1,453,691	2,456,332	2,120,073	891,366	1,991,673	2,367,430	2,640,895	1,765,080	1,621,632	265,300
2	16,779	9,811	22,819	13,169	18,995	17,282	5,814	16,127	27,119	23,399	15,652	12,632	2,169
3	4,655	2,467	5,125	2,796	4,230	4,133	1,288	3,562	6,196	4,416	3,540	2,753	497
4	2,068	884	2,032	1,069	1,654	1,771	562	1,479	2,922	1,683	1,465	1,090	225
5	511	239	560	276	469	492	154	336	761	377	287	217	68
6	163	99	163	110	168	145	47	111	197	90	99	99	23
7	54	24	74	37	85	50	26	33	37	27	31	41	11
8	22	13	18	11	30	25	9	14	14	16	15	6	1

EUC Band	% of the population that is part of the sample (Oct 2017)												
	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	0.01%	0.02%	0.01%	0.02%	0.01%	0.01%	0.03%	0.01%	0.01%	0.01%	0.01%	0.01%	-
2	0.91%	1.22%	0.72%	0.95%	1.07%	0.86%	1.51%	1.30%	0.84%	1.02%	1.18%	1.23%	0.14%
3	5.11%	5.15%	2.50%	5.01%	4.18%	3.31%	2.72%	5.19%	3.08%	6.16%	5.34%	4.61%	4.43%
4	19.63%	28.28%	13.88%	33.49%	13.91%	13.04%	14.23%	16.16%	10.30%	25.73%	22.46%	11.83%	11.56%
5	50.49%	51.46%	27.86%	55.07%	29.42%	33.94%	27.92%	28.87%	20.11%	47.48%	48.08%	31.80%	33.82%
6	61.35%	55.56%	33.74%	52.73%	35.12%	46.21%	51.06%	36.04%	24.87%	65.56%	55.56%	46.46%	34.78%
7	59.26%	41.67%	24.32%	78.38%	43.53%	42.00%	23.08%	30.30%	24.32%	62.96%	45.16%	46.34%	18.18%
8	50.00%	61.54%	33.33%	54.55%	50.00%	56.00%	33.33%	14.29%	42.86%	43.75%	60.00%	66.67%	-

(Population sizes as of Jan 2018)

Population vs. Sample Size

All Data		Population Size												
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
1	2,051,653	1,273,638	2,860,996	1,453,691	2,456,332	2,120,073	891,366	1,991,673	2,367,430	2,640,895	1,765,080	1,621,632	265,300	
2	16,779	9,811	22,819	13,169	18,995	17,282	5,814	16,127	27,119	23,399	15,652	12,632	2,169	
3	4,655	2,467	5,125	2,796	4,230	4,133	1,288	3,562	6,196	4,416	3,540	2,753	497	
4	2,068	884	2,032	1,069	1,654	1,771	562	1,479	2,922	1,683	1,465	1,090	225	
5	511	239	560	276	469	492	154	336	761	377	287	217	68	
6	163	99	163	110	168	145	47	111	197	90	99	99	23	
7	54	24	74	37	85	50	26	33	37	27	31	41	11	
8	22	13	18	11	30	25	9	14	14	16	15	6	1	

All Data		Current Sample Size												
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
1	245	240	250	250	255	250	255	250	250	245	250	250	-	
2	140	135	165	140	145	150	115	150	220	160	120	135	10	
3	110	95	150	80	130	80	75	120	185	165	85	100	20	
4	425	270	450	340	445	335	175	445	480	520	360	260	50	
5	420	235	445	250	340	415	130	275	515	355	280	240	55	
6	116	83	186	104	148	177	61	110	157	94	85	90	20	
7	34	24	45	45	64	66	21	31	26	20	35	32	7	
8	14	16	25	10	25	32	10	8	5	10	8	8	3	

Note: cells highlighted in red show where the sample size requirement is larger than the population size

