Anglian Water: Base modelling consultation Response Appendix

Contents

Population thresholds to account for economies of scale in sewage treatment

Table A.1 – Ofwat's proposed SWT models including alternative SWT2 models with different population thresholds at Sewage Treatment Works (STWs)

Model	SWT1	SWT2-100K	SWT2-150K	SWT2-175K	SWT2-200K	SWT2-250K	SWT3
Load (log)	0.653*** (0.000)	0.723*** (0.000)	0.778*** (0.000)	0.798*** (0.000)	0.758*** (0.000)	0.713*** (0.000)	0.788*** (0.000)
Load treated with ammonia consent ≤3mg/l	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)
Load treated in size bands 1 to 3 (%)	0.029 (0.211)						
Load treated in STWs ≥ 100,000 people (%)		-0.008*** (0.007)					
Load treated in STWs ≥ 150,000 people (%)			-0.009*** (0.000)				
Load treated in STWs ≥ 175,000 people (%)				-0.010*** (0.000)			
Load treated in STWs ≥ 200,000 people (%)					-0.009*** (0.000)		
Load treated in STWs ≥ 250,000 people (%)						-0.007*** (0.003)	
Weighted average treatment size (log)							-0.242*** (0.000)
Constant	-3.734*** (0.004)	-4.072*** (0.000)	-4.814*** (0.000)	-5.075*** (0.000)	-4.626*** (0.000)	-4.142*** (0.000)	-3.001*** (0.000)
Model Robustness Tests and Additional Diagnostics							
Adjusted 2	0.854	0.869	0.896	0.895	0.901	0.899	0.911
RESET	0.056	0.272	0.192	0.443	0.499	0.463	0.849
VIF	5.337	5.347	4.676	4.702	4.439	4.443	4.339
Pooling	0.999	1	0.998	0.997	0.996	0.998	0.997
Normality	0.024	0.221	0.11	0.254	0.181	0.057	0.064
Heteroskedasticity	0.417	0.764	0.541	0.83	0.988	0.927	0.865
LM	0	0	0	0	0	0	0
Efficiency Score Distribution							
Minimum	0.82	0.87	0.88	0.89	0.91	0.91	0.91
Maximum	1.50	1.41	1.23	1.27	1.31	1.36	1.24
Range	0.68	0.53	0.35	0.37	0.39	0.45	0.33
Upper Quartile	0.94	0.91	0.95	0.91	0.92	0.94	0.95
Sensitivity Analysis							
Removal most efficient company	A	G	G	G	G	G	G
Removal least efficient company	A	G	G	G	G	G	G
Removal first year	G	G	G	G	G	G	G
Removal last year	G	G	G	G	G	G	G

As demonstrated by the various metrics across the model robustness tests and additional diagnostics, models that replace the 'load treated in STWs ≥ 100,000 people' variable with a higher population threshold perform better statistically. This is apparent in the material improvements in adjusted R2 (by three percentage points for all the different thresholds presented above). In addition, the range between the most and the least efficient company reduces materially, from 0.53 in Ofwat's SWT2 model to 0.35-0.39 for models using a population threshold of 150k, 175k or 200k.

As the models exhibit a tighter efficiency range and a higher adjusted R2, these models better capture the cost differences between companies than the models with the threshold chosen by Ofwat. Overall, it is clear that the models that incorporate an economies of scale variable at large STWs at higher thresholds than Ofwat's 100,000 are materially more robust and reliable.

Table A.2 - Distribution of efficiency scores and rankings across population thresholds

Model	SWT2-100K	SWT2-150K	SWT2-175K	SWT2-200K	SWT2-250K
Efficiency Scores					
ANH	1.05	1.00	0.98	0.98	0.99
NES	0.99	1.08	1.14	1.12	1.03
NWT	1.10	1.10	1.10	1.05	1.08
SRN	1.41	1.23	1.27	1.31	1.36
SVH	0.90	0.88	0.89	0.92	0.93
SWB	0.95	0.95	0.90	0.91	0.95
TMS	0.87	0.91	0.93	0.93	0.91
WSH	1.18	1.20	1.20	1.17	1.12
WSX	0.89	0.96	0.90	0.92	0.93
YKY	1.10	1.14	1.13	1.14	1.11
Range	0.53	0.35	0.37	0.39	0.45

Table A.3 – Variations in efficiency scores across the different SWT2 models presented in Table A.1 above.

Company	Range
SRN	0.18
NES	0.15
WSH	0.08
ANH	0.07
WSX	0.07
TMS	0.06
NWT	0.05
SWB	0.05
SVH	0.04
YKY	0.04

While we have shown that the SWT2 models with alternative thresholds perform better, one major note to address is the substantial variation in efficiency scores within the thresholds themselves. Such ranges in scores, as exhibited by companies such as Southern and Northumbrian, are not intuitive and illustrates the main drawback of using such arbitrary thresholds. As "Data analysis shows that unit costs continue to fall as the size of STWs increase." (p. 38 of the consultation), it is not intuitive to retain a single 'step change', beyond which economies of scale starts to occur. Choosing one particular threshold over another implies randomly favours some companies and penalise others without any particular operational justification behind it and creates these significant swings in efficiency scores for individual companies.

Therefore, we strongly encourage and favour the use of the WATS variable as it accounts for different thresholds and most importantly, accounts for the continuous relationship between unit costs and the size of STWs. We do not consider that arbitrary thresholds, imposing 'step-like' change in unit costs represent an appropriate alternative when a continuous relationship is now available. This continuous relationship is perfectly illustrated below as we can see that the average unit cost in STWs of size 25,000-250,000 is about 90% higher than the average unit cost in STWs of size greater than a population equivalent of 1,000,000. Only the WATS is able to capture such a decrease in unit cost as the size of STWs increase.

Table A.4 Average unit cost per STWs size in 2019

New Band	Population equivalent	Weighted average unit costs (£/p.e.)
Bands 1–3	0–2,000	**£2.86**
Band 4	2,000–10,000	**£1.19**
Band 5	10,000–25,000	**£0.84**
Band 6	25,000–250,000	**£0.59**
Band 7	250,000–500,000	**£0.41**
Band 8	500,000–1,000,000	**£0.36**
Band 9	>1,000,000	**£0.31**

Note: The figures in this table are given for 2019. Average unit cost is derived as the sum of industry direct costs and general and support OPEX as a proportion of total industry load.

