ST Classification: UNMARKED

PR24 Cost modelling consultation

Severn Trent submission

12 May 2013

WONDERFUL ON TAP

SEVERN TRENT

Contents

Executive Summary

We appreciate the open and transparent process that Ofwat has adopted to develop the models for PR24. Undertaking the consultation early and providing opportunity to propose improvements, is an important means to developing a set of models that are better specified / reflect engineering logic whilst also being suitably simplistic.

There are several features of Ofwat's proposed set of base cost models that we welcome, including the use of average pumping head and the triangulation over several alternative density measures. However, having reviewed the PR19 models and consultation models alongside colleagues from our Chief Engineer department, Atkins Engineering, Frontier Economics and Professor Ron Smith from Birkbeck University there is a compelling case for further strengthening the engineering logic of key models. We offer a more detailed review of Ofwat's PR24 consultation models in chapter 1.

By incorporating genuine cost drivers based on engineering logic, the models can be better specified and most importantly better describe the inherent external cost drivers that lead to a company incurring more or less cost than it would be allocated under a simple unit cost approach. This ensures that companies are funded for the costs they incur. Absent these changes, it creates the risk of artificial winners and losers (i.e., not reflecting actual efficiency but the extent to which the external drivers that impact on the company has been specified in the models).

In our response we have focused on four changes that have a material improvement in the specification of the models:

- Topography - Reflecting the two distinct ways that topography drives costs.
- Missing data – Accounting for the c40% of pumps that are before treatment works.
- Density – Removing the squared term to better reflect the cost drivers and addressing the impact that TMS as an outlier has on the data set.
- P removal - Ensuring phosphate removal is appropriately reflected to support the delivery of ecologically good rivers.

In this paper we summarise the key improvements below and highlight some important considerations regarding triangulation, before discussing in more detail in the following chapters these targeted opportunities to improve the modelling for PR24. In several cases we also plan to submit a similar cost adjustment claim in June in the event that a modelled remedy is not considered acceptable.

1. Topography cost drivers needs to account for both height and the spread of hills

The topography of a region influences water costs in two distinct ways:

- **Height -** The greater the difference in height between the abstraction point and consumption point the more water needs to be pumped, thereby driving higher energy costs. This cost is captured by pumping head.
- **Spread -** In hilly regions more boosters are needed per 1km of mains to move the water around to customers at a constant pressure. This drives higher capex costs associated with the boosters and higher costs associated with pressure management devices (e.g. PRVs). This driver is captured by boosters per 1km length of mains.

We welcome Ofwat's work to improve pumping head data and its inclusion in some of the draft models. **However, there is a compelling case to include this alongside boosters per length because they describe different costs.**

We note that Ofwat considers "both variables proxy for network topography". However, this is too simplistic a view of topography. In practice topography is a very broad term with these two complementary metrics describing different aspects.

For example, a company cannot typically supply an undulating supply area by simply installing one powerful pump at the WTW– excessive pressures would cause bursts on the network and within customers pipes. Instead, the only way to manage hilliness is with lots of boosters (and corresponding pressure reduction values) to deliver water within a constant pressure range. This more asset heavy configuration may or may not have a similar total pumping head.

Conversely, there are examples were populations are supplied from a water resource is lower than the supply area. For example, Coventry where water is largely sourced from the River Severn at Strensham WTW and then pumped 185m vertically to Meriden for distribution into the city. Such configurations require a small number of very large pumps. The material costs of this pumping (more than £4.4m per year) are poorly accounted for by the boosters/length variable – The metric does not describe pumping head; and the asset assemblage adds only one booster. Therefore, these costs are likely to be considered as inefficiency.

The fact that pumping head and boosters/length are fundamentally different cost drivers is also shown in statistical test results. For example, if they were proxies for the same driver, then they would be correlated and produce similar efficiency scores. However, unlike water density variables, these metrics show very weak Pearson correlations (0.2 versus 0.9), indicating that the presence of one does not give rise to the presence of the other. Importantly, they also produce different efficiency scores, which is why they need to be reflected in the modelling to capture genuine cost drivers otherwise it creates winners and losers for reasons other than efficiency.

In terms of the statistical performance of the models, they are also unquestionably stronger from the inclusion of both variables.

- All of the Ofwat models see an improvement in R-squared through the inclusion.
- The RESET test is passed in all cases.
- The AIC and BIC offer substantial improvements over Ofwat's PR24 suite, suggesting there is little likelihood of overfitting.
- CEPA (Ofwat's cost assessment consultants) trialled this in its report and found that models that included both these variables worked better than models that included just one individually.

Overall, there appears to be a very compelling case for including both variables in the model set.

Water network plus costs drivers are described further in chapter 2.

2. The explanatory variable for (height) topography excludes c50% of pumping head

The amount of pumping a company needs to undertake is a direct function of the height difference between the source of the water and service reservoir that serves the customer's property. Under the draft suite of models, 42% of pumping head is disregarded because the pumps are located before the treatment works rather than after.

The engineering logic for the inclusion of WRP APH is clear – it is the same as TWD APH which has been included. In the current models, companies that have to pump a lot of water between the source and the treatment works are being clearly disadvantaged. Despite 57.7% of sector wide APH being in the treated water distribution network, for some companies more than 50% of their pumping is incurred in the water resources or treatment business units.

We can infer from the CEPA report that the basis for excluding the data is because it "was not a significant driver of costs". Clearly this doesn't hold in practice, pumping is a material driver of costs and even from a statistical perspective it is significant for one of the draft models and only weakly insignificant for the remainder. Importantly, just because it is not statistically significant from a modelling perspective shouldn't necessarily be a reason for disregarding – as Ofwat notes with respect to weighted average complexity, what is important is that the coefficient is sensible and the engineering logic is sound.

We also understand that there are concerns about the quality of the data for water treatment APH. However, Ofwat's analysis shows that the data quality of APH before the network is just likely of better quality than the APH data from after the works (which has been included in some model specifications).

For this reason, we don't think it is sensible to exclude c50% of pumping head when power makes up 14.4% of water resources and treatment (WRP) costs in the historical data.

This is described further in chapters 3 and 4 (as WRP APH can be a robust proxy of water treatment complexity and water resources cost drivers).

3. Density cost drivers need to be better accounted for

Water Resources and Treatment (WRP)

There is an opportunity to improve the engineering logic of the water resource plus (WRP) models by better reflecting the benefits of serving highly dense regions of population.

Under the draft approach a squared term is used to account for density in water resources plus models. This means that as density increases cost initially falls (with as companies benefiting from economies of scale) and then after a certain point, costs increase (ie, there is a U shape to the relationship between WRP costs and density). We strongly agree with the premise that there are economies of scale that benefit water resources and treatment costs. However, as the supplier of some of the largest cities in the UK we can confidently state that the disbenefit on water resources is not realised.

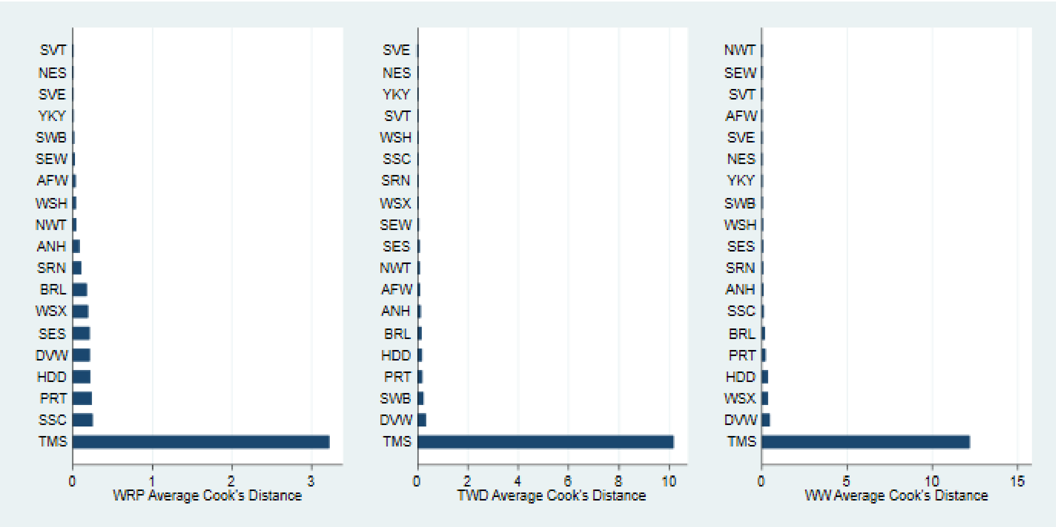
The only rationale for giving companies more totex for water resources at high levels of density is because "property, rental and access costs are higher". However, as a company that operates across Birmingham, Leicester, Coventry, Nottingham and Derby, we have seen no such evidence. Water resources and treatment assets are not located in the centre of urban areas but rather in rural surrounding areas or on the outskirts of towns in which property and access costs are not materially different. Equally, such cost pressures are likely to be immaterial relative to the underlying costs direct cost of abstraction, raw water transport and treatment (i.e. pumping, chemicals and capital maintenance).

The only potential difference is London and the cost of wages. However, given wages only account for 20% of totex, it is unlikely to be a material driver. Ultimately if this is proven otherwise, then a company specific adjustment would be more suited.

This gets to the heart of the issue – that Thames can be shown to be a statistical outlier - and the current approach effectively fits a model to Thames against engineering logic. This can be illustrated in two ways:

- **Cook's distance analysis**:
  - This helps to show the influence of specific observations, or groups in panel data, on coefficients. As a rule of thumb, Cook's distances that are substantially larger than any others in the sample should be treated with caution and manual checks undertaken to test the influence of removing that particular group from the sample. Some authors suggest that values greater than 0.5 are potentially concerning, while values greater than 1 are considered extreme[1].
  - Thames appears to *significantly* change the estimated relationship between density and Botex. Thames' Cook's distances are extremely large, with the average values across WRP, TWD and WW models being 3.22, 10.15 and 12.17 respectively as per ***Figure 1*** below.

Figure 1: Average Cook's distances in Ofwat's proposed PR24 models.

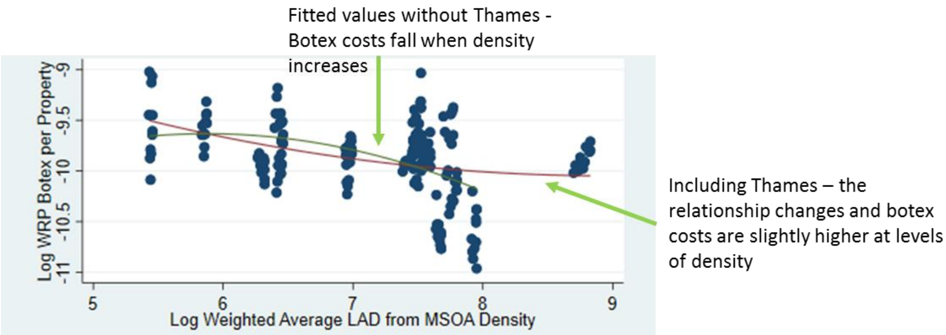


  - The implication is that Thames clearly have a substantial influence on the coefficients estimated for the density drivers.

- **Absence of U-shaped relationship:**
  - If the squared term reflects engineering logic then we should see some evidence of that U-shaped relationship when we produce a scatter of Botex unit cost (Botex per property) and density. This relationship should hold as observations are removed or added.
  - This is not the case, as shown in ***Figure 2***. When TMS is included in models, a curve withan inflection is seen (costs falling, then rising). However, when we exclude Thames from the data set, all of the proposed WRP models suggest a relationship in which density drives a reduction in costs at all levels of density.

[1] Fahrmeir et al. (2022) *Regression: Models, Methods and Applications*. 2nd edn. Berlin: Springer, p.166.

Figure 2: Scatter graph showing the relationship between WRP Botex per Property and Weighted Average LAD from MSOA Density with and without Thames included.



The Thames observations are legitimate and ideally should not be excluded. However, it would be wrong to include the squared specification as it suggests higher densities are detrimental when they are clearly not.

We understand CEPA was reluctant to exclude the squared term based on statistical results. However, an increase in R-squared when that increase comes from fitting to an outlier that exhibits a relationship with a variable that is not consistent with the rest of the industry is not a valid basis for the inclusion of a particular variable. This is, in essence, overfitting.

By removing the squared density variable we can show that there is effectively no change in R-squared, the RESET tests are similar and the Cook's distances are moderated.

Taken alongside the engineering logic, the squared term is mis-specified and should be removed. This would improve the overall specification of the water resources plus (WRP) models and ensure companies are funded for efficient levels of totex, rather than being rewarded and penalised for misspecifications.

Treated water distribution (TWD)

On face value a non-linear density term has the right engineering justification for TWD. This is because costs are likely to rise in both very urban (driven by issues of congestion and complexity in urban settings) and very rural areas (driven by additional assets and distance between assets in rural settings). This would point to a curved relationship between cost and density where the inflection is within the distribution at the theoretical optimal level of density.

However, the effects of rural companies may not be being adequately reflected in the MSOA and LAD from MSOA density drivers. Instead, density variables that describe overall measures of density (population/length and population/area) are likely to better describe rurality cost drivers.

The other key issue is that TMS distorts the results given it is such an extreme outlier.

For treated water distribution the Cook distances for Thames are up to 36.45, which is 89 times higher than the next most "influential company". This value is very extreme and it is clear that Thames unduly influences the assumed U-shape of the relationship between density and Botex. When a linear functional form is assumed, this Cook's distance falls to just 1 which is still high but far more acceptable than with the inclusion of the squared term.

There are two potential options:

- **Remove TMS (or introduce a TMS dummy) and retain the squared term.**

- **Retain TMS and moderate the effect of the outlier by removing the squared term in some models**. Namely:
  - **Use linear density only for models that have a MOSA and LAD from MSOA density specification (TWD models 1, 2, 4 and 5), or remove these models entirely.** The preferential selection of non-linear density model specifications is fundamentally a function of the effect of TMS rather than describing the distribution across the sector observations. Also, despite rurality being a legitimate TWD cost driver, there is not a strong case to retain as the rurality cost drivers are poorly accounted for in weighted average density variables, and rural companies are also likely receiving benefit from the length scale driver – they appear larger than would otherwise appear.
  - **Retain density squared for pop/length (and include population per area as an additional point of triangulation) i.e. as per Ofwat model (TWD3 and 6).** These models better reflect rurality drivers as shown by the material difference in the distribution of model that use linear and non-linear density once TMS is removed. The inclusion/exclusion of TMS is much less sensitive to these models (i.e. TMS is less of an outlier if using this variable).

Density cost drivers are described further in chapter 5.

4. Operating sewage works with tight P consents need to be better reflected

The increased focus on rivers and addressing the reasons for why rivers do not achieve good ecological status means P-consents are set to tighten substantially over the course of the current AMP and into the next AMP. We welcome Ofwat's recognition of this issue in the draft consultation and its desire to account for increased costs as a result of tightening P-consents. The sector will only be able to deliver sustained ecological improvements if the base models appropriately account for the higher costs associated with phosphorus treatment, which accelerates when consents move below 1mg/l.

When P-consents are set at very low levels, additional processes must be installed, as has been the case at Finham STW where we have had to install a bespoke tertiary solids removal plant in order to satisfy a 0.22mg/l P consent. This comes with additional capex and opex requirements. At Finham alone we are incurring additional opex of more than £0.8m per year as a result of the tighter consent. These processes require additional chemical dosing (e.g. ferric) and additional power, primarily for mixing, to achieve the tighter consents. Even at slightly slacker consents (of less than ~0.75mg/l) effluent standards cannot be achieved biologically and require significant chemical dosing and mixing adding cost pressures that were not previously there.

Our view is that the models can and should accommodate a variable to account for the percentage of load treated with phosphorus consents of less than 0.5mg/l. This could be achieved through inclusion as part of a composite variable.

We propose that either P≤0.5mg/l is included in the sewage treatment models as an additional variable, triangulated with the current NH3≤3mg/l or included as part of a composite variable. The composite variable assumes the Botex response to a unit (1 percentage point) increase in either variable is the same. We can write the composite variable with its coefficient as:

$$\beta Composite \equiv \beta \frac{\%NH3 \leq 3mg/l + \%P \leq 0.5mg/l}{2} \equiv 0.5\beta\%NH3 \leq 3mg/l + 0.5\beta\%P \leq 0.5mg/l$$

Whilst this is likely to not be perfectly case – our experience is that treating high P consents is likely to be significantly more costly than Ammonia – given the lack of disaggregated cost data with which to calculate appropriate weightings, this seems to be a pragmatic solution if we are to model Phosphorous consents with the data currently available.

The effect of tight P consents on operational cost drivers are described further in chapter 6.

Appropriate use of triangulation

Finally, we strongly support the use of additional models to triangulate an output where identified models are different but equally valid. Where it is clear that a particular cost driver should be accounted for, but there are a range of appropriate explanatory variables, triangulation is highly effective. However, triangulating over several models, some or all of which individually are 'missing' a significant cost driver only serves to average out the noise. This is not the same as describing the true underlying efficiency of companies within our sample.

Density is a good example of where triangulation should be considered – different density measures produce very different company rankings and it is unclear which of the measures is 'right'.

Triangulation is not appropriate to assimilate model outputs that have varying engineering logic included or excluded. Either the engineering logic is appropriate, or it is not. An example of where we believe triangulation should not be used is in the network plus wastewater models. Here six models include economies of scale in sewage treatment but two do not. In our view, all models should have some consideration of this. Therefore, the two models where it is not present should be excluded rather than triangulated.