Band 3 to 8 – Sample Size

	Target Sample Size (TMS)						Actual Sample Size 2016 (TMS)						Actual Sample Size 2017 (TMS)					
	EUC						EUC						EUC					
	3	4	5	6	7	8	3	4	5	6	7	8	3	4	5	6	7	8
LDZ																		
SC	110	425	420	116	34	14	234	424	273	101	30	13	238	406	258	100	32	11
NO	95	270	235	83	24	16	122	294	134	54	13	10	127	250	123	55	10	8
NW	150	450	445	186	45	25	165	294	170	56	19	6	128	282	156	55	18	6
NE	80	340	250	104	45	10	159	384	180	64	28	11	140	358	152	58	29	6
EM	130	445	340	148	64	25	202	248	162	71	39	13	177	230	138	59	37	15
WM	80	335	415	177	66	32	147	265	180	72	24	14	137	231	167	67	21	14
WS	75	175	130	61	21	10	42	85	49	22	7	4	35	80	43	24	6	3
EA	120	445	275	110	31	8	223	265	98	45	14	4	185	239	97	40	10	2
NT	185	480	515	157	26	5	214	310	194	50	10	6	191	301	153	49	9	6
SE	165	520	355	94	20	10	286	434	191	58	20	3	272	433	179	59	17	7
SO	85	360	280	85	35	8	203	335	141	57	14	9	189	329	138	55	14	9
SW	100	260	240	90	32	8	139	150	78	47	18	4	127	129	69	46	19	4
WN	20	50	55	20	7	3	24	33	22	9	2	0	22	26	23	8	2	0

Note: the cells highlighted in red show where the sample size does not meet the target sample size

- Band 3 appears to be meeting the current target sample size.
- Although it appears to show that the majority of EUC bands are not meeting current targets, the following slide explores how well the sample performs in the modelling process.

Modelling Results vs. Sample Size

Model R² - (All days) Single Year Modelling 2017/18

No of supply points used in the modelling

EUC Band	SC	NO	NW/WN	NE	EM	WM	WS	EA	NT	SE	SO	SW
1	98%	98%	98%	97%	98%	99%	98%	99%	99%	99%	99%	99%
	188	179	185	189	208	201	185	233	194	196	218	201
2	97%	96%	95%	96%	97%	96%	96%	96%	97%	97%	97%	97%
	132	105	141	117	169	131	81	173	182	170	162	147
3	97%	96%	94%	96%	97%	95%	95%	96%	97%	98%	97%	97%
	172	114	153	132	166	119	32	181	173	210	152	111
4	96%	97%	96%	96%	98%	96%	97%	98%	98%	98%	98%	97%
	319	231	265	320	221	215	73	217	239	299	275	122
5	97%	97%	98%	97%	98%	98%	97%	98%	98%	99%	98%	98%
	239	114	158	145	131	144	33	79	147	145	104	70
6	97%	97%	98%	94%	98%	98%	97%	96%	98%	97%	96%	97%
	84	60	92	77	79	80	21	46	56	46	48	54
	SC	NO	NW/WN	NE	EM	WM	EA	NT	SE/SO	WS/SW	-	-
7 & 8	77%	89%	95%	93%	95%	95%	87%	97%	94%	91%	-	-
	36	38	92	70	93	74	42	36	45	66	-	-

Note: the cells highlighted are where the R² are lower than 90%. This appears to be in the less weather sensitive models.

Selecting a Sample Size

- Key terms:
 - Population – the total no. of people/consumers in the group
 - Sample – a group that represents the population
 - Margin of Error – a percentage that tells you how much you can expect the results to reflect the views of the overall population. The smaller the margin of error, the closer you are to having an accurate answer at a given confidence interval
 - Confidence Level – how confident you want to be that the actual mean falls within your confidence interval. The most common confidence intervals are 90% confident, 95% confident and 99% confident. Your confidence level corresponds to a z-score.
 - z-score - the number of standard deviations a given proportion is away from the mean.

Selecting a Sample Size cont...

The formula used to calculate the sample sizes is as follows:

$$\text{Sample Size} = \frac{\frac{z^2 \times p(1 - p)}{e^2}}{1 + \left(\frac{z^2 \times p(1 - p)}{e^2 N}\right)}$$

Population size = N

Margin of error = e

z-score = z

standard deviation = p

The following tables explore the suggested sample sizes at both the 95% and 90% confidence levels (with 5% Margin of Error)

Suggested Sample Sizes – All Data

All data														All data														
Sample Size														Target Sample Size - 95%														
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
1	245	240	250	250	255	250	255	250	250	245	250	250	-	1	384	384	384	384	384	384	384	384	384	384	384	384	384	384
2	140	135	165	140	145	150	115	150	220	160	120	135	10	2	376	370	378	373	377	376	360	375	379	378	375	373	326	
3	110	95	150	80	130	80	75	120	185	165	85	100	20	3	355	332	357	338	352	351	296	347	362	353	347	337	217	
4	425	270	450	340	445	335	175	445	480	520	360	260	50	4	324	268	323	283	312	316	228	305	340	313	304	284	142	
5	420	235	445	250	340	415	130	275	515	355	280	240	55	5	219	147	228	161	211	216	110	179	255	190	164	139	58	
6	116	83	186	104	148	177	61	110	157	94	85	90	20	6	114	79	114	86	117	105	42	86	130	73	79	79	22	
7	34	24	45	45	64	66	21	31	26	20	35	32	7	7	47	23	62	34	70	44	24	30	34	25	29	37	11	
8	14	16	25	10	25	32	10	8	5	10	8	8	3	8	21	13	17	11	28	23	9	14	14	15	14	6	1	

Confidence level = 95%

Cells highlighted red show where the current sample size does not meet the suggested sample size.