Source : Anglian analysis provided during the CMA appeal.

Population density as a substitute to the load treated in bands 1-3 (but as a complement to a model with WATS)

Table A.5 SWT1 model with population density as the economies of scale variable

Model	SWT1	SWT1A
Load (log)	0.653*** (0.000)	0.643*** (0.000)
Load treated with ammonia consent ≤3mg/l	0.006*** (0.000)	0.007*** (0.000)
Load treated in size bands 1 to 3 (%)	0.029 (0.211)	
WAD LAD from MSOA (log)		-0.193*** (0.002)
Constant	-3.734*** (0.004)	-2.126*** (0.000)
Model Robustness Tests and Additional Diagnostics		
Adjusted R2	**0.854**	**0.877**
RESET	0.056	0.328
VIF	5.337	7.476
Pooling	0.999	0.999
Normality	0.024	0.004
Heteroskedasticity	0.417	0.273
LM	0	0
Efficiency Score Distribution		
Minimum	0.82	0.91
Maximum	1.50	1.48
Range	0.68	0.57
Upper Quartile	0.94	0.95
Sensitivity Analysis		
Removal most efficient company	A	G
Removal least efficient company	A	G
Removal first year	G	G
Removal last year	G	G

As mentioned elsewhere we believe that the model with WATS is clearly superior to any model that includes a population threshold variable or a density variable. However, *if*, in addition to a required model with WATS, Ofwat were to consider another model with *load treated in size bands 1 to 3 as a* driver, then we think that a model with the WAD LAD from MSOA variable would represent an improvement over such a model, although still insufficient to fully capture economies of scale.

The population density variable is highly significant and holds the expected sign, where companies with sparser regions clearly have to use smaller STWs, and therefore cannot benefit from, or exploit economies of scale. In terms of model performance and robustness, there is a clear improvement in adjusted R2 of two percentage points compared to SWT1 with no issues regarding other diagnostics across multicollinearity, heteroskedasticity etc. The coefficients in SWT1A do not change in significance or sign when the most/least efficient company is removed, unlike in the original SWT1. Furthermore, the original SWT1 'fails' the RESET test, indicative of the model suffering from misspecification, potentially from an incorrect functional form. The substitution of this problematic variable with population density addresses this issue as it 'passes' the RESET test, indicative of no misspecification problems.

However, as mentioned in our consultation response, the aim of such a model is only to replace SWT1 model if Ofwat thinks that a single model with WATS itself is not sufficient. **In our view, WATS remains the best option to capture economies of scale so a model with it is absolutely required.**

APH and BPL comparison

In the graphs collected in this section, we compare the relationship of APH and BPL (booster pumping stations per length of mains) to companies' energy consumption over a 11-year period from 2011/12 to 2021/22. The data are taken from the Excel file "PR24 Cost Assessment Master Dataset, Wholesale Water Base Costs v4.xlsx", published alongside the base cost consultation.

We plotted the 'topography' measure (APH and BPL) against energy usage measures (both in terms of MWh and £m), normalised by distribution input (BN1000_CA22_A, i.e. "Water balance – Distribution input"). We first focus on power costs related to TWD and then extend it to total power costs and total energy consumption.

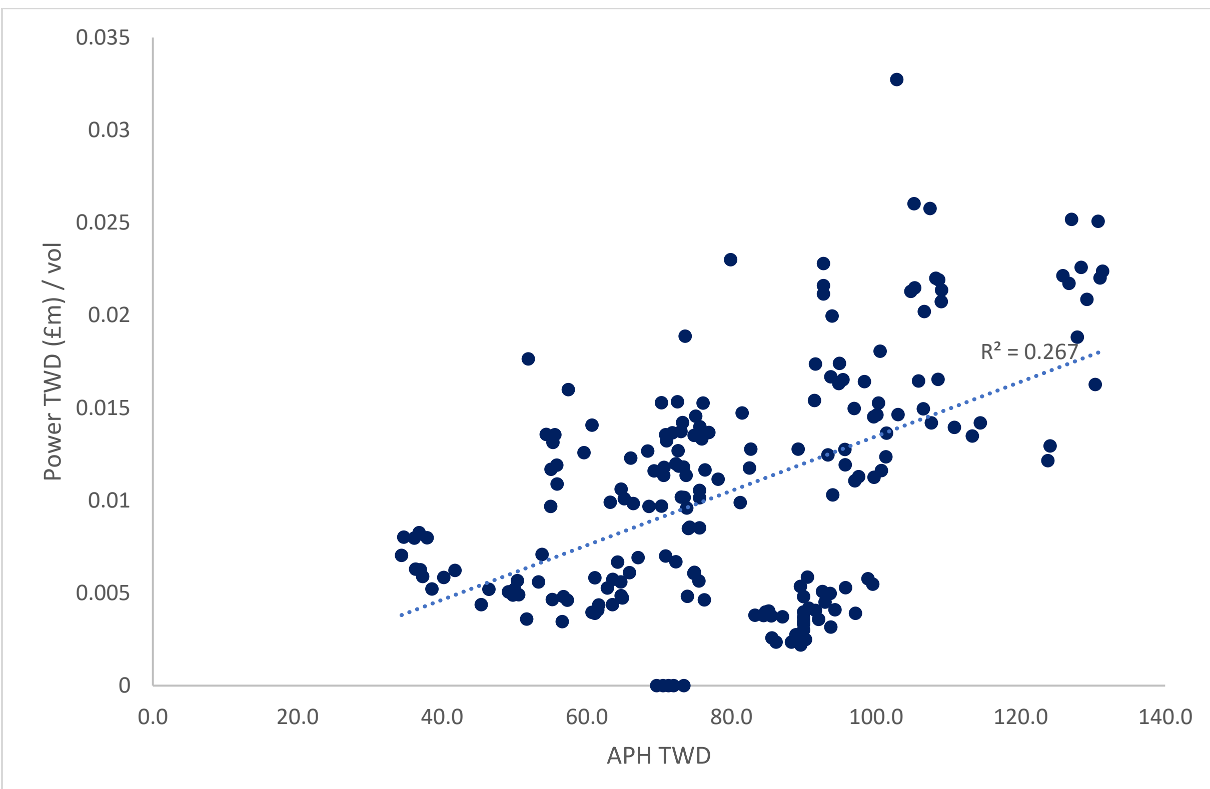
In all cases, APH presents a consistently stronger positive relationship with energy expenditure or usage than BPL.