We consider that where there is a great deal of movement in the modelled efficiency of companies where something is removed or included for triangulation, this can serve as a warning that triangulation is not appropriate – this is the case for boosters per length and average pumping head.

1. Overview of Ofwat's PR24 consultation models

In this chapter we review the models that Ofwat has presented in its consultation.

We fully support Ofwat's Principles for PR24 base cost assessment which highlight the need for models to be "consistent with engineering, operational and economic rationale". We also agree that "Robust econometric cost models" should perform appropriately against relevant statistical tests; however, it is our strong view that this should not be used to override the first principle.

We have reviewed both the coverage of engineering expectations, and statistical performance of the draft PR24 models that Ofwat has presented. The way that we have done this is summarised below.

Engineering expectations

In January we set out in detail our 'engineering expectations' at a subservice level, the summary tables are set out in appendix 1. We have used these to review the coverage of "engineering, operations and economic rationale in Ofwat's models. As summarised in ***Table 1***, our engineering expectations show a need to take adequate account of:

- the scale of the company;
- opportunities for economies of scale (as often proxied by population density);
- the extent to which geographical/geological circumstance drive complexity;
- weather effects;
- cyclical expenditure driven by the price setting process; and
- the effects of deprivation meter reading and population transience on retail costs.

Table 1: Coverage of Severn Trent engineering expectations across subservices

Engineering Expectation type	Water (Resources + Treatment)	Water (Network)	Waste (Network)	Waste (Treatment)	Bioresources	Retail
Scale	EE1	EE1	EE1	EE1	EE1	EE1
Density / Economies of scale	EE2, EE3	EE2	EE2	EE2	EE2, EE3, EE4, EE5	
Geographical / Geological Complexity	EE4, EE5	EE3, EE4	EE3	EE3	EE6, EE7, EE8, EE9, EE10	
Weather effects	EE7	EE5	EE4	EE4		
AMP effects	EE8	EE6	EE5	EE5		
Retail specific (Ability to pay, meter reading, population transience)						EE2, EE3, EE4, EE5, EE6, EE7, EE8, EE9

Statistical performance

To review the statistical performance of models in our January submission, we highlighted the following statistical performance tests:

- **Statistical fit:** R-squared describes the proportion of variance described by the model. We also reviewed the R-squared of unit cost configuration given that scale drivers will dominate the explanatory power of models.

- **Functional form:** RESET test considers the need for higher order terms.
- **Predictive power:** AIC describes that benefit of adding extra parameters while guarding against over-fitting.
- **Efficiency range:** Efficiency ranges need to be plausible. Poorly specified models will generate very large efficiency ranges. Our view is that if modelled efficiency ranges reduce substantially in response to the addition of a variable supported by engineering logic, then legitimate costs have been incurred which should not be attributed to inefficiency. We are concerned about Ofwat's presuppositions on the nature of modelled inefficiency[2]
- **Number of parameters:** To deliver 'sensibly simple' models, we are trying to deliver that maximum number of engineering expectations using the minimum number of parameters

In our January submission, we colour coded statistical test results to show improving (or reducing) statistical performance of proposed models relative to PR19 model specifications.

Review of Ofwat's consultation models to engineering expectation and statistical performance and opportunities for improvements

We have set out the extent to which Ofwat's models improve the coverage of engineering expectations and statistical performance following the same process that we followed in our January submission. This is set out in Appendix 2 and summarised in ***Table 2*** below.

Table 2: Summary of coverage of engineering expectations and statistical performance of Ofwat PR24 consultation models relative to the PR19 models.

Model scope	Engineering expectations	Statistical performance	Opportunity to make material improvement by simple remedy
Water Wholesale (WW)	No change	No change	High (Combine WW APH & Boosters)
Water resources and treatment (WRP)	No change	Reduced	High (Add WT APH and change WAC weightings, remove squared density)
Water Network (TWD)	Limited improvements (properties/length density)	Moderately improved	High (Combine TWD APH & Boosters, change density)
Waste Network Plus (NPWW)	N/A – Not at PR19	N/A – Not at PR19	Moderate (Consistency in application of economies of scale at treatment works)
Waste Collection (SWC)	Limited improvements (rainfall)	Moderately improved	Limited
Waste Treatment (SWT)	Limited improvements (weighted average size)	Improved	High (Some allowance for tight P consents)
Bio	Reduced	Reduced	High (Intersiting, treatment, disposal)
Retail	No change	Moderately improved	Limited

In summary, we acknowledge that there have been improvements to some models relative to PR19, but there are also areas where models have regressed.

[2] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.33.

However, we consider that there are some clear opportunities to make simple, targeted remedies to the Ofwat consultation models that will lead to material improvements in both their coverage of engineering expectations and statistical performance.

The case for these remedies are set out in the chapters 2-6. The performance (engineering expectation and statistical performance) of these updated models is then set out in chapter 7.

2. Network Complexity (and the wider impacts of pumping)

In this chapter we focus on how Network Complexity is accounted for in the proposed PR24 Wholesale Water modelling suite. We do not think that the separation of Boosters per Length and Average Pumping Head (APH) in the Treated Water Distribution and Wholesale Water models is sensible. To rectify this, we propose a simple remedy of allowing Boosters per Length and APH to feature in the same models.

We also consider that there is a case to better account for complexity at water treatment works and the preceding water resources processes. These are discussed further in ***Chapter 3*** and ***Chapter 4*** where we set out that the relevant pumping head metrics can be used to account for the cost drivers that are present but currently not well accounted for.

Separate remedies could be implemented by adding in further explanatory variables for each of these complexity cost drivers (i.e. a TWD APH explanatory factor as set out here, then a Water treatment APH explanatory factor, then a Water Resources APH explanatory factor). Whilst each can be justified separately, a simpler, econometrically more 'efficient' and more pragmatic remedy is to account for these factors together. This would mean that the use of APH explanatory variables would be as follows.

- Combining TWD APH with boosters per length in the TWD models (as discussed in this chapter).
- substituting Treated Water Distribution APH for Wholesale Water APH in the Wholesale Water models (accounting for the findings of chapters 2, 3 and 4).
- Adding Water Resources Plus or Water Treatment APH to the Water Resources Plus models (accounting for the findings of chapter 3 and 4).

2.1 Boosters per Length and TWD APH are fundamentally different variables

We contest that the 'boosters per length' and 'APH' explanatory variables are not sufficiently similar to be considered as proxying for the same unobservable exogenous driver of cost. More than this, we suggest that one driver makes little sense without the presence of the other. We will be submitting a cost adjustment claim to this effect as an alternative way to mitigate this material issue.

Boosters per Length

Boosters per Length is defined as the number of booster pumping stations per km of mains. This is the 'true' topography proxy – lots of small boosters are necessary to move water through hilly terrain to serve customers that live in these largely rural areas. Conceptionally, it is unlikely that high volumes of water will be moved through these areas because it would be particularly inefficient to do so. Where there are large population centres within such regions, water resources have traditionally been sought and developed such that gravity fed distribution networks that limit the impact on APH have developed. The need for boosters will to some extent depend on where sources and treatment works are located.

Relative network length increases with rural areas. Therefore, this will act to moderate the impact of boosters per length in rural areas. In essence, boosters/length is therefore a measure of the 'intensity' of pumping requirements within the network. However, this measure gives no indication of how much water is being pumped, or how high it's being pumped.

Boosters per length is primarily associated with higher Capex costs – more assets to inspect and maintain. Also, boosters and associated pressure release valves generate pressure fluctuations and create weak points in the network that require more repairs.

TWD Average Pumping Head (APH)

APH is the weighted average height that each Ml of water is pumped. This in turn is a function of the physical height lifted (static head) and the frictional effects acting on the pump (dynamic head, driven by the volume and speed of flow and the specific configuration of assets).

Most TWD pumping head is located on water treatment sites (be it borehole pumps after treatment, or high lift pumps from surface water WTWs to service reservoirs). For SVE this is approximately 65% of TWD APH - with the remaining 35% largely being boosters located within the network. This means that, whilst the APH values are explicitly driving cost, they are less related to the 'hilliness' of the supply area.

On average, just 30% of pumps contribute 90% of APH but this varies between 17% and 58% of pumps[3]. Therefore, pumping energy (Opex) requirements will have a limited relationship with boosters per length. This is because these energy intensive pumps contribute just a single booster per water treatment works.

Increasing APH values are associated with topography but also source type and density. For example, a city fed by a river (with associated requirement for low lift pumping) is likely to contribute more to APH than several villages in a hilly area. There is also a requirement in flat areas to generate sufficient pressure for the distribution network. This is often done with a single large booster rather than multiple small boosters. Fundamentally, pumping a lot of water a little contributes equally APH as does pumping a little water a lot – the opposite of boosters per length in this regard.

While APH will have some association with Capex cost, APH is much more strongly associated with Opex, particularly Power.

Functional Form

For the PR24 cost models we are assuming a Cobb-Douglas functional form. Ignoring squared density, this can be written as:

$$\ln(Botex) = \alpha + \beta_1 \ln(Scale) + \beta_2 \ln(Density) + \beta_3 \ln(Network\ Complexity_1) + \beta_4 \ln(Network\ Complexity_2)$$

$$\therefore Botex = e^{\alpha} Scale^{\beta_1} Density^{\beta_2} Network\ Complexity_1^{\beta_3} Network\ Complexity_2^{\beta_4}$$

When the functional form is considered, the reason for including both variables becomes clear – Botex is currently being scaled by the intensity of pumping assets within the network OR the intensity at which those pumping assets are working, rather than a function of the two. As an extreme example, a company in which the entire distribution input moves through a single high lift pump is going to have a very different cost profile to a company in which the same distribution input flows through many

[3] Average Pumping Head: data quality improvement, Turner & Townsend for Ofwat, 24 March 2022, available at: https://www.ofwat.gov.uk/wp-content/uploads/2022/05/Average-Pumping-Head-Data-Quality-Improvement-Final-Report-.pdf, p.38.

shorter network pumps. The models as they stand do not reflect this fact, but the combined model will.

2.1.1 Additional evidence for Boosters per Length and TWD APH measuring different facets of Network Complexity

We would expect to see at least one of two things if Boosters per Length and APH were different ways to describe the same feature – firstly, they should be highly correlated and secondly, when they are substituted for one another in the same model, any changes in efficiency should be modest. Neither of these expectations are met.

Correlation

The Pearson correlations between logged boosters per length, logged treated water distribution APH and, for completeness, logged wholesale water APH are shown in ***Table 3***.

Table 3: Pearson correlations between proposed water network complexity variables

	ln(Boosters per Length)	ln(TWD APH)	ln(WW APH)
ln(Boosters per Length)	1.0000		
ln(TWD APH)	**0.1894**	1.0000	
ln(WW APH)	**0.2715**	0.7119	1.0000

The correlations between boosters per length and the APH variables are weak, indicating that the relationship between the variables is not very strong, i.e. a high APH value does not imply a high boosters per length value – were they two alternative variables measuring the same exogenous factor, it would be likely that a high value for one would suggest a high value for the other.

By way of comparison, we show the Pearson correlations of 'Percent water treated in bands 3-6' and log weighted average complexity in ***Table 4***, and the Pearson correlations between the three logged density measures suggested by Ofwat for its PR24 consultation in ***Table 5***. These are variables that we accept are describing the same costs in different ways, with it being unclear which is the 'correct' option to describe those costs.

Table 4: Pearson correlations between proposed water treatment complexity variables

	% water treated 3-6	ln(wac)
% water treated 3-6	1.0000	
ln(wac)	0.9117	1.0000

Table 5: Pearson correlations between proposed water density variables

	ln(LAD from MSOA)	ln(MSOA)	ln(properties/length)
ln(LAD from MSOA)	1.0000		
ln(MSOA)	0.9552	1.0000	
ln(properties/length)	0.9145	0.9159	1.0000

The correlations displayed in ***Table 4*** and ***Table 5*** are clearly very strong, and significantly higher than we see in ***Table 3***. Fundamentally, we can accept that the treatment complexity and density variables are suitable for triangulation on this basis – they are very similar and describing the same cost. In short, they are alternative ways to describe the same costs driver, we are just unsure of exactly how that variable should be defined.

This does not hold for the network complexity variables – the absence of either misses a legitimate component of costs entirely and as a result, where we triangulate we only cancel out noise and don't improve the ability to expose true efficiency.

Efficiency changes on substitution

If boosters per length and APH were two different ways to describe the same feature, we would also expect that substitution between the two in the same model would only produce fairly modest changes in efficiency. Again, this is not the case.

We assess changes in efficiency by considering both the efficiency scores and the efficiency ranks. For each company we define the average efficiency change in network complexity by taking the average absolute difference in efficiency score and rank for each equivalent model in which only the network complexity variable changes. For example, we define the average absolute difference in rank caused by network complexity for the treated water distribution models to be:

$$AVERAGE(ABS(TWD1_{Rank} - TWD4_{Rank}), ABS(TWD2_{Rank} - TWD5_{Rank}), ABS(TWD3_{Rank} - TWD6_{Rank}))$$

We then take the average of these average changes across the industry.

We assessed these absolute differences for both efficiency rank and efficiency score for network complexity in treated water distribution and wholesale, treatment complexity in water resources plus and wholesale, and density for all three levels of aggregation. The latter two assessments again provide a comparison against variables we consider to be suitable for triangulation. We display our findings in ***Table 6***.

Table 6: Average absolute changes in efficiency rank (score) as a result of changes in definition of a variable intended to describe the same feature

	Network Complexity	Density	Treatment Complexity
TWD	4.14 (0.13)	2.04 (0.06)	N/A
WRP	N/A	1.41 (0.06)	0.39 (0.03)
WW	3.34 (0.12)	1.32 (0.04)	0.64 (0.02)

We can see that the change in the network complexity driver produces far greater swings in efficiency than does density or treatment complexity. In both TWD and WW, the change in the network complexity variable is responsible for more than twice the efficiency change than changes in the density variable, both in terms of rank and score.

We reiterate that our stance is that this is because the network complexity drivers are fundamentally different, whereas the treatment complexity and density drivers are not.

2.1.2 Our proposed remedy: Boosters per length and TWD APH produce better TWD models when combined rather than considered separately

We have found that combining APH and boosters per length in the same model produces better performing models than using APH and boosters per length separately[4]. This is a finding corroborated by CEPA in their *PR24 Wholesale Base Cost Modelling* report[5]. It should be noted that CEPA only considered using both network complexity variables alongside Weighted Average LAD Density as per PR19. However, the same holds for models that consider Ofwat's PR24 density specifications.

[4] Models with both network complexity variables show improvements in R-squared, R-squared in unit cost specifications which indicates more or the variation around a simple unit cost model is accounted for, AIC and BIC. The RESET test is satisfied in all specifications, and other robustness tests are passed.

[5] PR24 Wholesale Base Cost Modelling, CEPA for Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/CEPA_Ofwat_Base_Cost_Models_Final_Report.pdf, p.60.

We show our output for TWD and WW models that include both APH and Boosters per Length but otherwise follow Ofwat's PR24 specifications in ***Table 7*** and ***Table 8*** respectively. We can see that all of these models provide improvements in R-squared over and above those put forward by Ofwat in its PR24 consultation. The RESET test is passed in all cases. Whilst we have not reported the AIC or BIC here, these tests of the relative merit of adding extra variables both offer substantial improvements over Ofwat's PR24 suite. This suggests there is little likelihood of overfitting, and the engineering logic has been improved.

Table 7: TWD PR24 specifications combining boosters per length and TWD APH in the same model with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	TWD1	TWD2	TWD3
Log lengths of main	1.063*** (0.000)	1.017*** (0.000)	1.060*** (0.000)
Log boosters per length	0.346*** (0.004)	0.305*** (0.004)	0.387*** (0.002)
Log APH TWD	0.314*** (0.000)	0.380*** (0.000)	0.339*** (0.000)
Log LAD from MSOA	-2.754*** (0.000)		
Log LAD from MSOA sq	0.219*** (0.000)		
Log MSOA		-5.788*** (0.000)	
Log MSOA sq		0.405*** (0.000)	
Log Properties/Length			-14.232*** (0.000)
Log Properties/Length sq			1.811*** (0.000)
Constant	2.549** (0.029)	14.570*** (0.000)	21.933*** (0.000)
Observations	187	187	187
R-squared Adjusted	0.965	0.968	0.971
RESET	0.325	0.858	0.851
AIC	-136	-142	-148
Efficiency Range	0.62	0.61	0.56

*Table 8: WW PR24 specifications combining boosters per length and TWD APH in the same model with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WW1	WW2	WW3	WW4	WW5	WW6
Log	1.069*** (0.000)	1.061*** (0.000)	1.047*** (0.000)	1.043*** (0.000)	1.037*** (0.000)	1.031*** (0.000)
% water treated in bands 3-6	0.003** (0.011)		0.003** (0.028)		0.003*** (0.006)	
Log weighted average complexity		0.321** (0.034)		0.287** (0.050)		0.334** (0.019)
Log boosters per length	0.358** (0.011)	0.354*** (0.009)	0.420*** (0.003)	0.409*** (0.003)	0.322** (0.030)	0.306** (0.027)
Log APH TWD	0.285*** (0.007)	0.270** (0.015)	0.298*** (0.007)	0.285** (0.012)	0.249** (0.037)	0.233* (0.062)
Log LAD from MSOA	-1.931*** (0.000)	-1.769*** (0.000)				
Log LAD from MSOA sq	0.136*** (0.000)	0.124*** (0.000)				
Log MSOA			-5.015*** (0.000)	-4.748*** (0.000)		
Log MSOA sq			0.322*** (0.000)	0.303*** (0.000)		
Log Properties/Length					-10.787*** (0.000)	-10.044*** (0.000)
Log Properties/Length sq					1.263*** (0.000)	1.171*** (0.000)
Constant	-3.207*** (0.008)	-3.814*** (0.002)	10.080*** (0.002)	8.964*** (0.010)	13.447*** (0.003)	11.811*** (0.008)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.970	0.971	0.969	0.970	0.969	0.97
RESET	0.645	0.544	0.693	0.577	0.700	0.394
AIC	-165	-167	-165	-167	-164	-166
Efficiency Range	0.57	0.61	0.58	0.62	0.56	0.57

3. Water Treatment Complexity

We are concerned that the water treatment complexity explanatory variables used in Ofwat's consultation models do not accurately capture the true cost pressures associated with water treatment. We also think that there are some issues being created by outlier observations.