Suggested Sample Sizes – All Data

All data														All data														
Sample Size														Target Sample Size - 90%														
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
1	245	240	250	250	255	250	255	250	250	245	250	250	-	1	271	271	271	271	271	271	271	271	271	271	271	271	271	270
2	140	135	165	140	145	150	115	150	220	160	120	135	10	2	266	263	267	265	267	266	259	266	268	268	266	265	241	
3	110	95	150	80	130	80	75	120	185	165	85	100	20	3	256	244	257	247	254	254	224	251	259	255	251	246	175	
4	425	270	450	340	445	335	175	445	480	520	360	260	50	4	239	207	239	216	233	235	183	229	248	233	228	217	123	
5	420	235	445	250	340	415	130	275	515	355	280	240	55	5	177	127	182	137	172	175	98	150	200	158	139	120	54	
6	116	83	186	104	148	177	61	110	157	94	85	90	20	6	102	72	102	78	104	94	40	79	114	68	72	72	21	
7	34	24	45	45	64	66	21	31	26	20	35	32	7	7	45	22	58	33	65	42	24	29	33	25	28	36	11	
8	14	16	25	10	25	32	10	8	5	10	8	8	3	8	20	12	17	11	27	23	9	13	13	15	14	6	1	

Confidence level = 90%

Cells highlighted red show where the current sample size does not meet the suggested sample size.

Suggested Sample Size – Domestic

Domestic		Population Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	2,014,503	1,248,986	2,796,538	1,417,383	2,403,832	2,074,276	873,133	1,951,171	2,306,299	2,584,837	1,727,445	1,584,555	258,460
2	2,990	1,624	3,643	2,502	2,995	2,549	720	3,197	6,734	5,509	2,591	1,859	214

Domestic		Sample Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	384	384	384	384	384	384	384	384	384	384	384	384	384
2	340	311	348	333	340	334	251	343	363	359	335	318	137

Confidence level = 95%

Domestic		Sample Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	271	271	271	271	271	271	271	271	271	271	271	271	270
2	248	232	252	244	248	245	197	249	260	258	245	236	119

Confidence level = 90%

Suggested Sample Size – Non Domestic

Non Domestic		Population Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	37,150	24,652	64,458	36,308	52,500	45,797	18,233	40,502	61,131	56,058	37,635	37,077	6,840
2	13,789	8,187	19,176	10,667	16,000	14,733	5,094	12,930	20,385	17,890	13,061	10,773	1,955

Non Domestic		Sample Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	380	378	382	380	381	381	376	381	382	382	380	380	364
2	374	367	377	371	375	374	357	373	377	376	373	371	321

Confidence level = 95%

Non Domestic		Sample Size											
EUC Band	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN
1	269	268	269	269	269	269	267	269	269	269	269	269	260
2	265	262	267	264	266	266	257	265	267	267	265	264	238

Confidence level = 90%

Suggested Sample Size – Pre-Payment Domestic

Domestic		Population Size												
PrePayment Dom	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
EUC Band 1	284,058	149,270	382,136	153,519	256,565	253,004	128,514	180,058	293,708	288,751	125,927	143,831	40,591	

Domestic		Sample Size												
PrePayment Dom	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
EUC Band 1	384	383	384	383	384	384	383	383	384	384	383	383	381	

Confidence level = 95%

Domestic		Sample Size												
PrePayment Dom	SC	NO	NW	NE	EM	WM	WS	EA	NT	SE	SO	SW	WN	
EUC Band 1	270	270	270	270	270	270	270	270	270	270	270	270	269	

Confidence level = 90%

DESC Thoughts

- Do we continue with the current target sample sizes or use the suggested sample size numbers (95% or 90% confidence level)?
- If DESC would like to boost the current sample size then there will be a requirement for third party data as Xoserve are struggling to recruit and retain sample sites.
 - By 3rd April 2018 we require additional NDM sample data covering the period from 24/03/2017 to 01/04/2018 inclusive for use in the Spring 2018 modelling (Band 01B in particular)
 - Full requirements document is available on the Joint Office website (<https://www.gasgovernance.co.uk/desc>)