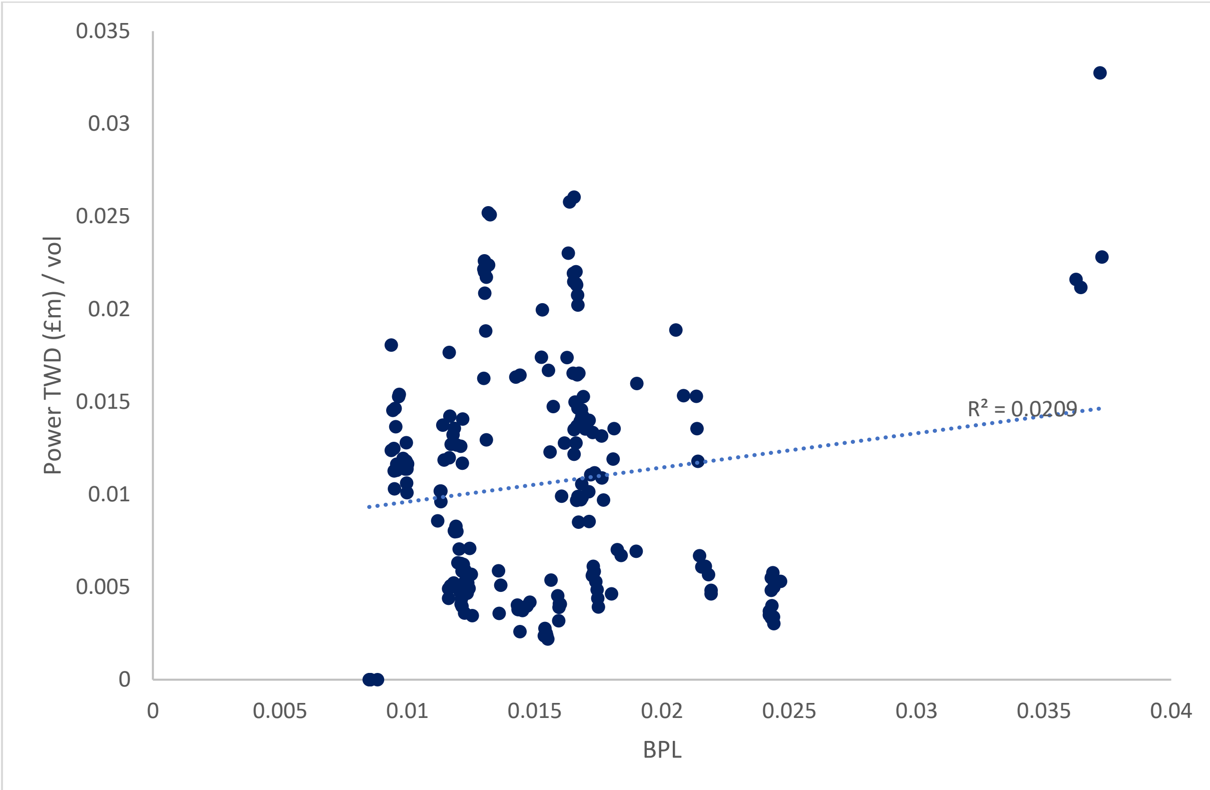
This supporting evidence further demonstrates the clear superiority of APH as a measure of network topography and energy needs, thus reinforcing the case for its exclusive use in both TWD and WW models.

Comparison 1 (APH TWD and BPL vs TWD power costs)

Both APH TWD and BPL show positive coefficients, but APH TWD's R^2 is over ten times higher than BPL.