We set out two remedies to the current modelling approach that enable more engineering logic to be accounted for, provide improved statistical performance, and moderate the outlier issues. These are:

- Adding Water Treatment APH, whether individually or as part of Water Resources Plus APH, to capture the differences in cost pressures amongst processes within current treatment complexity bands.
- Re-specifying weighted average complexity to more accurately capture the cost pressure differences between bands.

In its PR24 cost modelling consultation, to justify the inclusion of percentage of water treated in bands 3-6, Ofwat state that "the complexity bands are the closest approximation of water treatment complexity available in companies' annual reporting, and are well established and understood in the sector."[6] To justify the inclusion of the current weighted average complexity specification, Ofwat state that "the current weights… are simple and intuitive"[7].

We accept the above statement, but we do not think that this data is effectively describing the cost pressures that companies face with regards to water treatment complexity. The data being well established and understood does not necessarily mean it is effectively describing what the cost drivers intended. We also consider that the current weights used in calculating the weighted average complexity measure are no simpler, better supported or more intuitive that any other set of weightings that could be used. Whilst we do not necessarily have a problem with the inclusion of these variables, if they are used we must then account for the differences in unit cost that transpire that are a result of legitimate cost drivers.

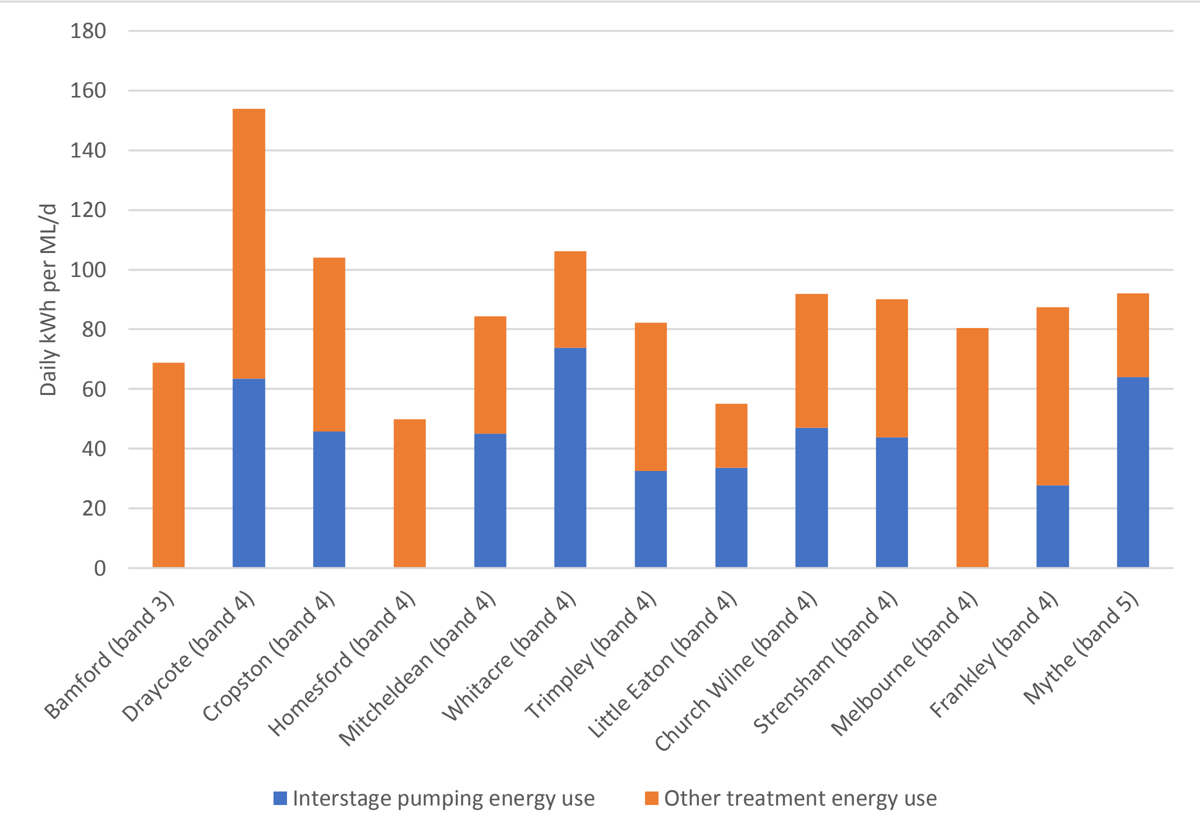
There are material treatment cost differences between companies that have corresponding levels of the water treatment complexity. This is illustrated in ***Figure 3*** which shows differences in power usage across our surface water WTWs[8]. The analysis shows the following.

- The variance does not appear to be driven by the Ofwat treatment complexity categorisation.
  - Mythe WTW (band 5) and Bamford WTW (band 3) are not materially more (less) power intensive.
  - There is material variance across the remaining band 4 WTWs.
- Treatment power costs are dominated by interstage pumping costs (ranging between 32% and 70% of total WTW power usage). This is pumping at WTWs to allow sufficient head for treatment processes to work (this is largely driven by the need for a GAC treatment process for pesticide removal).

[6] Ibid. p.22.
[7] Ibid.
[8] Power is the largest component of our Water treatment opex budget (ranging between 40-50% depending on the input price pressure of energy).

- There is significant variance between interstage pumping requirements. This is largely a function of the topography of the site and the need to add retrospective treatment processes (GAC) that cannot be accommodated in the hydraulics of the existing WTW.
- The other power costs largely relate to Clarification, Filtration, GAC, Disinfection and sludge treatment. Variance typically relates to:
  - the type of clarification installed (DAF is a more power intensive clarification process than HBC);
  - the method of disinfection used (i.e. on-site production of Sodium Hypochlorite (a power cost) or purchase of chlorine (a chemical cost)); and
  - the amount of sludge produced and how it is disposed of (on site treatment to cake or disposal of sludge to sewer).

Figure 3: Power usage across surface water WTWs ordered by complexity band then ML/d output. On site power usage is split between power used in interstage pumping and all other onsite power usage.



If these observed differences in costs at WTWs of equivalent Ofwat treatment complexity band are not accounted for in the models, this resultant variance is seen as inefficiency. The current models generate efficiency scores as low as 0.47 and as high as 2.02, implying that companies are spending as little as half or as much as double what they should be spending. This seems unrealistic and strongly points to a modelling misspecification.

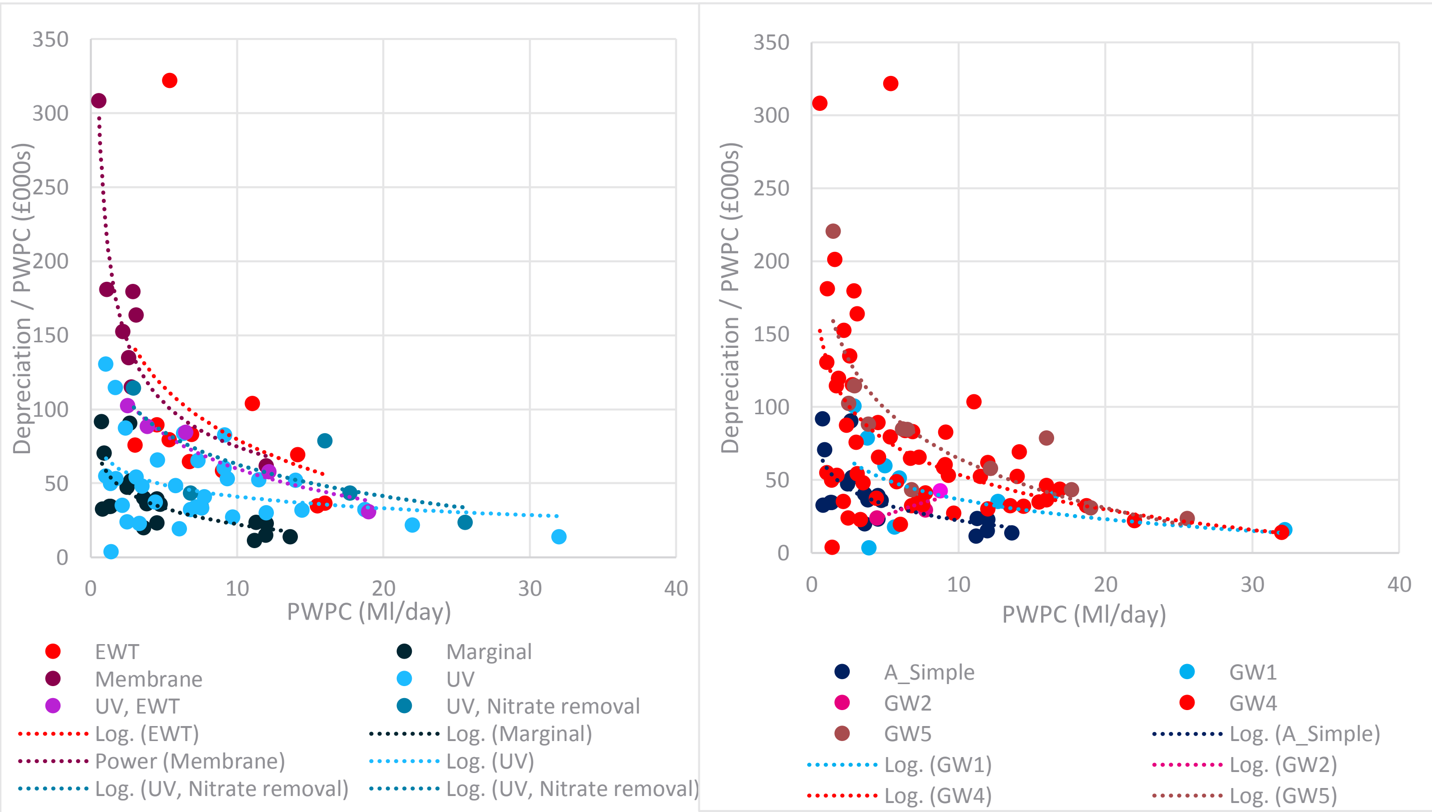
Variables exist which are consistent with engineering logic, provide better overall statistical performance, and go some way to moderating these extreme efficiency scores. Where variables backed by engineering logic are available that constrict the range of efficiency scores, this strongly suggests that the cost differences are driven by legitimate cost drivers and should not be attributed to efficiency.

We are also concerned about Ofwat's assertion that taking logs of the current weightings implies "lower levels of complexity are more expensive than higher levels of complexity"[9]. Rather it implies a diminishing *increase* in costs as we progress through the bands. This seems somewhat sensible given an additional process in a higher band generally requires processes in the lower bands to be operated regardless. As an example, all of our surface water sites that use Activated Carbon (W4) require some form of filtration and disinfection (W1). Treatment works of a given size will also have a fixed base level of cost common to all works regardless of the treatment complexity present on site.

3.1 Differentiating water treatment complexity cost drivers

Our main concern with the current water treatment complexity variables is that they do not adequately differentiate between treatment processes within bands. We showed in our January submission that there are material cost differences between the treatment processes within bands, most notably that UV treatment is considerably cheaper than other processes listed as a band 4/5 treatment process. This is illustrated in ***Figure 4*** below which shows the relationship between Depreciation, Size, Treatment process and Ascribed Treatment complexity band at Groundwater sites.

Figure 4: Graphs showing the relationship between Groundwater depreciation values by size (Peak Week Production Capacity PWPC), installed treatment process (left) and treatment complexity band (right)



[9] Ibid.

There is also some academic evidence that Ozone is more expensive on a unit cost basis than UV[10], and evidence from the United States EPA that membrane filtration and adsorption are both significantly more expensive than UV on an annualised unit cost basis for general treatment works designs[11].

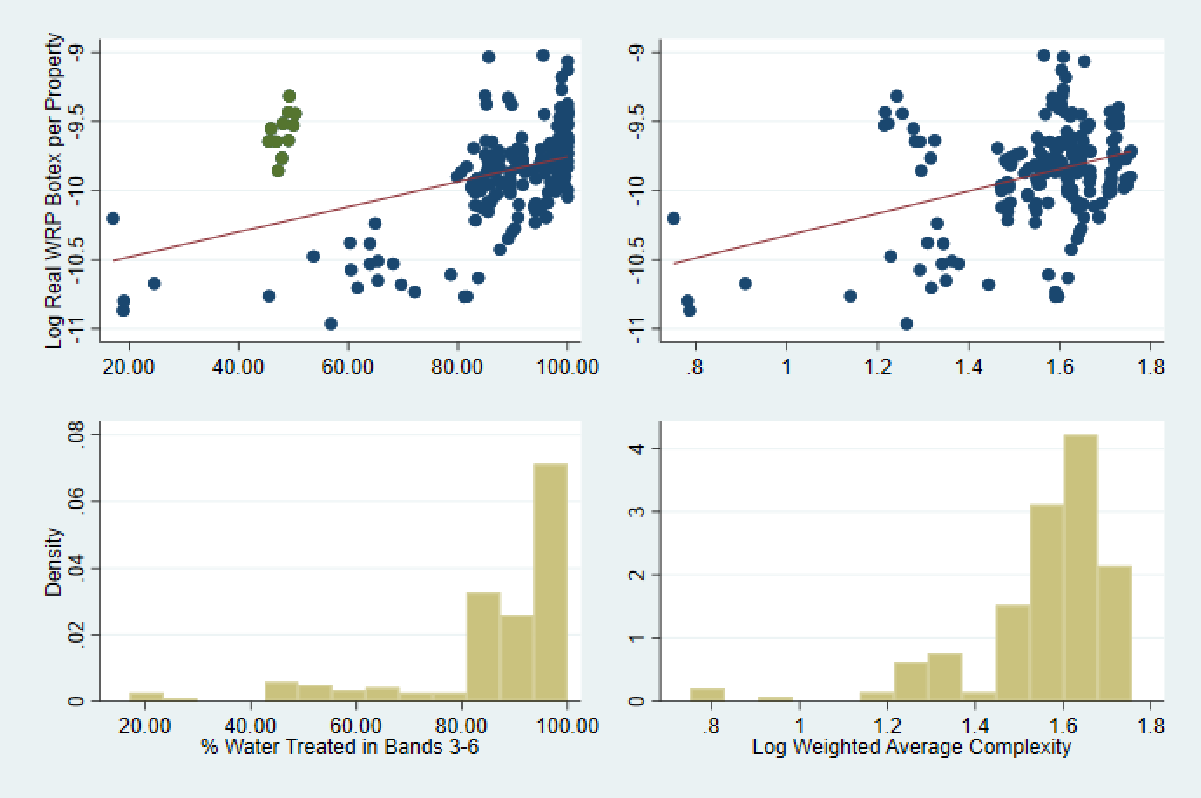
Water treatment processes are driven by raw water quality – some processes are only suitable where certain conditions are met. Different processes have different power, maintenance, and labour requirements. Additionally, where an additional process needs to be added because of changes in statutory requirements or raw water deterioration, this has to be done within a confined space that often leads to interstage pumping that would not have been originally foreseen or designed.

We can show that there is a vast range of unit costs associated with the same level of complexity as defined by defined by the current treatment complexity variables. ***Figure 5*** shows the relationship between the current treatment complexity variables and WRP unit costs. There is a great deal of variation unaccounted for when a linear relationship is assumed here (as per the red trendline, which equates to adding these variables into models). Wessex is highlighted for the % Water Treated in Bands 3-6 because it is found to be an outlier by Cook's distance (see ***Chapter 5.3***) in the PR24 models that include this variable.

Figure 5 also shows that the variation in each of these variables is limited, with the majority of observations grouped towards the right. This is also shown in the histograms in ***Figure 5***. Only three companies have a mean percentage of water treated in bands 3-6 under 80% over the course of the historical data (PRT, WSX, SSC). The ability for the model to differentiate between company water treatment cost pressures will therefore be limited as they are simply describing the difference in costs between these atypical companies and all other companies collectively. This limits the variation in unit costs away from a basic unit cost model. This is responsible for the relatively low R-squared values in the Water Resources Plus models.

[10] Arian Khaleghi Moghadam & Mohammed Dore, 2012. *Cost and Efficacy of Water Disinfection Practices: Evidence from Canada*, Review of Economic Analysis, Digital Initiatives at the University of Waterloo Library, vol. 4(2), pp. 209-223, December.
[11] United States Environment Protection Agency, *Drinking Water Treatment Technology Unit Cost Models*. Available at: https://www.epa.gov/sdwa/drinking-water-treatment-technology-unit-cost-models, Accessed: 03 May 2023

Figure 5: Scatter graphs and histograms showing the relationship between Ofwat's chosen treatment complexity variables & Botex, and the distribution of those variables respectively.