Figure A.1 APH TWD and BPL vs TWD power costs

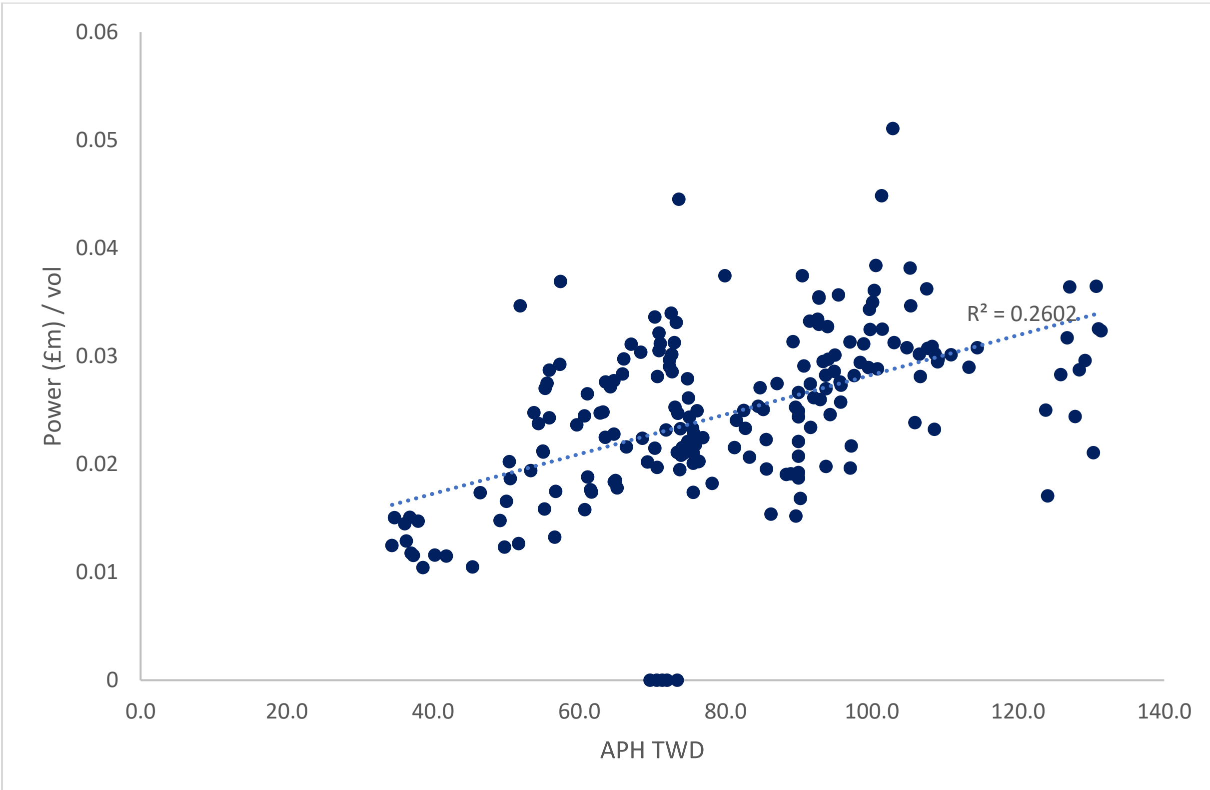


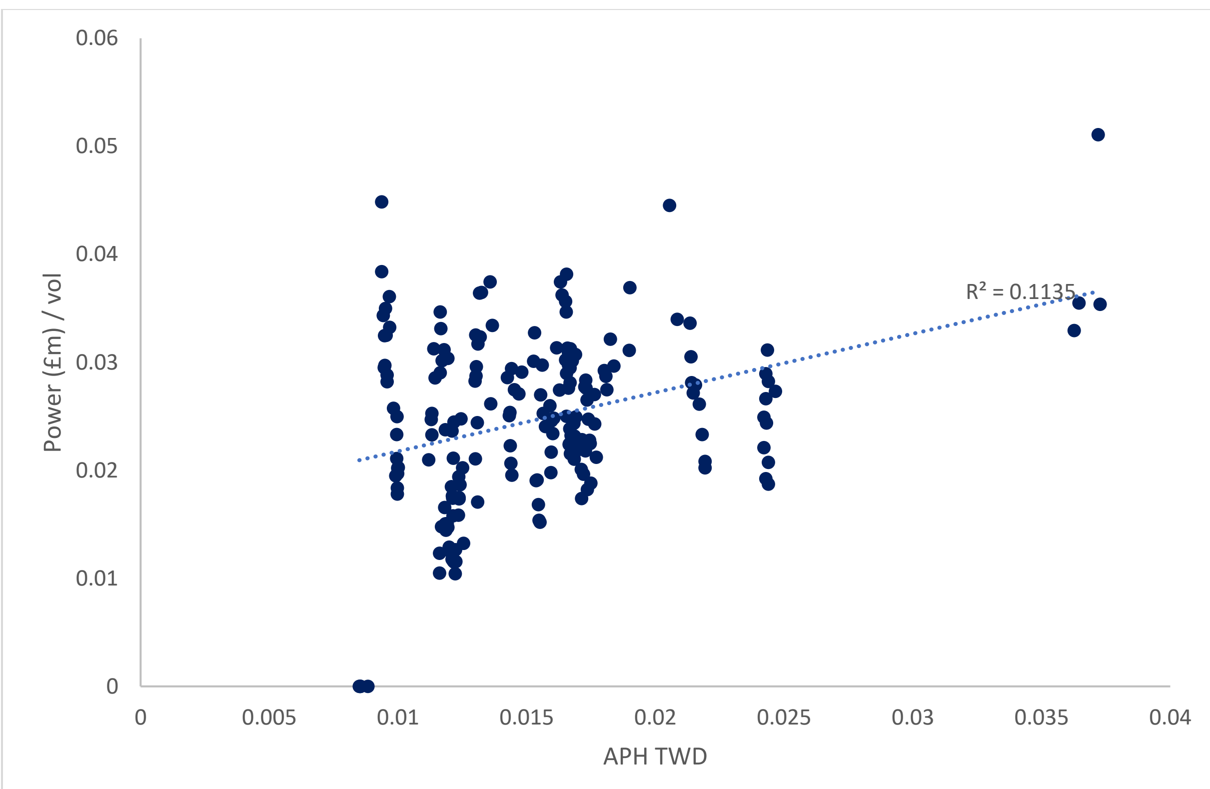


Comparison 2 (APH TWD and BPL vs total power costs)

We also examine the explanatory power of these two measure for energy consumption in MWh, which removes the impact of changes energy prices over time and thus provides a 'cleaner' comparison. APH TWD's explanatory power of total energy consumption improves over energy costs (with an R^2 of 0.41 compared to 0.26). In contrast, BPL's R^2 worsens (to 0.08) and is five times lower than APH TWD's. When we use total APH instead, its explanatory power increases to 0.51, which further increases the gap with BPL.