3.2 Proposed remedies

3.2.1 Remedy 1: Using Water Treatment APH

The inclusion of Water Treatment APH, whether as a standalone variable or as part of Water Resources Plus APH, could be a simple remedy to the concerns that we have identified. The addition of pumping before the works (i.e. water resources) will be discussed in ***Chapter 4***, and its engineering logic is clear – there is a trade-off between pumping before the works and after the works, but more pumping incurs more cost. There needs to be an allowance for companies that pump before the works where we consider that treated water APH is a legitimate cost driver.

As already illustrated in ***Figure 3*** above, the addition of Water Treatment APH can be considered a way to differentiate between the processes in each complexity band. It is also a way to account for retrospective treatment installation – and the typical corresponding increase in interstage pumping – as a result of evolving DWI/raw water quality requirements. This incremental increase in treatment pumping head as quality requirements increase has been seen across many of our water treatment works. For example, at Strensham WTW where interstage pumping was introduced as a result of the need to add a GAC process to manage growing water quality risks.

We can show that the combination of Water Treatment APH or Water Resources Plus APH with the current treatment complexity variables improves the relationship with unit WRP Botex substantially and provides a great deal more variation with which to reasonably distinguish between companies.

Figure6 and ***Figure 7*** show scatter graphs and histograms for the water treatment variables multiplied by Water Treatment APH and Water Resources Plus APH respectively. When compared to the current variables (***Figure 5*** above) These show substantially improved relationships between the augmented variables and unit WRP Botex, and a greater distribution between the variables that provides a better opportunity for the models to differentiate between companies. Treatment APH can be allowed for alongside water resources or separately.

- ***Table 14*** in ***Chapter 4*** shows improvements in R-squared when WRP APH is accounted for (i.e. Water resources and treatment APH together). The addition of WRP APH into the WRP models also moderates the extreme efficiency challenges somewhat, with a low of 0.48 (just 0.01 above the proposed PR24 suite) and a high of 1.86 (0.16 below the proposed PR24 suite).
- The inclusion of Water Treatment APH as a standalone variable offers further improvements, however, with R-squared improvements in the region of 0.02 in each model, and a substantial tightening of the extreme efficiency scores, with a low of 0.54 (0.07 above the proposed PR24 suite) and a high of 1.70 (0.32 below the proposed PR24 suite). The maximum efficiency score range is reduced from 1.51 to 1.14, a reduction of 0.37. In all models, it is significant and the RESET test is satisfied. These results are displayed in ***Table 9***.

Either remedy represents a substantial swing in efficiency scores that is driven by a legitimate driver of costs. Swings in efficiency this large cannot be ignored when we have a variable to describe that change. We accept Ofwat's assertion that some companies have increased expenditure in recent years in an effort to catch up with the sector[12], but these cost increases are still exposed as inefficiencies with these better specified models.

Ofwat argued in its consultation that "the quality of disaggregated APH data in water treatment is also too poor to consider including in the models"[13], but ***Table 12*** and ***Table 13 (Chapter 4)*** shows that this lack of confidence in Water Treatment APH has limited justification. While Water Treatment APH does have the lowest measurement rates, these do not appear to be sufficiently lower than TWD APH to reject its use if we accept that TWD APH is now reasonable to include.

[12] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.33.
[13] Ibid., p.22.

Figure 6: Scatter graphs showing unit WRP costs against treatment complexity variables multiplied by Water Treatment APH and histograms for these augmented water treatment variables.

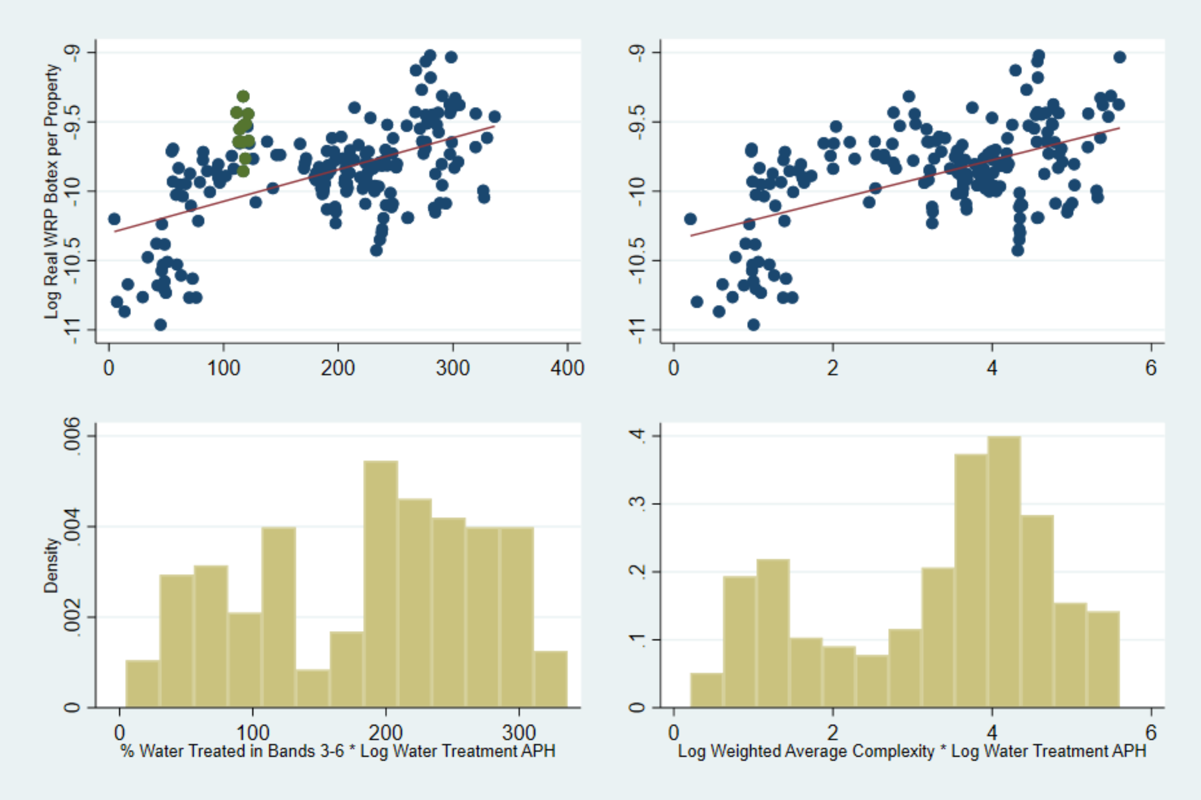


Figure 7: Scatter graphs showing unit WRP costs against treatment complexity variables multiplied by Water Resources Plus APH and histograms for these augmented water treatment variables.

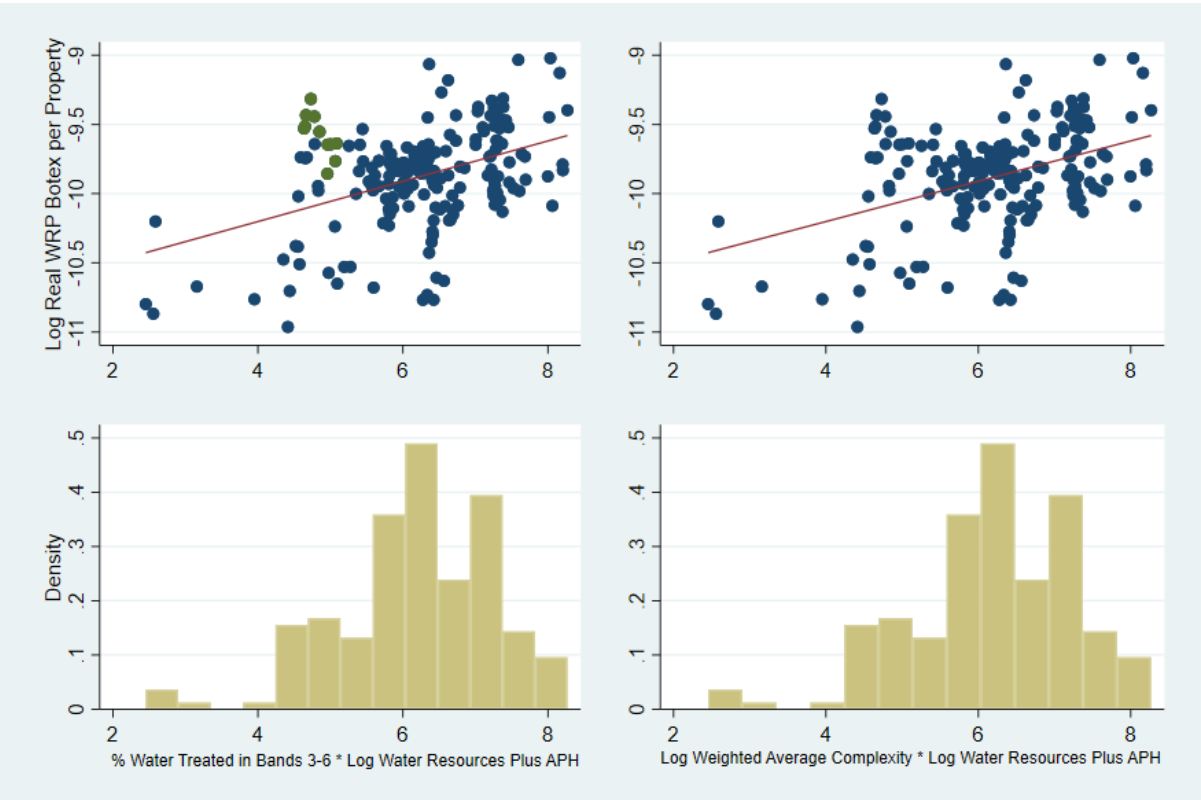


*Table 9: WRP PR24 specifications with the addition of WT APH with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.066*** (0.000)	1.064*** (0.000)	1.049*** (0.000)	1.050*** (0.000)	1.028*** (0.000)	1.027*** (0.000)
% water treated in bands 3-6	0.004** (0.019)		0.004* (0.052)		0.004** (0.020)	
Log weighted average complexity		0.303 (0.243)		0.275 (0.309)		0.315 (0.219)
Log APH WT	0.114* (0.066)	0.120* (0.054)	0.112* (0.097)	0.116* (0.087)	0.122** (0.045)	0.127** (0.038)
Log LAD from MSOA	-1.140** (0.018)	-1.030* (0.051)				
Log LAD from MSOA sq	0.069** (0.027)	0.061* (0.075)				
Log MSOA			-3.833** (0.014)	-3.808** (0.033)		
Log MSOA sq			0.230** (0.014)	0.228** (0.034)		
Log Properties/Length					-6.279** (0.024)	-5.733** (0.044)
Log Properties/Length sq					0.680** (0.039)	0.614* (0.065)
Constant	-6.841*** (0.000)	-7.276*** (0.000)	4.726 (0.418)	4.569 (0.502)	3.472 (0.551)	2.258 (0.704)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.927	0.923	0.920	0.918	0.929	0.926
RESET	0.515	0.387	0.743	0.686	0.546	0.342
AIC	-24	-22	-21	-20	-24	-22
Efficiency Range	1.11	1.09	1.15	1.15	1.08	1.07

3.2.2 Remedy 2: Re-specifying Weighted Average Complexity

Three companies (TMS, YKY, SVE) put forward alternative weightings for the weighted average complexity variable. In our view, Thames' suggestion is limited because it assumes the same cost at all works in bands 3-6, and the same costs in bands s-2. This doesn't seem reasonable. Yorkshire Water and Severn Trent both put forward variables that improve the models, are significant in all models, and allow for increasing costs across bands.

To date, there has been no evidence put forward as to what the correct weightings are. We know that, in general, costs increase as the band increases, but the imposition of a linear relationship on the basis of it being simple and intuitive does not seem sensible. Rather, given the information we have available to us at present, we should attempt to select the best set of weightings subject to certain criteria being met, i.e. cost does not decrease as the band increases, that best fits the data we observe.

Table 10 shows the proposed PR24 WRP specifications where weighted average complexity was included with Ofwat's specification replaced by YKY's or SVE's. YKY's shows R-squared improvements across the board, while SVE's does not. Crucially, both of these Weighted Average Complexity drivers are far more statistically significant that Ofwat's, which was considered to be of high importance by Ofwat in its model selection process. These variables may offer suitable alternatives to simple linear weightings that we urge Ofwat to consider.

Remedy 1, the addition of Water Treatment pumping, is better supported than simply changing the weighted average complexity weightings and serves as a remedy to our concerns across all models, not just those with weighted average complexity. However, it should be noted that these two remedies are not mutually exclusive. The addition of Water Treatment APH to the models improve those with alternative weighted average complexity measures, as shown in ***Table 11***. These models all show substantial improvements over the proposed PR24 suite by R-squared and offer further reductions in the efficiency range, with the maximum efficiency range reduced by 0.3.

Table 10: WRP PR24 WAC specifications substituting Ofwat's Weighted Average Complexity variable for YKY's (WRP1, WRP2, WRP3) and SVE's 'altwac2' (WRP3, WRP4, WRP5) with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.069*** (0.000)	1.051*** (0.000)	1.023*** (0.000)	1.035*** (0.000)	1.001*** (0.000)	0.989*** (0.000)
Weighted average complexity with Log Weightings	0.314* (0.078)	0.283 (0.134)	0.33* (0.052)			
Log 'altwac2'				0.408* (0.053)	0.409 (0.107)	0.445** (0.037)
Log LAD from MSOA	-1.398** (0.025)			-1.348** (0.044)		
Log LAD from MSOA sq	0.086** (0.030)			0.085** (0.042)		
Log MSOA		-4.787** (0.037)			-3.984 (0.119)	
Log MSOA sq		0.289** (0.037)			0.245 (0.112)	
Log Properties/Length			-7.026** (0.036)			-7.611*** (0.10)
Log Properties/Length sq			0.762* (0.050)			0.850** (0.012)
Constant	-5.764*** (0.001)	8.692 (0.315)	5.385 (0.439)	-6.334*** (0.002)	5.041 (0.610)	5.936 (0.348)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.907	0.899	0.909	0.901	0.892	0.901
RESET	0.363	0.729	0.201	0.379	0.732	0.404
AIC	-20	-18	-19	-20	-18	-19
Efficiency Range	1.47	1.50	1.46	1.66	1.63	1.65

*Table 11: WRP PR24 WAC specifications substituting Ofwat's Weighted Average Complexity variable for YKY's (WRP1, WRP2, WRP3) and SVE's 'altwac2' (WRP3, WRP4, WRP5) and adding Water Treatment APH with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.059*** (0.000)	1.046*** (0.000)	1.024*** (0.000)	1.026*** (0.000)	1.002*** (0.000)	0.996*** (0.000)
Weighted average complexity with Log Weightings	0.273 (0.132)	0.245 (0.211)	0.284* (0.117)			
Log 'altwac2'				0.348** (0.045)	0.349 (0.100)	0.363** (0.034)
Log WT APH	0.117* (0.065)	0.114* (0.098)	0.123** (0.049)	0.126* (0.058)	0.124* (0.073)	0.132** (0.045)
Log LAD from MSOA	-0.995** (0.047)			-0.843 (0.122)		
Log LAD from MSOA sq	0.059* (0.071)			0.050 (0.148)		
Log MSOA		-3.634** (0.034)			-2.671 (0.148)	
Log MSOA sq		0.217** (0.036)			0.161 (0.146)	
Log Properties/Length			-5.533** (0.046)			-5.444** (0.27)
Log Properties/Length sq			0.591* (0.069)			0.593** (0.037)
Constant	-7.278*** (0.000)	3.994 (0.539)	1.945 (0.737)	-8.094*** (0.000)	-0.177 (0.980)	1.249 (0.811)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.926	0.919	0.929	0.925	0.918	0.927
RESET	0.410	0.701	0.381	0.524	0.683	0.597
AIC	-23	-20	-23	-24	-21	-24
Efficiency Range	1.09	1.14	1.07	1.20	1.21	1.18

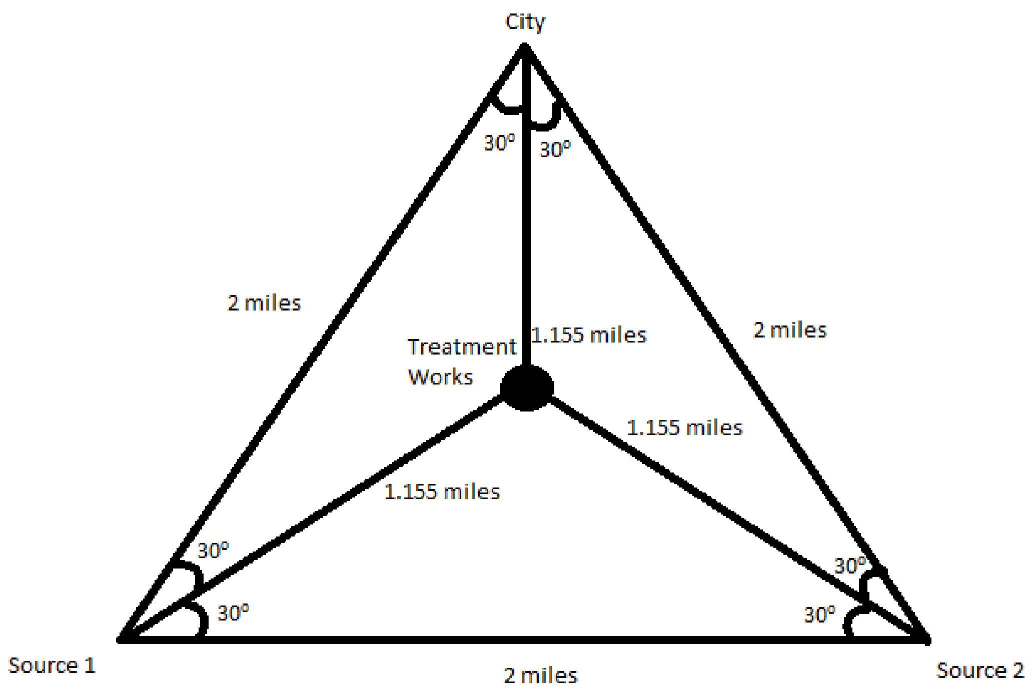
4. Water resources cost drivers

4.1 Pumping before the distribution network (Water resources APH)

We are concerned about the lack of pumping before the distribution network (i.e. WRP models that account for water resources and treatment). Ofwat's claim that 'TWD APH is **by far** the largest contributor to wholesale water APH (57.7%)'[14] is too bold an assertion. It feels wrong to not account for 42.3% of pumping where is has been accepted that the remaining 57.7% is a legitimate driver of cost in TWD.