Figure A.2 APH TWD and BPL vs total power costs

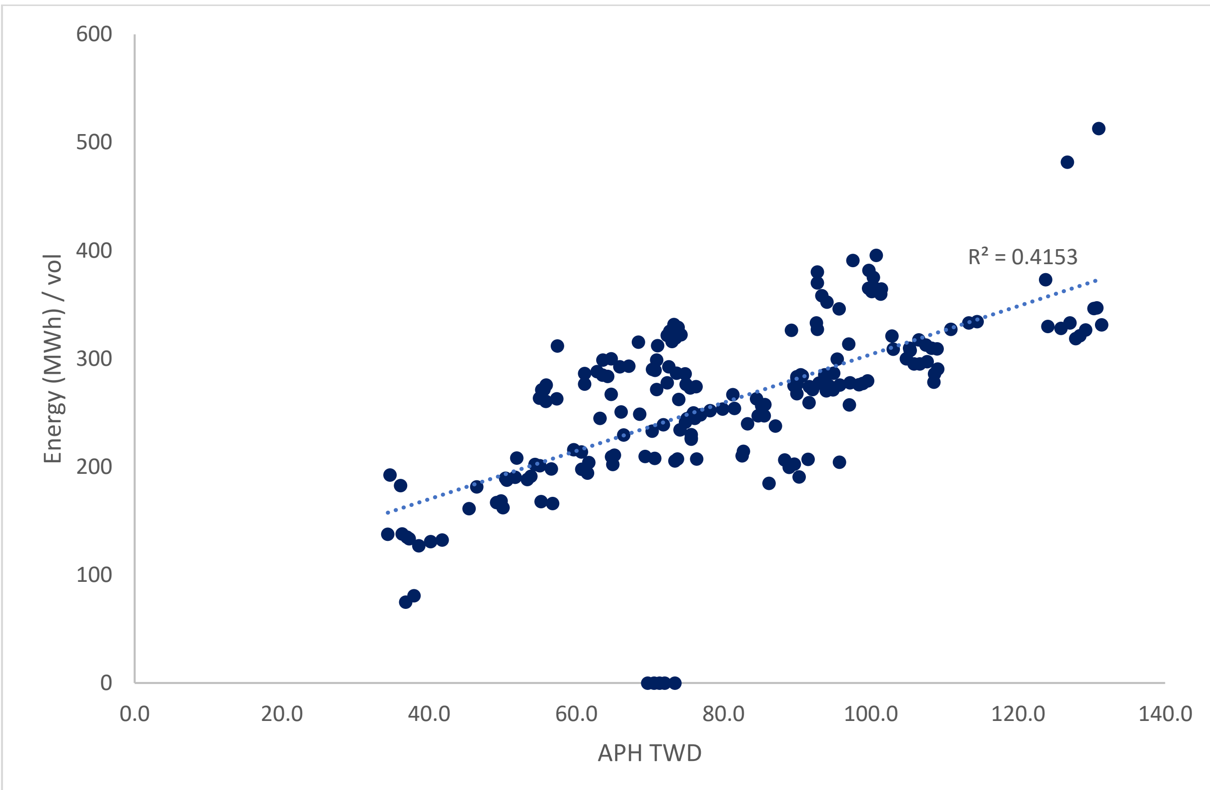


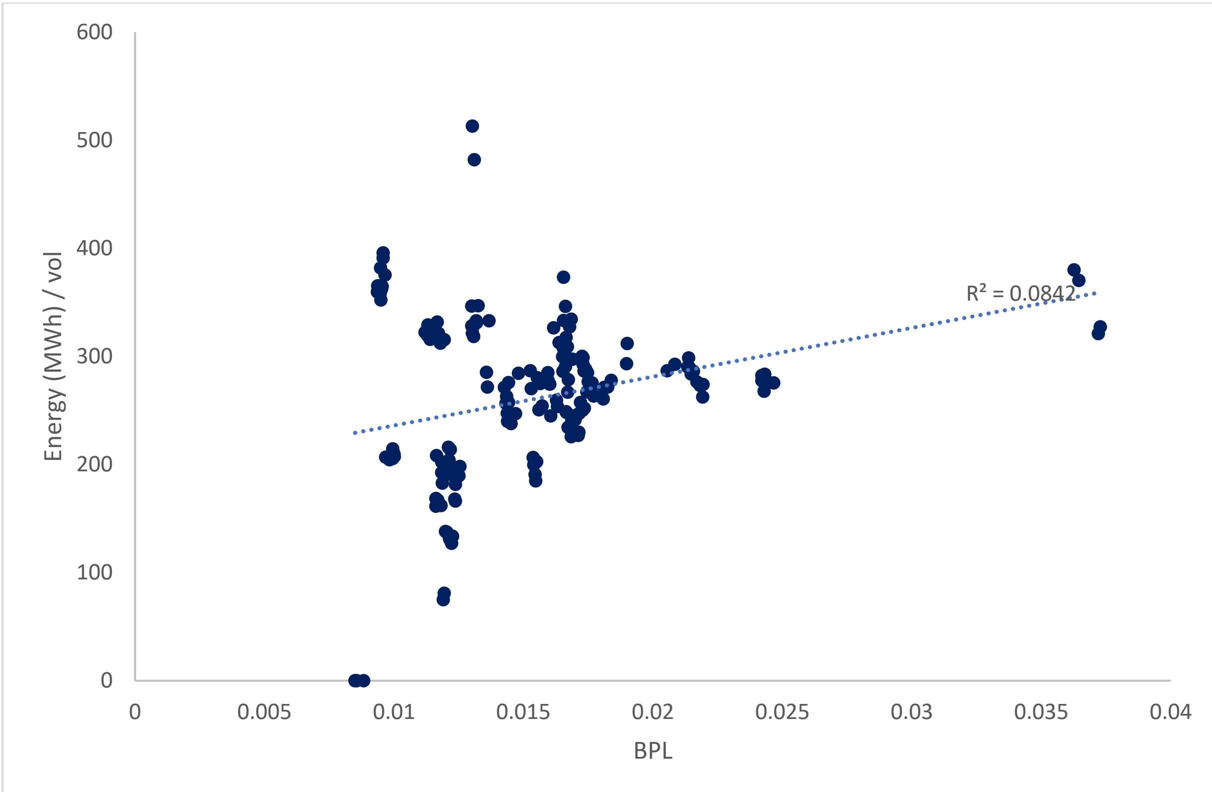


Comparison 3 (APH TWD and BPL vs total energy consumption)

APH TWD performance improves significantly, presumably as energy consumption in MWh removes the impact of prices. BPL's R2 worsens, and is five times lower than APH TWD's. We note that when we use total APH instead, its explanatory power increases up to 0.51, which further increases the gap with BPL.

Figure A.3 APH TWD and BPL vs total energy consumption in MWh




600
500
400
300
200
100
0
Energy (MWh) / vol
R^2 = 0.0842
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
BPL

U-shape relationship between sewage collection costs and density

When we include the square term of our favoured density measure, the WAD LAD from MSOA, the estimated coefficients of both the density measure and its square term are, in addition, to being consistent with the economic intuition, significant at the 5% level.