There is a trade-off between pumping before the works and pumping after it, and for the most part, these decisions were made prior to privatisation (i.e. by the relative siting of WTWs between its water resource and supply location). There would be limited benefit and it would be extremely costly to change where the treatment works are situated, so the placement of the works has to be considered exogenous.

Figure 8: Illustrative example of two sources feeding a single works which feeds a single population centre, showing that a shorter distance is required when pumping occurs before the works. The location of the treatment works could be anywhere within the area of the triangle. This is typically driven by geographical constraints at the time of construction and becomes a sunk cost from that point. Where there is a single source and a single demand node, the location of the treatment works should be irrelevant. However, where Raw Water Pumping is not accounted for it will have a major impact on efficiency.



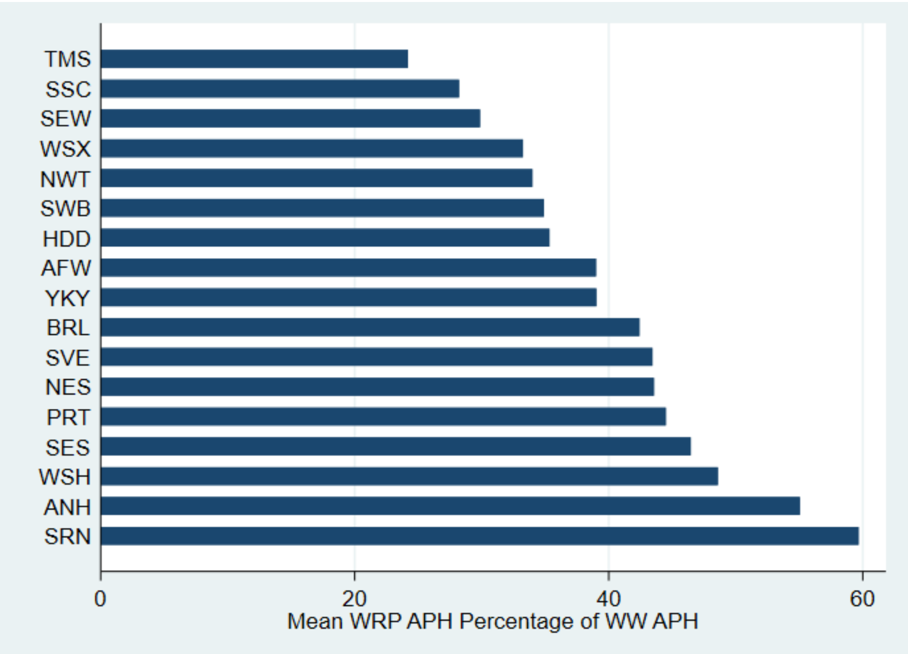
[14] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.23

In the current modelling suite we are accepting that pumping after the works is a legitimate cost but fail to make any allowance for companies that need to pump before the works. This assumes it is inefficient to pump before the works, which is simply not the case. Indeed, dependent on where sources are situated, there could be cases where it is more efficient to pump before the works, such as where multiple sources feed a works which in turn predominantly feeds a single city. This is shown in ***Figure 4***. In this example, if the works were co-located with a source, a total of 2 miles would need to be travelled, 50% before the works from one source to the other, and 50% after the works from the works to the city. ***Figure 4*** shows that where the works is situated at the centre of the triangle, 67% of travel is before the works (2.310 miles), and just 33% after the works (1.155 miles), despite a lower overall travel distance of 3.465 miles compared to 4 miles where a works is located with one of the sources. Assuming pumping requirements per mile are the same along each side of the triangle, a company that co-locates with a source will be deemed more efficient by the current models because more of their pumping requirements are in the network, despite the set-up being less efficient and requiring more pumping. Companies with larger APH levels before the works are being materially disadvantaged by this, while those with the good fortune to have inherited works close to sources are benefitting.

4.1.1 Our proposed remedy: Adding WRP APH as an explanatory variable

Engineering logic of WRP APH and potential for omitting major cost drivers.

We consider that the engineering logic for the inclusion of WRP APH is clear – it is the same as TWD



APH (as we have described above). In the current models, companies that have to pump a lot of water between the source and the treatment works are being materially disadvantaged. Despite 57.7% of APH being in the treated water distribution network at an industry level, for some companies, more than 50% of their pumping is performed before this stage. ***Figure 9*** shows the average percentage contribution of WRP APH to WW APH in the historical data. Companies for which WRP APH makes up less of the total WW APH measure will have a comparative advantage in the models in their current form. Figure 9: Historical average contribution of WRP APH to WW APH by company.The engineering logic would suggest a need to add WRP APH into the Water Resources Plus models and in some capacity into the Wholesale Water models, either as a third complexity variable, or as a composite WW APH variable calculated as the sum of the component WRP and TWD APH values.

Data Quality is not a sensible reason to exclude WRP APH

There have been some valid concerns surrounding the historic quality of APH data. These concerns were explored in detail in the Turner & Townsend APH Data Quality Improvement Report. When considering the extent to which sites are measured or inferred, ***Table 12*** and ***Table 13*** show the number of companies measuring less than a percentage threshold of their sites' lift and volume respectively to calculate their APH values.

The more sites that are measured, the better the APH data quality. This suggests that Raw Water Abstraction is the highest quality APH variable, followed by Raw Water Transport. Treated Water Distribution and Water Treatment APH then appear to have equivalent levels of measured data depending on which measure/ threshold is used[15].

This is made clearer when considering Table 5 in the Turner & Townsend report[16], which suggests that the estimation methods used to estimate sites where measurement is not available are, on the whole, the same for both WT and TWD.

Given this analysis, we do not consider that there is a basis to accept the quality of APH data for the TWD models but reject it for WRP (RW Abstraction, RW transport and Treatment combined) where the proportion of measured (rather than inferred) data appears to be higher.

Table 12: Number of companies measuring lift at less than a percentage threshold of their sites for their APH calculations. Numbers from the Turner & Townsend APH data quality report[17].

Threshold	RW Abstraction	RW Transport	Treatment	TWD
90%	9	11	12	13
70%	7	9	9	10
50%	4	6	8	7
Average contribution to WW APH	19.8%	14.2%	8.3%	57.7%

Table 13: Number of companies measuring volume at less than a percentage threshold of their sites for their APH calculations. Numbers from the Turner & Townsend APH data quality report[18].

Threshold	RW Abstraction	RW Transport	Treatment	TWD
90%	1	3	5	7
70%	1	3	5	4
50%	0	1	4	2
Average contribution to WW APH	19.8%	14.2%	8.3%	57.7%

[15] Average Pumping Head: data quality improvement, Turner & Townsend for Ofwat, 24 March 2022, available at: https://www.ofwat.gov.uk/wp-content/uploads/2022/05/Average-Pumping-Head-Data-Quality-Improvement-Final-Report-.pdf, pp.22-24

[16] Ibid, p.30.

[17] Ibid., pp.22-23.

[18] Ibid., pp.22-23.

Coherence of models when including WRP APH

Ofwat suggest that in the WRP models, CEPA found that WRP APH was not a statistically significant explanatory variable[19]. For most models this is true, but in general it is only weakly insignificant and in one model it is significant at the 10%. However, it is generally more significant than the selected weighted average complexity explanatory variable[20] where Ofwat has determined that the statistical performance is not a reason to 'disregard this variable'[21] given the sign and magnitude of the coefficient is sensible and the engineering logic is sound.

Again, we reiterate that if the engineering logic for including TWD APH is sound, then it must also be so for WRP APH. The coefficient in each WRP model is in the 0.129-0.191 range. Given APH is strongly associated with power, and power makes up ~14.4% of WRP costs historically, a roughly 0.15% increase in Botex following a 1% increase in WRP APH seems eminently sensible, particularly given that as a variable it accounts for costs over and above just power, e.g. maintenance costs and treatment complexity.

The same logic holds for the inclusion of WRP APH in the wholesale models. The inclusion of WRP APH in the Wholesale models alongside both boosters per length and TWD APH improve the models by R-squared above those proposed by Ofwat in its consultation and even above the combined network complexity models shown in ***Table 8 (Chapter 2.1.2)***. When WW APH is used in place of TWD APH there is still an improvement over the models proposed in Ofwat's consultation, but it offers little to no improvement in R-squared over the models suggested in ***Table 8 (Chapter 2.1.2)***.

Table 14, ***Table 15*** and ***Table 16*** show a series of WRP, WW with WW APH, and WW with WRP APH models respectively that account for WRP APH and otherwise follow the models proposed by Ofwat in its PR24 cost modelling consultation.

Table 14: WRP PR24 specifications with the addition of WRP APH with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.093*** (0.000)	1.094*** (0.000)	1.068*** (0.000)	1.071*** (0.000)	1.060*** (0.000)	1.062*** (0.000)
% water treated in bands 3-6	0.004** (0.023)		0.003* (0.052)		0.004** (0.026)	
Log weighted average complexity		0.263 (0.291)		0.234 (0.358)		0.253 (0.295)
Log APH WRP	0.129 (0.305)	0.139 (0.192)	0.143 (0.260)	0.148 (0.170)	0.185 (0.155)	0.191* (0.084)
Log LAD from MSOA	-1.657*** (0.005)	-1.610** (0.014)				
Log LAD from MSOA sq	0.106*** (0.006)	0.103** (0.015)				
Log MSOA			-5.082** (0.017)	-5.153** (0.030)		
Log MSOA sq			0.312** (0.017)	0.316** (0.028)		

[19] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.22.
[20] There is no WRP model where the Weighted Average Treatment Complexity variable is significant at the 10% level. This is only further weakened by the inclusion of WRP APH, a fact which Ofwat.
[21] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.21.

Log Properties/Length					-10.471*** (0.003)	-10.340*** (0.002)
Log Properties/Length sq					1.175*** (0.006)	1.158*** (0.003)
Constant	-5.688*** (0.002)	-5.948*** (0.005)	8.928 (0.274)	9.131 (0.322)	11.443* (0.098)	11.095* (0.094)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.911	0.905	0.903	0.900	0.914	0.910
RESET	0.672	0.581	0.865	0.804	0.645	0.512
AIC	-20.828	-19.341	-18.319	-17.389	-21.332	-20.084
Efficiency Range	1.335	1.322	1.312	1.312	1.269	1.263

Table 15: WW PR24 specifications, combining boosters per length and APH, and substituting TWD APH for WW APH. p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WW1	WW2	WW3	WW4	WW5	WW6
Log properties	1.091*** (0.000)	1.081*** (0.000)	1.067*** (0.000)	1.062*** (0.000)	1.072*** (0.000)	1.066*** (0.000)
% water treated in bands 3-6	0.002* (0.091)		0.002 (0.148)		0.002 (0.118)	
Log weighted average complexity		0.249 (0.115)		0.216 (0.155)		0.237 (0.117)
Log boosters per length	0.452*** (0.004)	0.443*** (0.003)	0.504*** (0.001)	0.491*** (0.001)	0.303** (0.013)	0.294** (0.021)
Log APH WW	0.319** (0.010)	0.290** (0.015)	0.348*** (0.005)	0.321*** (0.007)	0.365*** (0.003)	0.344*** (0.004)
Log LAD from MSOA	-1.898*** (0.000)	-1.757*** (0.000)				
Log LAD from MSOA sq	0.137*** (0.000)	0.127*** (0.000)				
Log MSOA			-4.670*** (0.000)	-4.441*** (0.000)		
Log MSOA sq			0.305*** (0.000)	0.289*** (0.000)		
Log Properties/Length					-13.740*** (0.000)	-13.060*** (0.000)
Log Properties/Length sq					1.611*** (0.000)	1.527*** (0.000)
Constant	-3.647** (0.038)	-4.073** (0.015)	8.127 (0.114)	7.286 (0.153)	18.510*** (0.000)	17.105*** (0.000)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.967	0.968	0.966	0.967	0.971	0.972
RESET	0.562	0.486	0.661	0.568	0.974	0.929
AIC	-163.345	-165.084	-164.425	-166.09	-170.568	-172.156
Efficiency Range	0.58	0.611	0.584	0.611	0.524	0.552

Table 16: WW PR24 specifications, combining boosters per length and APH, and adding WRP APH. p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WW1	WW2	WW3	WW4	WW5	WW6
Log properties	1.082*** (0.000)	1.073*** (0.000)	1.059*** (0.000)	1.054*** (0.000)	1.065*** (0.000)	1.059*** (0.000)
% water treated in bands 3-6	0.002** (0.048)		0.002* (0.100)		0.002* (0.069)	
Log weighted average complexity		0.268* (0.081)		0.230 (0.120)		0.236* (0.100)
Log boosters per length	0.386*** (0.004)	0.379*** (0.003)	0.447*** (0.001)	0.434*** (0.001)	0.305** (0.011)	0.295** (0.016)
Log APH TWD	0.265** (0.011)	0.255** (0.019)	0.276** (0.012)	0.269** (0.016)	0.207* (0.063)	0.199* (0.082)
Log APH WRP	0.090 (0.141)	0.077 (0.157)	0.103* (0.079)	0.090* (0.098)	0.156*** (0.005)	0.146*** (0.006)
Log LAD from MSOA	-2.014*** (0.000)	-1.865*** (0.000)				
Log LAD from MSOA sq	0.144*** (0.000)	0.133*** (0.000)				
Log MSOA			-5.136*** (0.000)	-4.899*** (0.000)		
Log MSOA sq			0.332*** (0.000)	0.316*** (0.000)		
Log Properties/Length					-13.644*** (0.000)	-12.959*** (0.000)
Log Properties/Length sq					1.601*** (0.000)	1.517*** (0.000)
Constant	-3.272** (0.020)	-3.777*** (0.006)	10.072*** (0.007)	9.146** (0.015)	18.647*** (0.000)	17.201*** (0.000)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.970	0.970	0.969	0.970	0.971	0.972
RESET	0.589	0.57	0.617	0.572	0.939	0.865
AIC	-164.613	-166.359	-165.321	-167.133	-168.163	-169.746
Efficiency Range	0.552	0.59	0.562	0.594	0.498	0.528

5. Taking appropriate account of population density

In our January submission we provided analysis about the risks and opportunities associated with various population density explanatory variables. On reviewing the Ofwat consultation models, we are still concerned that the way density has been treated is likely to lead to spurious results.

Fundamentally, the use of squared density terms implies either:

- that costs both increase (or decrease) in both very urban and rural settings (where the inflection in the cost/density curve is within the modelled observations); or
- that the rate that costs increase (or decrease) accelerate per unit of density (this is where the inflection in the cost/density curve is outside of the plausible range of observations.

For each of Ofwat's proposed water models, the inflection of the cost/density curve is within the observed data set. This means that we must infer that each of WRP, TWD and WW costs, exhibit both rural and urban cost pressures where the curve has a 'u' shaped distribution (or opportunities if the curve has an 'n' shape distribution). We do not believe that there is evidence to support this contention in WRP models.

Our primary concerns here are as follows.

- WRP models should not include a squared density term. We contend that the engineering logic simply doesn't hold for this.
- We accept the premise that TWD costs respond in a non-linear way to density (cost pressures in both urban and rural areas). Whilst this would suggest a case for a squared density term in the TWD models, we are of the opinion that this is not being correctly reflected by the current drivers.
- Density has problems with outliers which have a disproportionate impact on the coefficients. This is largely driving the fitting of the non-linear density drivers.
- There is a case to also use a population/area measure. This is an overall density measure that gives better coverage of rurality cost drivers.

In this chapter we set out the basis for these concerns: Section 5.1 describes the basis for our engineering expectations in both WRP, TWD and WW. Section 5.2 considers the benefit of different density measures. Then Section 5.3 describes to likely impact of outliers on non-linear density coefficients.

Finally, in sections 5.4 and 5.5 we identify some simple remedies that would improve the specification and robustness of the models. Given that density responds differently to Water Resources Plus and Treated Water Distribution costs, these are considered separately. These are summarised in the ***Table 17***.

Table 17

Models	Remedy	Reason
Water Resources & Treatment (WRP)	Remove squared density terms	• There is no engineering basis for squared term. • Inclusion of a squared term makes the model very sensitive to outlier companies (TMS). • There is no statistical performance improvement by including squared term.

Treated Water Distribution (TWD)	• Include models with Population/ area as an additional density variable. And;
	• Either:
	  ◦ Remove TMS from the models as an outlier and retain the squared density term; or
	  ◦ Retain TMS in models but remove the squared density term where weighted density variable[22] is used
		• Population / area provides an overall measure of density that will better allow for rurality cost drivers.
		• TMS is a statistical outlier that is biasing the models.
		• The squared term is currently fitting to TMS rather than describing variations between the wider modelling data set.

Wholesale water models (WW)	Retain squared density terms	• The opposing effects of density on WRP and TWD costs mean a nonlinear distribution should be expected.

5.1 Challenging the engineering basis for density having a non-linear distribution

Ofwat suggests the 'argument for a non-linear relationship between wholesale water costs and population density is clear'[23] and gives a number of reasons as to the engineering logic it considers holds with respect to density[24], which are laid out in ***Table 18***. We contest that population density has fundamentally different impacts on WRP and TWD costs. We set out our interpretation in the table below.

In summary, we expect a negative linear relationship for WRP costs, and non-linear relationship in TWD where costs increase in increasingly urban and rural settings.

We agree that for the wholesale models, these points taken together suggest that there is a case for the inclusion of a squared density term given the relative strength of the counter-acting effects of density in WRP and TWD.

Our assertions on engineering logic are also supported when we remove MSOA density and its squared term, and replace them with rural density and urban density, as defined by MSOA classifications[25]. Given the potential opportunities of these classifications we have appended of the data with this submission. When these variables are included in the models, there is a negative association with WRP Botex as densities increase in both rural and urban areas, which supports our assertion that the true

[22] The weighted average density drivers are 'MSOA' and 'LAD from MSOA'

[23] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.24.

[24] Ibid.

[25] MSOA classification available at: https://www.gov.uk/government/statistics/2011-rural-urban-classification-lookup-tables-for-all-geographies. We have added the classification to the MSOA density derivation file, and calculated urban and rural densities in the 'Water- WAD & WAD at MSOA' tab.

relationship is linear. For TWD Botex, there is a negative association with rural densities and a positive association with urban densities, which suggests a nonlinear relationship between density and cost.

Table 18: Table outlining Ofwat's logical Assertions with respect to a U-shaped density., Blue = Positive cost driver; Purple = Negative cost driver

Ofwat Logic Assertion	Impact on WRP costs	Materiality on WRP costs	impact of TWP costs	Materiality on TWD costs
Companies operating in densely populated areas may have the opportunity to source and treat water using larger and fewer sources / treatment works, leading to lower unit costs.	Costs fall as population density rises due to opportunities for economies of scale	High – Clear opportunities for EOS at Water resources and Treatment assets.	Not relevant to TWD costs	None
Companies operating in densely populated areas may be able to make more efficient use of resources, e.g. reduced travelling distances for maintenance & duplication of depots & spare parts to deliver good service.	Costs fall as population density rises due to opportunities for economies of scale	Low – No. of sites limited therefore journey numbers are low	Costs fall as population density rises due to opportunities for economies of scale	High – Large number of locations /jobs make logistics more important.
Companies operating in densely populated areas may bear higher property, rental, labour, and access costs[26].	Costs increase as population density rises due to greater input price pressures in urban areas	Low – WRP assets not likely to be located in urban areas	Costs increase as population density rises due to greater input price pressures in urban areas	Medium – Company procurement / resourcing strategies largely company wide.
Congestion of underground assets complicates access	Not relevant to WRP costs	None	Costs increase as population density rises	High – Operating in highly urban areas is complex (interaction with other utilities)
Higher electricity requirement to pump water to taller buildings[27]	Not relevant to WRP costs	None	Costs increase as population density rises	Medium – Ground topography is a bigger driver of pumping requirement than built environment
Traffic affects ground movement increasing the frequency of repairs	Not relevant to WRP costs	None	Costs increase as population density rises	High – Traffic loading is a major contributor to network failures.
Longer travel times due to congestion	Costs increase as population density rises due to greater input price pressures in urban areas	Low – Most journeys relate to TWD rather than WRP	Costs increase as population density rises due to greater input price pressures in urban areas	High – Large number of locations /jobs make logistics more important.
Added by SVE – In rural areas, distances between supply & demand increases relative length of network and likelihood for pumping / pressure variations	Not relevant to WRP costs	None	Costs fall as population density rises due to opportunities for economies of scale	High – Increased asset base & pumping requirements directly impact on power & maintenance costs.
Overall impact of density on cost (amalgam of above drivers)	Dominated by opportunities for EOS & geographical / geological circumstance. Other drivers not material. **Therefore, engineering expectation that density is a negative linear cost driver.**	N.A.	Multiple competing cost drivers that increase costs in both rural and urban settings. Urban likely to be stronger than rural. **Therefore, engineering expectation that population density is a complex and likely non-linear cost driver.**	N.A.

[26] We note that property rental and labour and access costs relate to some extent to all costs, but these are all subject to choice and are highly endogenous. For the most part, water company buildings are not situated in the densest areas of a company, and water sources / treatment assets certainly aren't (e.g. raw water reservoirs are largely not located in urban settings and treatment works tend to be positioned either close to the source they're fed from (e.g. Bamford WTW) or on the extreme outskirts of a city (e.g. Frankley WTW).

[27] We note that this cost driver can be explicitly covered by APH.

5.2 The benefits of different density specifications for describing particular density cost drivers

The rural areas are largely ignored by the Weighted Average Density (MSOA, LAD from MSOA) calculations – the 20 densest of Severn Trent's 1152 MSOAs, all of which are in Leicester, Birmingham, Nottingham, Coventry and Derby contribute to more than 11% of the numerator on Severn Trent's MSOA density measure, and the 576 densest MSOAs describe almost 90% of the numerator. This leaves very little density being actively described by the rural areas within a company.

The least dense company, HDD, has over 77% of its MSOA numerator described by 9 of its 62 MSOAs, all of which are in Wrexham. Its 15 densest MSOAs which cover Wrexham, Newtown and Saltney account for 93% of the numerator. In short, this shows that the impact of Rural areas where we assert that costs might increase because of longer travel times, proximity to depots, economies of scale at depots etc., have an extremely limited impact in the calculation of these weighted average measures.

Therefore, the Weighted Average Density measures are not capturing whether a company is more or less rural, but rather capturing the density of a company's most dense areas, all of which will come with the complications, or benefits, of increased density. In essence, these can be viewed as measures of the concentration of a company's population.

In contrast, uniform overall measures of density (Properties per Length, Population per Area) give an overall picture of a company's density. Large rural areas will moderate the impact of very dense population centres where they would not in the Weighted Average Density specifications. This is not perfect – given our contention that costs rise in very rural and very dense areas in Treated Water Distribution for example, then we would ideally not use a measure that allocates the same density allowance to companies with moderate density in their entire supply area as it does to companies that operate across both very dense and very rural areas, but as it stands, the uniform measures give us the best approximation of the overall density picture.

WRP

The benefits of operating in more densely populated regions would be best described by the weighted average measures of density given they reveal most accurately the density of those dense regions in which the company can benefit, despite the properties per length density specification appearing to be most performant by R-squared.

Economies of scale at treatment works are most readily achieved in these very dense areas where very large treatment works can supply large, dense populations. Therefore, it seems that weighting to these highly populated, dense areas (i.e. allowing for a measure that captures the *concentration* of population within a company's supply region, rather than just the overall density) would best expose the opportunities for economies of scale at treatment works.

TWD

Our assertion here is that costs increase in both very rural and very urban areas. The uniform measures of density therefore make more sense, intuitively, to describe the associated cost pressures here. Since the Weighted Average Density measures negate the effect of rural areas, and for all companies the most dense areas will come with problems of congestion, we will not be capturing the increased cost pressures of rurality. Rather, we would expect that costs are increasing monotonically with the Weighted Average Density, as there would be increasing costs associated with operating in denser urban centres.

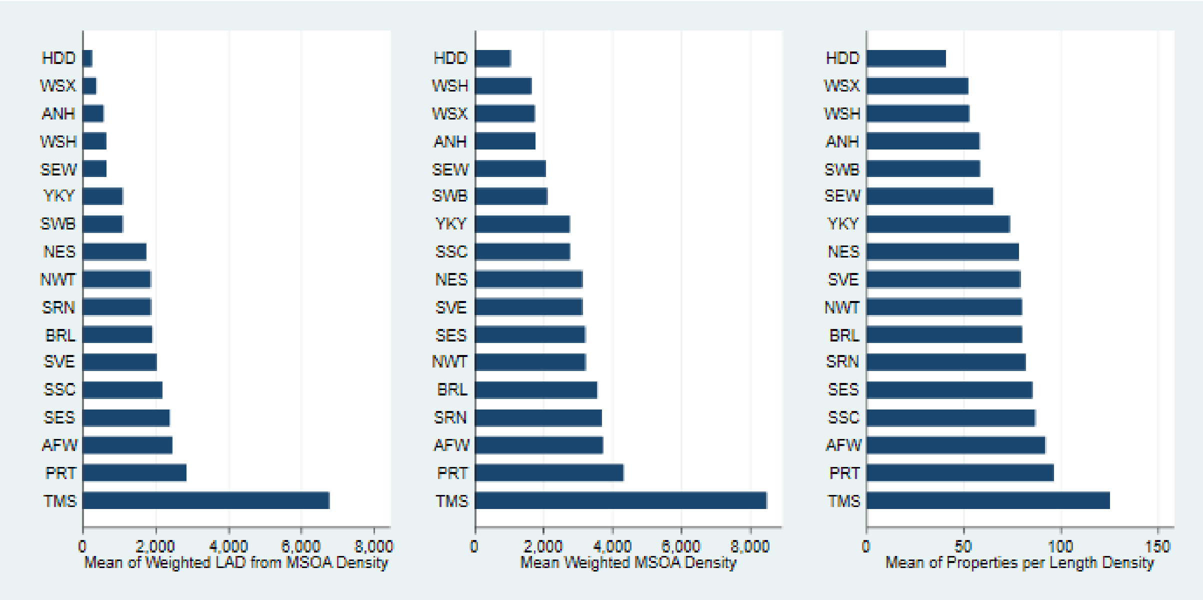
In contrast, the uniform density measures give a better picture of a company's overall density, and allow its rurality to be exposed to some extent. This will allow the increased costs in rural areas to be properly accounted for.

We made reference to government defined rural and urban classifications for each MSOA in ***Chapter 5.1***, and believe these can help to further expose the relationship between density and cost on either side of the distribution.

5.3 The effect of outliers is likely to be providing false confidence in the significance of non-linear density explanatory factors

Outliers are potentially problematic in statistical analysis, particularly where they change the relationships described by the data. ***Figure 10*** shows the mean density for each company in the sample. We can see that for the two weighted average density measures ('LAD from MSOA' and 'MSOA'), Thames is clearly an outlier. It has a greater difference between its density and the second highest density company than the second most dense and least dense company for each of those measures. This is because the weighted average density measures are heavily weighting towards urban, dense, highly populated areas – essentially these are a measure of the density of a company's most dense areas, and not a measure of overall density. Thames is not a clear outlier in Properties per Length, which is a uniform measure of density.

Figure 10: Bar chart showing mean density for each company in the sample for each density variable.



We can quantify the effect that outliers have on the estimated coefficients using Cook's distance, which is commonly used to measure the influence of removing an observation from the modelled sample (or, where we have panel data, removing an entire group from the sample). As a rule of thumb, Cook's distances that are substantially larger than any others in the sample should be treated with caution. Manual checks of the influence of removing that particular group from the sample should then be completed. Some authors suggest that values greater than 0.5 are potentially concerning, while values greater than 1 are considered extreme[28].

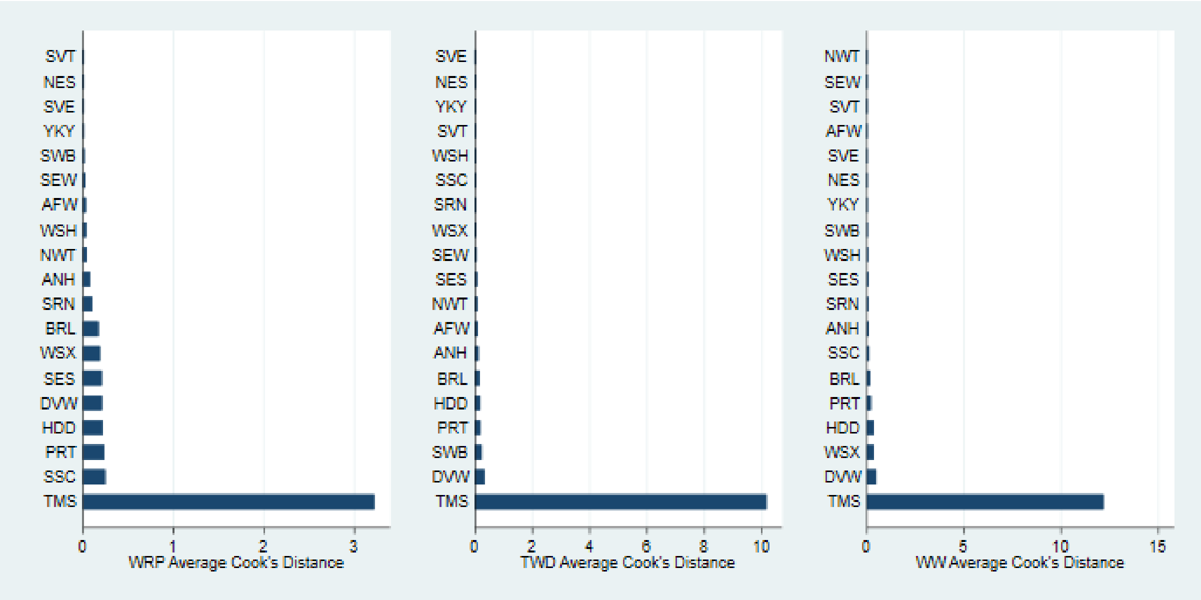
[28] Fahrmeir et al. (2022) *Regression: Models, Methods and Applications*. 2nd edn. Berlin: Springer, p.166.

Thames appears to *significantly* change the estimated relationship between density and Botex. Thames' Cook's distance values are very large – the average values across WRP, TWD and WW models being 3.22, 10.15 and 12.17 respectively. The second highest Cook's distance for WRP, TWD and WW are 0.25 (SSC), 0.24 (DVW) and 0.48 (DVW) respectively. By far the highest Cook's distances are in TWD and WW models that use MSOA density, in one case these models give a Cook's distance of over 30.

Figure 11[29] shows the average Cook's distances across WRP, TWD and WW in Ofwat's proposed PR24 suite. Thames clearly have a substantial influence on the coefficients estimated for the density drivers. We therefore have to be cautious when including these variables, or Thames, in the models.

This distribution means there is a significant risk that the model simply fits between the outlier and the rest of the distribution. Therefore, we need to consider the relationship we expect to see between density and Botex for the non-outlier companies.

Figure 11: Average Cook's distances in Ofwat's proposed PR24 models.

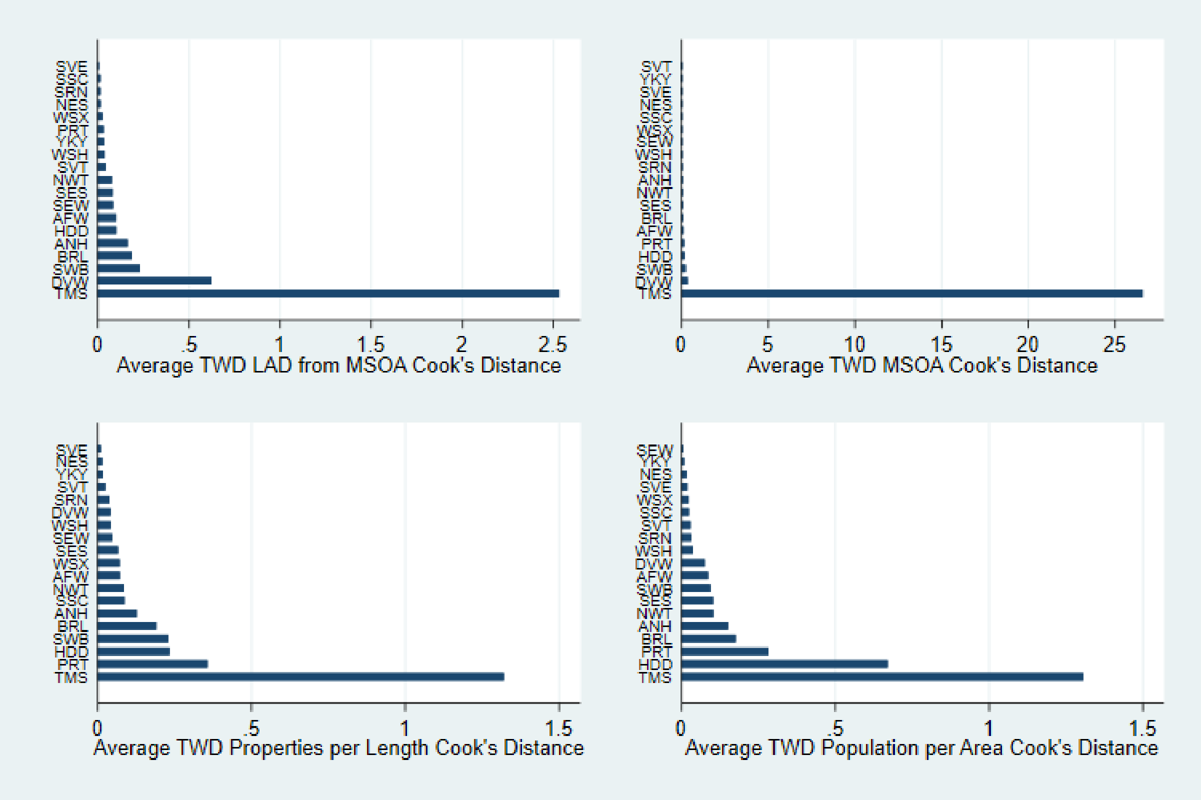


The very large Cook's distance values are significantly moderated where linear density specifications are considered. Population per Area as a density driver also appears to moderate the Cook's distances values somewhat. However, even where these remedies are applied, Thames still appears to unduly influence the coefficients on the density drivers. Where Thames is efficient or inefficient, the squared term will undoubtedly be fitting some of that efficiency.