Table A.6 – Alternative SWC2 model with the square term of the density variable

Model	SWC2A
Sewer length (log)	0.847*** {0.000}
Pumping capacity per length of mains (log)	0.594*** {0.000}
WAD LAD from MSOA (log) WAD LAD from MSOA (log) squared	-2.291** {0.041} 0.169** {0.021}
Constant	3.016 {0.501}
Model Robustness Tests and Additional Diagnostics	
Adjusted R2 RESET	0.897 0.326
Sensitivity Analysis	
Removal most efficient company	A
Removal least efficient company	G
Removal first year	G
Removal last year	G

Such model performance can be easily compared with Ofwat's WRP2 model where the estimated coefficient of both the density measure (the same one as above, i.e. WAD LAD from MSOA) and its square term are significant at the 5% level. However, while the removal of the most efficient company (Wessex) implies a slightly lower significance for the density coefficient at the 10% level, the removal of the least efficient company does not lower the significance of any of the variable (unlike the case with Ofwat's WRP2 model). Given the economic rationale behind the U-shape relationship between sewage collection costs and density and the strong statistical performance of models with a square term of the density, it is not clear on what basis such models have been dismissed.

Table A.7 – Ofwat's WRP2 model as a comparison point to SWC2A above

Model	WRP2
Connected properties (log)	1.075***
	{0.000}
Weighted average treatment complexity (log)	0.343
	{0.183}
WAD LAD from MSOA (log)	-1.468**
	{0.026}
WAD LAD from MSOA (log) squared	0.091**
	{0.031}
Constant	-5.660***
	{0.002}
Model Robustness Tests and Additional Diagnostics	
Adjusted R2	0.902
RESET	0.367
Sensitivity Analysis	
Removal most efficient company	A
Removal least efficient company	A
Removal first year	G
Removal last year	G

Metering

The addition of the 'proportion of metered households' to the RTC1 model improves it in certain respects (most notably improving the significance of the coefficient estimates on other accepted cost drivers), whilst performance remains similar on the other aspects where Ofwat's RTC1 model is already performing well (most pertinently, model fit and stability under sensitivity checks). The estimated coefficient of metering is of the expected sign, close to the PR19 estimate (though marginally higher, 0.005 vs 0.004) and has a satisfactory significance level when put in perspective with other variables included by Ofwat that are even less significant (such as the load treated in size bands 1 to 3 or the weighted average treatment complexity measure). Finally, we would add that the resulting efficiency scores from the RTC1 model plus metering results in a slightly narrower range of efficiency scores than Ofwat's current RTC1 model (0.49 versus 0.46 in the latter).

Table 08 - RTC1 model with and without metering

Model	RTC1	RTC1 + metering
Average bill size (£ per/household) (log)	0.651***	0.642***
	{0.000}	{0.000}
Equifax - Percentage of households with payment default (%)	0.025**	0.045***
	{0.012}	{0.003}
Total number of households (log)	-0.096***	-0.118***
	{0.002}	{0.000}
Covid-19 dummy for 2019-20 (nr)	0.176***	0.182***
	{0.000}	{0.000}
Covid-19 dummy for 2020-21 (nr)	0.058**	0.062**
	{0.022}	{0.015}
Proportion of metered households (%)		0.005
		{0.167}
Constant	0.405	0.036
	{0.255}	{0.933}
Adjusted R2	0.697	0.695
RESET	0.103	0.286
VIF	2.708	3.361
Pooling	1	1
Normality	0.036	0.146
Heteroskedasticity	0.041	0.011
LM	0	0
Minimum	0.83	0.81
Maximum	1.32	1.27
Range	0.49	0.46
Upper Quartile	0.95	0.92
Removal most efficient company	G	G
Removal least efficient company	G	G
Removal first year	G	G
Removal last year	A	A

Sewage treatment models relying on the coastal population as a cost driver

If we remove Southern Water from the analysis (i.e. the least efficient company), the modelling results clearly show that these models cannot be relied upon as they are not robust. The first model gives a negative counterintuitive sign which would mean that the higher the coastal population the lower the costs which is counter to Southern Water's view. The significance of the estimated coefficient of the coastal population for the other two models is very low which indicates that the estimated relationship between sewage treatment costs and the coastal population is spurious. It seems likely that models including Southern are simply capturing part of the inefficiency of the company and the different sensitivities tests show the high inaccuracy and the inability of the coastal population to explain companies' base costs on sewage treatment.

Table 09 - Sewage treatment models including coastal population (excluding the least efficient company)

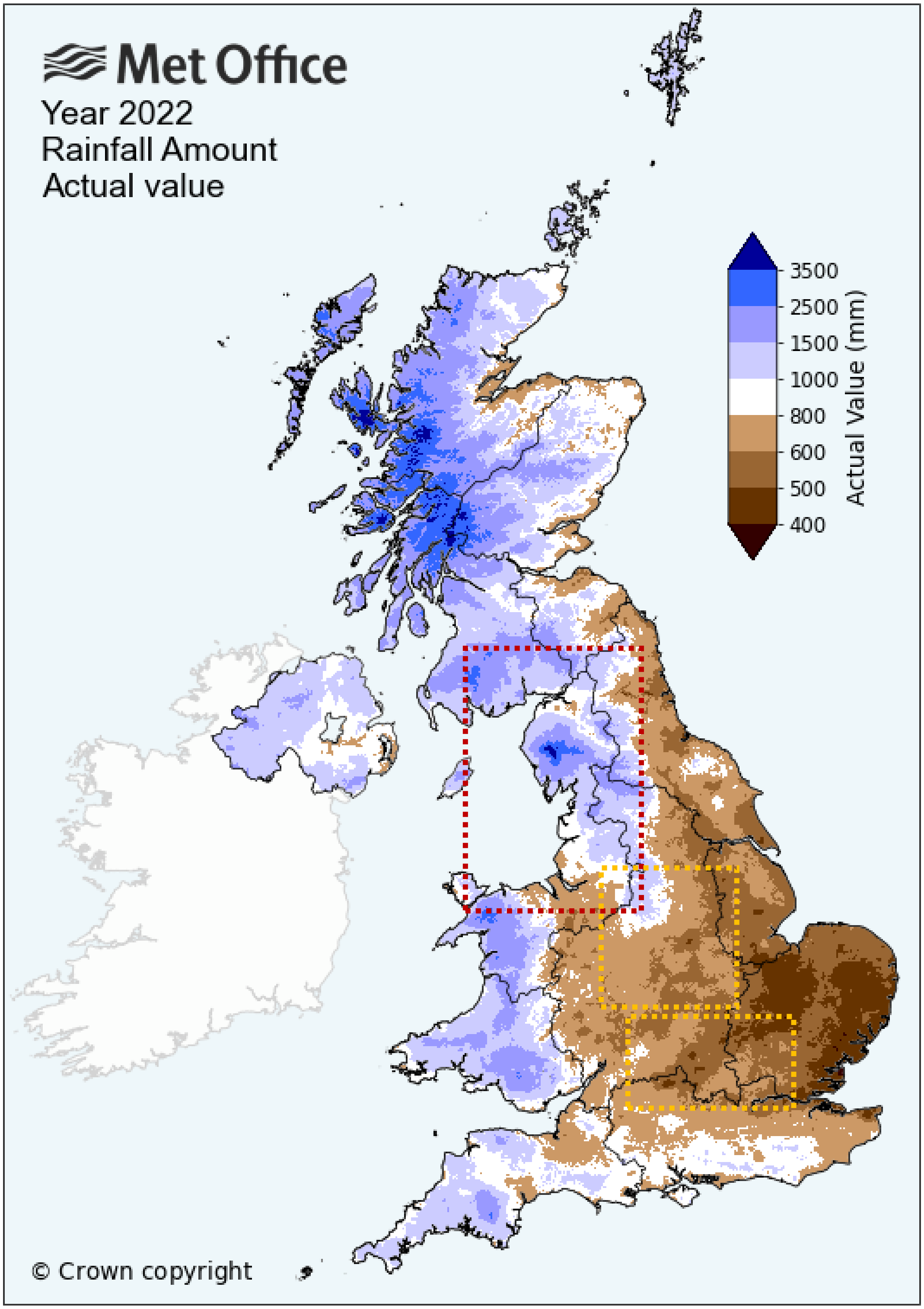
Model	SWT1	SWT2	SWT3
Load (log)	0.582** {0.033}	0.820*** {0.001}	0.927*** {0.000}
Load treated in size bands 1 to 3 (%)	0.047* {0.077}		
Load treated with ammonia consent ≤3mg/l	0.006*** {0.000}	0.006*** {0.000}	0.006*** {0.000}
Population living in coastal areas (%)	-0.007 {0.645}	0.004 {0.738}	0.008 {0.379}
Load treated in STWs ≥ 100,000 people (%)		-0.009*** {0.000}	
WATS (log)			-0.227*** {0.000}
Constant	-2.826 {0.436}	-5.335 {0.123}	-5.040** {0.041}

Urban rainfall

Urban rainfall is currently defined as the “average rainfall falling in a company area (mm) multiplied by the urban company area (squared kms)”. However, as rainfall is not homogeneous over a company’s area, this definition may lead to an inaccurate representation of the rainfall received by a company’s urban areas. An example is found in the case of UUW (highlighted in red), the company with the highest average urban rainfall over the 2012-22 period. In UUW’s case, areas with the most rainfall are located in the Lake District area, far away from the main urban centres, which receive rainfall at similar levels to large parts of England. As a consequence, UUW will have higher urban rainfall values than are actually received by its network in its urban areas.

Moreover, SVE and TMS, the second and third companies with the highest urban rainfall (highlighted in yellow), appear to be located in low-precipitation areas. In this case, urban rainfall seems to only be capturing the level of urbanity, thus failing in controlling for the impact of excessive rainfall on sewerage networks.

Figure A.4 UK annual rainfall (mm, 2022)



Source: Met Office, available here: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-actual-and-anomaly-maps.

Moreover, by looking at the companies with the highest urban rainfall values (coloured in light blue, i.e. NWT, SVH and TMS—see Figures A.5 and A.6), it is clear that they are not disproportionately affected both in terms of OPEX (power expenditures) or number of blockages, as shown by the subsequent scatter plots presented below (for 2022). If these companies had higher costs or were negatively impacted in some way by high 'urban rainfall', we would expect them to appear as outliers significantly *above* the regression line.

Indeed, when regressing power expenditure against sewerage length over the period 2012-2022 (the related scatter for 2022 is provided in Figure A.7), a dummy variable indicating the three companies with the highest urban rainfall is found to have a *negative* coefficient (although non-significant). In the case of the number of blockages (the related scatter for 2022 is provided in Figure A.8), the coefficient is both *negative* and significant. Similar results are obtained when using a dummy indicating the companies with the highest level of rainfall. The coefficients are presented in table A.10.

These results are contrary to the assumption that a high level of urban rainfall (or rainfall) leads to higher expenditures and incidents, thus confirming the issue raised concerning data quality and operational dynamics. In fact, the significant *negative* effect of 'urban rainfall' on blockages is consistent with the operational insight that rainwater flow *assists* in the transportation of wastewater and the clearance of solids that might otherwise block pipes.

Figure A.5 Annual rainfall by company (mm, 2012-2022 average)

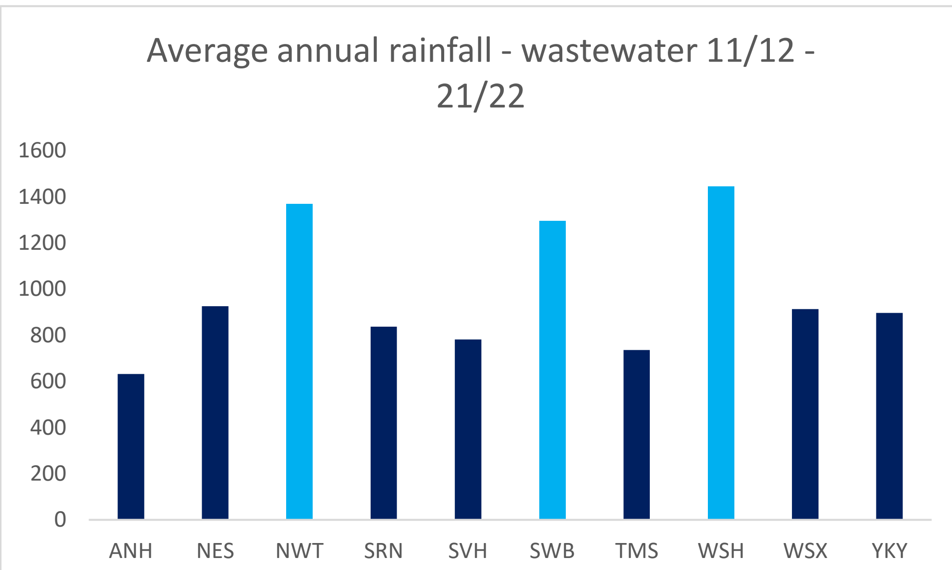


Figure A.6 Urban rainfall (2012-2022 average)