The extent to which this is a problem depends on the choice of density driver, this is seen in ***Figure 12***. In TWD the uniform measures of density are far less influenced by the inclusion of Thames than are the Weighted Average Density measures, although we should still note that Thames has a significant influence on these variables too.

[29] This uses 'cooksd2' in Stata to calculate Cook's distances with respect to the density variables after estimating each of the proposed PR24 models. The 'panel' options is used which bases the Cook's distance on the removal of an entire group from the models, so for each group (company) the Cook's distance is the same for each year. More information on this module is available at: https://www.stata.com/meeting/uk22/slides/UK22_Vincent.pdf and https://ideas.repec.org/c/boc/bocode/s459149.html

Figure 12: Cook's distance for TWD models for each density specification. Averages triangulating over inclusion of APH or boosters per length.



Thames' observations are legitimate, consequently attempts should be made to retain it in the models. But we do not think that it is appropriate to not address the issue of overfitting that it causes and the skewed coefficients that it will likely generate. We set out potential remedies to WRP and TWD models in sections 5.4 and 5.5.

We note that HDD may also be an outlier in properties per length but there are insufficient observations for them to influence the coefficients in a significant way. The Cook's distance for HDD is above 1 when a regression of just the last 4 years is considered.

5.4 Our proposed remedy in WRP models: Linear density only

We propose that WRP models should include a linear density explanatory variable only. This is because:

- we do not consider that there is a strong engineering basis for the squared term;
- the inclusion of a squared term makes the model very sensitive to outlier companies (TMS); and
- there is no material statistical performance improvement when including the squared term.

As set out in section 5.1, engineering expectation points to falling WRP costs as density increases. This is because companies with large, dense population centres can develop larger water resources and treatment assets that benefit from economies of scale. These are typically remote from urban centres therefore they don't suffer from the 'congestion effects' that you see in the treated water network. Therefore, on an engineering basis, a linear density term should be used.

If a squared term is applied the curve of the term should be very shallow and only be describing subtle variances at the margins rather than showing inflections in the relationship between density and cost

However, as shown in ***Figure 13*** and ***Figure 14*** below, the key weakness with including a squared variable is that it becomes very sensitive to outliers. For example, the exclusion of TMS in a WRP model changes the modelled relationship between cost and density from a ‘u’ shaped curve to an ‘n’ shaped curve. This is highly undesirable as the fundamental relationship between cost and cost driver should not be dependent on one observation.

Figure 13: The relationship between WRP cost and LAD from MSOA density. The curve colours are as follows.

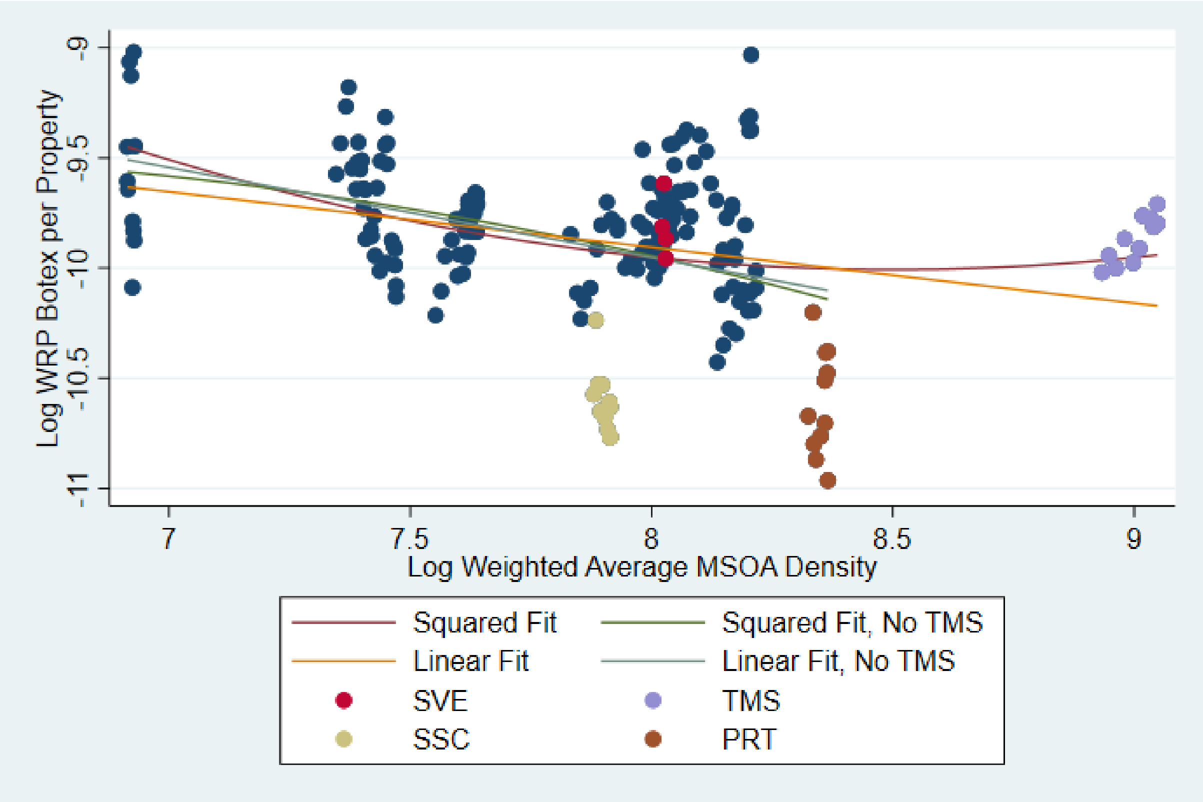
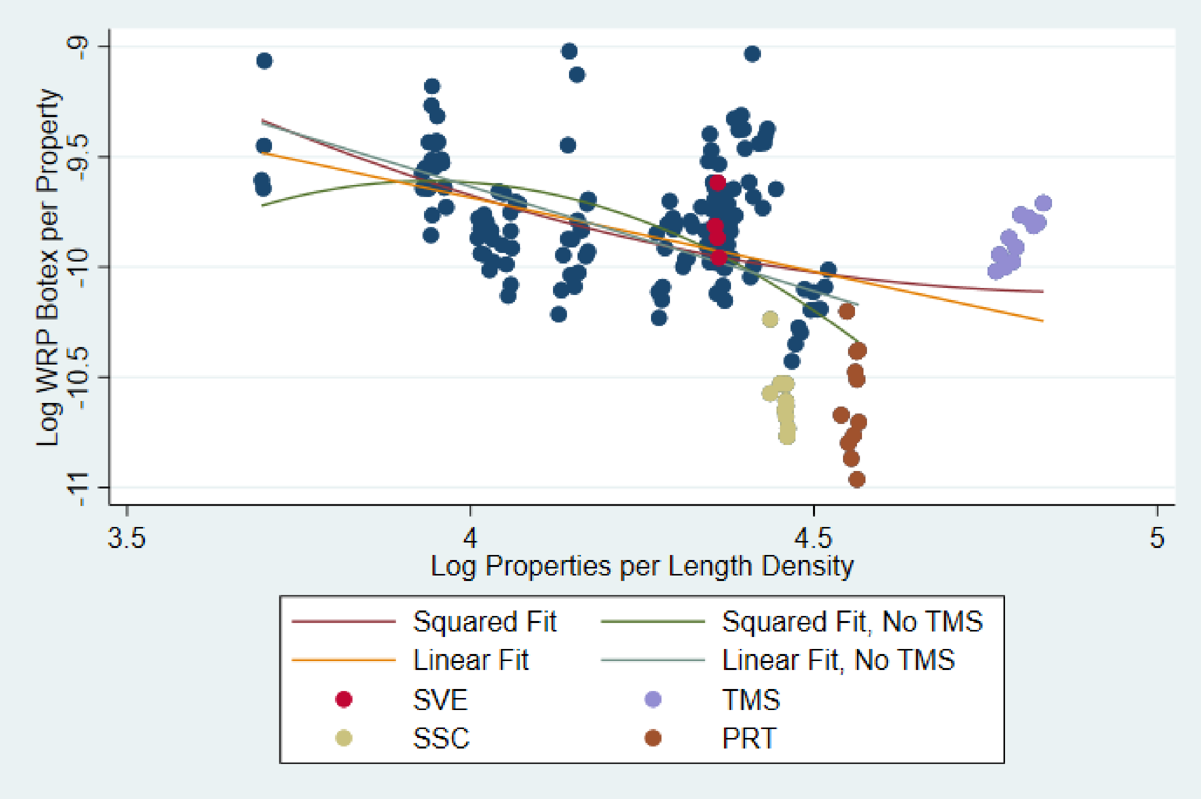


Figure 14: The relationship between WRP cost and LAD from Population/Length density. The curve colours are as follows.



This distortion in the distribution is also likely to be a contributary factor for why PRT and SSC appear to be so efficient in WRP. Assuming either linear or non-linear relationships following the exclusion of TMS shows the benefit that they are likely to be receiving from having relatively high population densities. Both PRT and SSC are shown to still be highly efficient in WRP models with Thames removed but the extent of this efficiency is moderated to a more practicable level.

We reject CEPA's assertion that 'the statistical results of the models excluding the squared term are weak'[30], our analysis shows that statistical performance remains materially constant. However, we reiterate the point that whether to include a variable or not should prioritise on the engineering logic, not on statistical performance. An increase in R-squared that derives from fitting to an outlier exhibiting a relationship that is not consistent with the rest of the industry is not a valid basis for the inclusion of a particular variable. This is, in essence, overfitting.

Model regressions

Table 19 shows WRP model results with the removal of the squared density term. The R-squared values for WRP1 and WRP6 increase by 0.001, stay the same for WRP2 and WRP5, and reduce by 0.004 and 0.003 for WRP3 and WRP4 respectively. These are not considerable changes by any measure and we have to conclude that there is no change in R-squared. The RESET test is similar across the board also, with some improvements and some reductions. Taken alongside the engineering logic, it is clear that the squared term is mis-specified.

[30] PR24 Wholesale Base Cost Modelling, CEPA for Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/CEPA_Ofwat_Base_Cost_Models_Final_Report.pdf, p.35.

Table 20 and ***Table 21*** show the WRP models with and without squared terms respectively, and with TMS excluded from the models. These show the influence that TMS has on these models. ***Table 16*** shows that following the exclusion of both TMS and the squared density variable, there is an improvement by R-squared over the models in ***Table 20*** that exclude TMS but retain the squared density variables. This suggests that a linear specification is a better fit to the rest of the industry. Models WRP5 and WRP6 in ***Table 20*** also show the 'n-shaped' density relationship, albeit with both coefficients being insignificant.

Table 19: WRP PR24 specifications with squared densities removed with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.045*** (0.000)	1.044*** (0.000)	1.014*** (0.000)	1.016*** (0.000)	1.014*** (0.000)	1.013*** (0.000)
% water treated in bands 3-6	0.005*** (0.001)		0.005*** (0.002)		0.005*** (0.001)	
Log weighted average complexity		0.376 (0.136)		0.364 (0.155)		0.396 (0.110)
Log LAD from MSOA	-0.185* (0.050)	-0.191* (0.054)				
Log MSOA			-0.18 (0.303)	-0.203 (0.253)		
Log Properties/Length					-0.521* (0.097)	-0.555* (0.073)
Constant	-9.615*** (0.000)	-9.699*** (0.000)	-9.068*** (0.000)	-9.056*** (0.000)	-8.289*** (0.000)	-8.291*** (0.000)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.907	0.902	0.897	0.893	0.910	0.906
RESET	0.573	0.29	0.546	0.312	0.586	0.339
AIC	-21.21	-19.902	-17.38	-16.419	-20.083	-19.013
Efficiency Range	1.421	1.42	1.452	1.46	1.411	1.411

*Table 20: WRP PR24 specifications with Thames removed with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.063*** (0.000)	1.064*** (0.000)	1.103*** (0.000)	1.112*** (0.000)	1.000*** (0.000)	0.999*** (0.000)
% water treated in bands 3-6	0.005*** (0.001)		0.004** (0.023)		0.005*** (0.000)	
Log weighted average complexity		0.345 (0.174)		0.286 (0.255)		0.386 (0.114)
Log LAD from MSOA	-1.026 (0.618)	-1.024 (0.615)				
Squared	0.059 (0.700)	0.059 (0.698)				
Log MSOA			-10.492 (0.352)	-11.379 (0.288)		
Squared			0.663 (0.368)	0.719 (0.302)		
Log Properties/Length					1.042 (0.921)	1.228 (0.907)
Squared					-0.209 (0.870)	-0.233 (0.857)
Constant	-6883 (0.250)	-7.001 (0.230)	29.796 (0.476)	333.023 (0.405)	-10.946 (0.601)	-11.452 (0.584)
Observations	176	176	176	176	176	176
R-squared Adjusted	0.901	0.893	0.884	0.878	0.910	0.904
RESET	0.623	0.413	0.876	0.736	0.66	0.353
AIC	-10.706	-9.209	-8.241	-7.6	-10.509	-9.183
Efficiency Range	1.478	1.472	1.423	1.447	1.466	1.46

*Table 21: WRP PR24 specifications with squared densities and TMS removed with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6
Log properties	1.039*** (0.000)	1.040*** (0.000)	1.017*** (0.000)	1.019*** (0.000)	1.005*** (0.000)	1.006*** (0.000)
% water treated in bands 3-6	0.005*** (0.001)		0.005*** (0.002)		0.005*** (0.000)	
Log weighted average complexity		0.359 (0.162)		0.353 (0.180)		0.383 (0.131)
Log LAD from MSOA	-0.224** (0.027)	-0.226** (0.042)				
Log MSOA			-0.279 (0.179)	-0.290 (0.190)		
Log Properties/Length					-0.709** (0.033)	-0.719** (0.041)
Constant	-9.262*** (0.000)	-9.377*** (0.000)	-8.338*** (0.000)	-8.401*** (0.000)	-7.371*** (0.000)	-7.478*** (0.000)
Observations	176	176	176	176	176	176
R-squared Adjusted	0.903	0.895	0.892	0.887	0.909	0.903
RESET	0.541	0.345	0.719	0.538	0.663	0.363
AIC	-12.646	-11.157	-8.674	-7.716	-12.332	-10.999
Efficiency Range	1.467	1.462	1.539	1.537	1.469	1.463

5.5 Our proposed remedy in TWD models: Adding population/area to better reflect rurality drivers and managing outlier effects

We propose that the TWD models should also include a specification that includes a population/area density variable. This is because network cost drivers relating to rurality will be better accounted for by this variable (as described in section 5.2).

We also believe that there is clear evidence that TMS is an extreme outlier with respect to weighted average density relative to the rest of the industry which is significantly distorting the models. We suggest that this can be moderated by either: removing TMS from the model; or, removing the squared term in the models that have weighted average density variables ('LAD from MSOA' and 'MSOA', i.e. TWD models 1,2,4 and 5).

5.5.1 The interaction between TWD cost and density is more complex than WRP (cost pressures in rural and urban areas)

As described in section 5.1, costs are likely to rise in both very urban (issues of congestion and complexity in urban settings), and very rural areas (additional assets and distance between assets in rural settings).

This would point to a curved relationship between cost and density where the inflection is within the distribution at the theoretical optimal level of density. Therefore, on an engineering basis, a non-linear density term should be used. However, as set out in section 5.2, the effects of rural companies may not be being adequately reflected in the 'MSOA' and 'LAD from MSOA' density drivers. In such cases the use of a squared term is not likely to adequately describe rurality cost drivers. Therefore, we are proposing that an additional density specification that includes another overall level of density (population/area) which better accounts for rurality is also included.

The choice of scale driver also has some influence non costs borne in rural areas. Such companies will have longer networks relative to their customer base than companies that operate in denser areas. As a result, length of mains as the scale driver of choice over properties will generate additional allowance for more rural companies. The longer travelling distances as an argument for increased costs in rural areas will be covered, at least partially, by length of mains as the choice of scale driver making rural companies appear comparatively larger than they would seem were properties to be used as the scale driver.

5.5.2 The need to moderate the outlier effects of TMS

In section 5.3 we set out analysis which clearly shows that TMS is a material outlier relative to the rest of the sector that is significantly skewing model coefficients. This is demonstrated by the sensitivity of the models to the inclusion/exclusion of TMS. We set out two potential remedies.

Option 1: Remove TMS from the TWD models or include a TMS dummy and retain squared density

Whilst the exclusion of the legitimate observations should be avoided where possible, the fit of models improves materially and conforms much more to the engineering expectations set out in section 5.1.

If TMS is excluded, we consider that the squared density term can be legitimately included. This is because the choice between linear and non-linear density becomes relatively immaterial in the MOSA models (i.e. rurality isn't really being described (no inflection within the distribution), and the model is no longer being fitted to the outlier observations). This is shown by the similarity between the green and blue lines in ***Figure 15***.