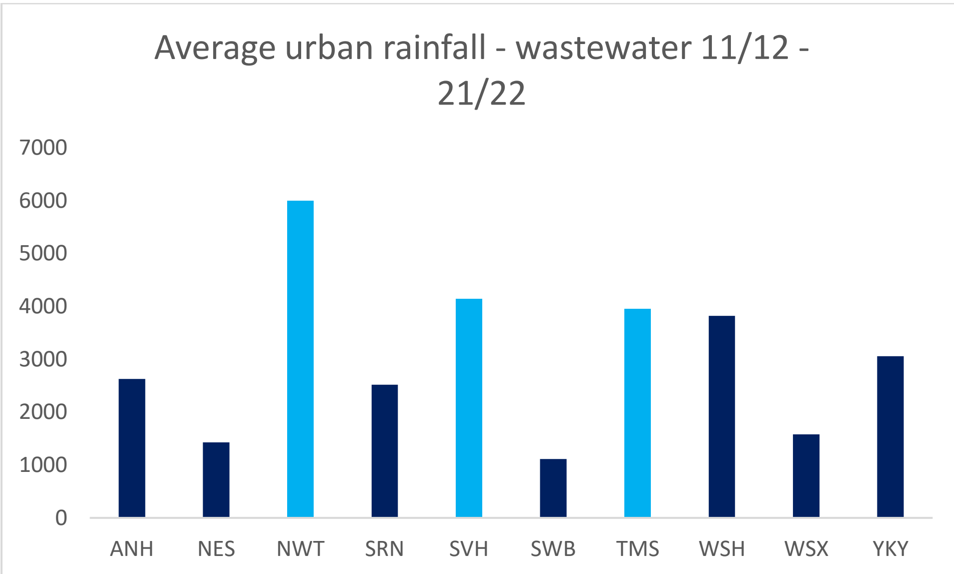


Figure A.7 Power expenditure vs sewer length (2022)

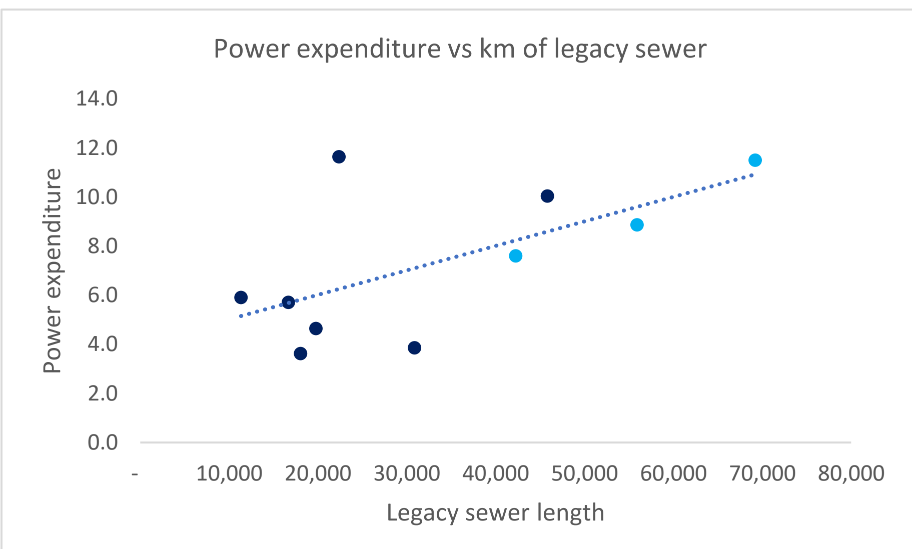


Figure A.8 Blockages vs sewer length (2022)

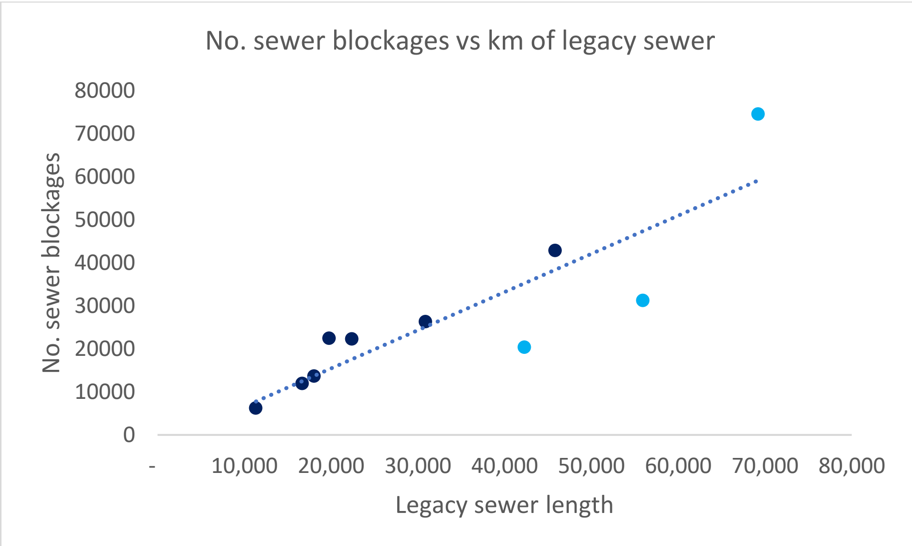


Table 010 – Impact of high urban rainfall on power costs and blockage

Model	Blockages	Blockages	Power	Power
Length	1.279*** {0.000}	0.955*** {0.000}	0.0001*** {0.000}	0.0001 {0.000}
Dummy (high urban rainfall)	-14218*** {0.000}		-1.22 {0.227}	
Dummy (high rainfall)		-4565** {0.021}		-0.629 {0.316}
Constant	-8689*** {0.000}	-911 {0.658}	3.58*** {0.000}	4.39 {0.000}