We note that the choice between linear and non-linear density remains more material in the population/(length or area) models. This is shown in the difference between green and blue lines in ***Figure 16***. This reinforces our point that these density drivers better account for rurality and supports the suggestion that a population/area specification should be added.

An alternative approach to removing TMS would be to add a TMS dummy. This would acknowledge the atypical circumstance of TMS and not force the models to fit to it. However, at face value this would then be assuming that the TMS observations were efficient. Therefore, a clear approach of how to address the dummy coefficient would be required.

A CAC will be required for TMS if it is excluded from the models. The modelled efficiency range for TMS cannot be entirely attributed to inefficiency – there are likely to be legitimate company specific circumstances that should be allowed for (but outside of models).

Option 2: Retain TMS in TWD models but moderate adverse effects of outlier

If the TMS observations are to be retained, we consider that a further remedy to moderate the issue of overfitting of the squared term to outlier observations is needed. We suggest the following.

- Use linear density only For MOSA and LAD from MSOA specifications, or remove these models entirely (TWD models 1,2,4 and 5). We give the following reasonings.
  - Whist the inclusion of the squared term improves the statistical performance of the models, this is purely a function of the effect of TMS (as shown by the equivalent performance when TMS is removed).
  - As set out in section 5.2, there is not a strong case to retain the squared terms at account for rurality cost drivers because, the rurality cost drivers are poorly accounted for in MSOA/LAD from MSOA. This is shown by the lack of variance between the blue and green lines in ***Figure*** 15. We have also noted that rural companies are receiving benefit from the length scale driver (such companies will appear larger than would otherwise).
- Retain density squared for models use an overall level of density (population/length or population per area) i.e. as per Ofwat model TWD3 and 6. We give the following reasonings.
  - As per section 5.1, the engineering expectation is that both rural and urban areas drive costs. Therefore, we should expect an inflection in the TWD cost / density curve within the observed distribution.
  - There is a material difference in the shape of the curves between linear and non-linear once TMS is removed (difference between green and blue lines in ***Figure 16***). This suggests that (unlike in the weighted average specifications in ***Figure 15***) the squared term is capturing variance that would otherwise be omitted.
  - The inclusion/exclusion of TMS is much less sensitive to these models (i.e. TMS is less of an outlier if using this variable)

Figure 15: The relationship between unit TWD cost and LAD from MSOA density. The curve colours are as follows.

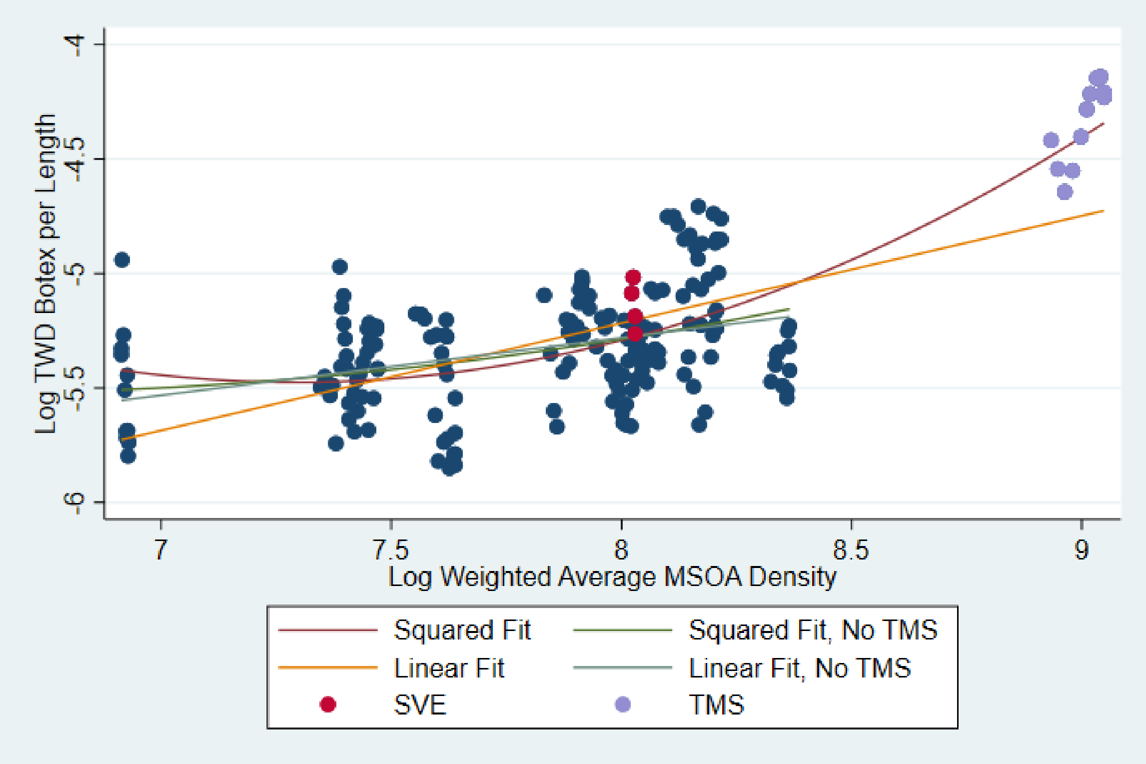
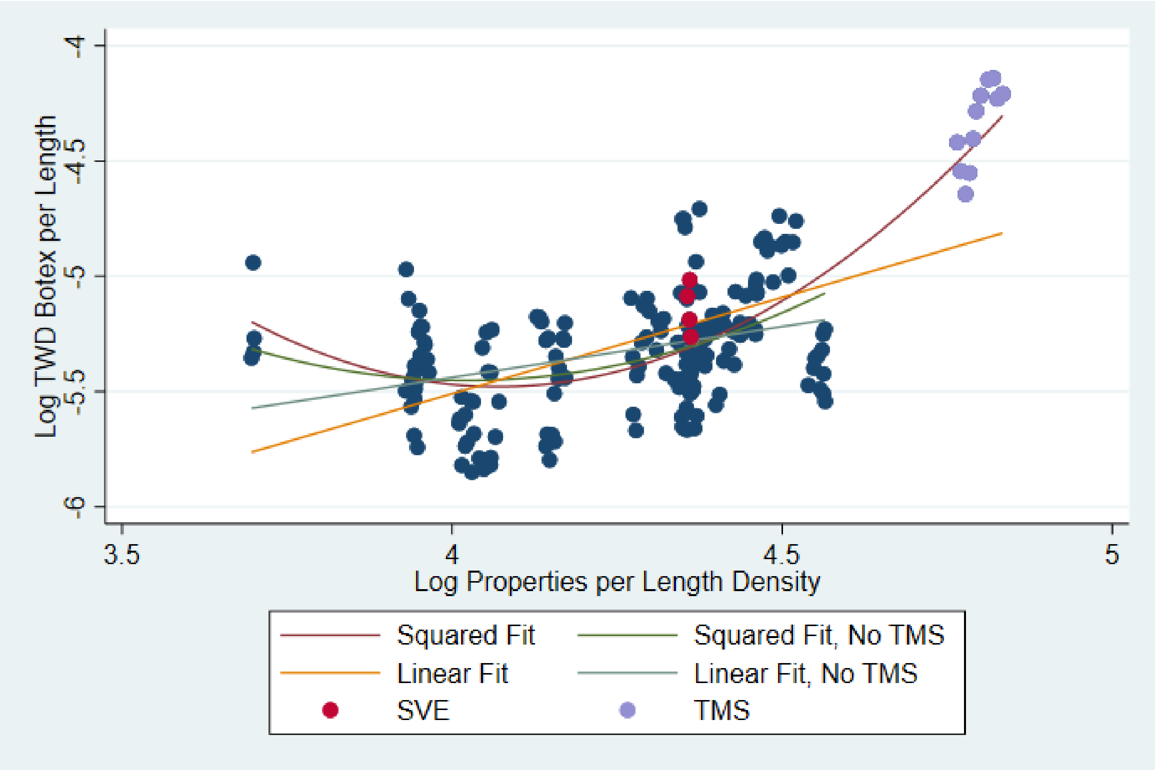


Figure 16: The relationship between unit TWD cost and LAD from Population/Length density. The curve colours are as follows.



Model regressions

Table 22, ***Table 23*** and ***Table 24*** show regressions without the inclusion of Thames, without the squared density terms, and without both Thames and the squared density terms respectively. ***Table 22*** shows that in the boosters per length specifications the density terms are not significant and the significance is reduced in TWD4. ***Table 23*** shows that the R-squared reduces by around 0.02-0.03 with the removal of the squared term, which is fairly significant. The RESET tests are also failed by the removal of those squared terms. However, ***Table 24*** shows that this is not the case when Thames is removed. The R-squared is reduced by at most 0.009, and the RESET test is satisfied when both the squared term and Thames are removed from the models, with the squared properties per length seemingly providing the most additional information to the model which is consistent with the engineering logic covered above.

Taken together, these regressions suggest that we are facing problems with the Thames being a significant density outlier. The true data generating process may be squared in TWD but the weighted average density measures don't capture the relationship effectively. Thames is also making the relationship far more pronounced than it should be. To that end, it may be prudent to either remove the squared term from the models using weighted average density, or remove those models entirely.

'Overall' measures of density are more strongly backed by engineering logic, giving a more complete picture of a company's density profile, rather than just being a measure of their most dense areas (see section 5.2). Therefore, including Population per Area gives an additional point of triangulation of an overall view of density which is more strongly supported by engineering logic than the weighted average density measures.

Table 22: TWD PR24 specifications without Thames included. p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	TWD1	TWD2	TWD3	TWD4	TWD5	TWD6
Log lengths of main	1.030*** (0.000)	0.987*** (0.000)	1.058*** (0.000)	1.073*** (0.000)	1.071*** (0.000)	1.051*** (0.000)
Log boosters per length	0.482*** (0.001)	0.469*** (0.000)	0.516*** (0.002)			
Log APH TWD				0.331*** (0.000)	0.496*** (0.000)	0.349*** (0.000)
Log LAD from MSOA	-1.311 (0.485)			-3.301* (0.095)		
Log LAD from MSOA sq	0.114 (0.409)			0.253* (0.083)		
Log MSOA		-0.452 (0.914)			-13.372*** (0.002)	
Log MSOA sq		0.06 (0.826)			0.893*** (0.002)	
Log Properties/Length			-9.991 (0.145)			-17.962*** (0.000)
Log Properties/Length sq			1.307 (0.112)			2.218*** (0.000)
Constant	-0.113 (0.984)	-3.406 (0.827)	15.065 (0.273)	3.068 (0.607)	41.682*** (0.008)	28.851*** (0.001)
Observations	176	176	176	176	176	176
R-squared Adjusted	0.949	0.948	0.952	0.953	0.961	0.959
RESET	0.397	0.321	0.536	0.473	0.531	0.948
AIC	-126.619	-127.838	-134.145	-131.531	-137.728	-139.949
Efficiency Range	0.899	1.186	0.608	0.584	0.876	0.535

*Table 23: TWD PR24 specifications without squared density terms. p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	TWD1	TWD2	TWD3	TWD4	TWD5	TWD6
Log lengths of main	0.995*** (0.000)	0.978*** (0.000)	1.054*** (0.000)	0.983*** (0.000)	0.961*** (0.000)	1.012*** (0.000)
Log boosters per length	0.561*** (0.001)	0.555*** (0.000)	0.741*** (0.000)			
Log APH TWD			0.348*** (0.004)		0.379*** (0.001)	0.378*** (0.001)
Log LAD from MSOA	0.370*** (0.000)			0.240*** (0.006)		
Log MSOA		0.712*** (0.000)			0.520*** (0.000)	
Log Properties/Length			1.478*** (0.000)			0.919*** (0.006)
Constant	-5.513*** (0.000)	-8.337*** (0.000)	-9.008*** (0.000)	-8.322*** (0.000)	-10.634*** (0.000)	-10.957*** (0.000)
Observations	187	187	187	187	187	187
R-squared Adjusted	0.928	0.939	0.938	0.931	0.946	0.942
RESET	0.049	0.019	0.587	0.095	0.131	0.314
AIC	-111.095	-119.367	-119.623	-112.355	-121.325	-118.12
Efficiency Range	1.321	0.956	0.99	1.3	0.909	0.981

*Table 24: TWD PR24 specifications without TMS or squared density terms. p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	TWD1	TWD2	TWD3	TWD4	TWD5	TWD6
Log lengths of main	0.986*** (0.000)	0.980*** (0.000)	1.029*** (0.000)	0.975*** (0.000)	0.965*** (0.000)	0.995*** (0.000)
Log boosters per length	0.516*** (0.000)	0.477*** (0.000)	0.618*** (0.000)			
Log APH TWD				0.272*** (0.000)	0.309*** (0.000)	0.309*** (0.000)
Log LAD from MSOA	0.246*** (0.000)			0.131*** (0.001)		
Log MSOA		0.477*** (0.000)			0.317*** (0.000)	
Log Properties/Length			0.993*** (0.000)			0.532*** (0.000)
Constant	-4.775*** (0.000)	-6.863*** (0.000)	-7.248*** (0.000)	-7.189*** (0.000)	-8.810*** (0.000)	-8.875*** (0.000)
Observations	176	176	176	176	176	176
R-squared Adjusted	0.949	0.948	0.951	0.947	0.952	0.951
RESET	0.467	0.323	0.515	0.816	0.802	0.837
AIC	-111.095	-119.367	-119.623	-112.355	-121.325	-118.12
Efficiency Range	1.854	1.418	1.438	1.799	1.418	1.531

6. P Consents

Wastewater treatment complexity is currently only accounted for through the percentage of load treated with Ammonia (NH_3) consents of less than 3mg/l. Phosphorus consents are set to become increasingly tight over the remainder of AMP7 and throughout AMP8.

Phosphorus treatment costs accelerate when consents move below 1mg/l. Consents of less than ~0.75mg/l cannot be achieved using conventional biological processes and require significant additional chemical dosing and mixing. At tighter consent levels, additional processes must be installed, as has been the case at Finham where we have had to install a bespoke COMAG tertiary solids removal plant in order to satisfy a 0.22mg/l P consent. Over the 4 years we have operated the plant, we incurred an additional 1.3MwH per day of energy and more than used 3,000 tonnes per year of Ferric, Polymer and Magnetite chemicals. Together these have led to an additional annual operating cost at Finham alone of more than £0.8m.

We welcome Ofwat's desire to account for increased costs as a result of tightening P-consents, and accept that fixing any forecasts at 24-25 levels seems sensible given the acceleration of tightening P-consents and the accounting for new P-consents over the course of AMP8 in the WINEP lines. Our view is that the models can accommodate a variable to account for the percentage of load treated with phosphorus consents of less than 0.5mg/l. We will also be submitting a cost adjustment claim to this effect to allow for a scenario in which Phosphorus consents are not considered in the base cost models given the materiality of the driver for us in AMP8.

We propose that either:

- $P \leq 0.5mg/l$ is included in the sewage treatment models as an additional variable;
- it is triangulated with the current $NH3 \leq 3mg/l$ models; or
- it is included as part of a composite variable.

In January, we proposed a simple average of $P \leq 0.5mg/l$ and $NH3 \leq 3mg/l$. $P \leq 0.5mg/l$ as a standalone variable is insignificant, although given the limited data currently present in the historical dataset, an additional year of data could change this.

A composite variable defined as an average of the NH3 and P variables is significant. Logically, this assumes the Botex response to a unit (1 percentage point) increase in either variable is the same. We can write the composite variable with its coefficient as:

$$\beta Composite \equiv \beta \frac{\%NH3 \leq 3mg/l + \%P \leq 0.5mg/l}{2} \equiv 0.5\beta\%NH3 \leq 3mg/l + 0.5\beta\%P \leq 0.5mg/l$$

This is likely to not be the case – our experience is that treating to $P \leq 0.5mg/l$ is likely to be significantly more costly – but given the lack of disaggregated cost data with which to calculate appropriate weightings, this seems to be a pragmatic solution if we are to model Phosphorous consents with the data currently available.

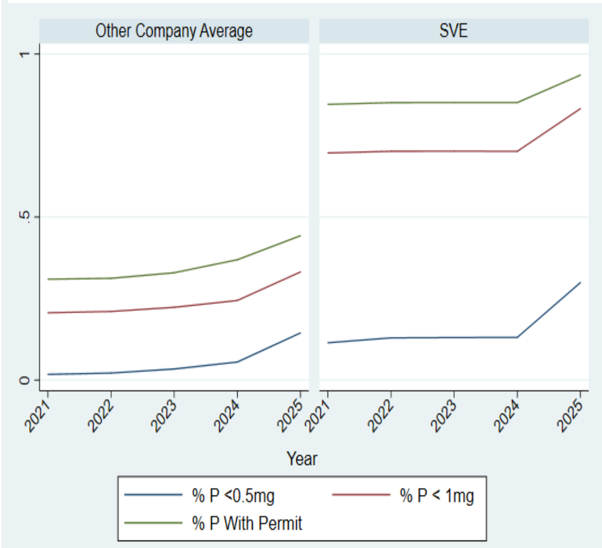
We should also note that tight phosphorous consents will have a negative impact on Bioresources revenues, which is also not accounted for, because of the lower quality of sludge that is produced when operating to such tight consents.

6.1 Tightening P-consents

P-consents are set to tighten substantially over the course of the current AMP and into the next AMP. This will be particularly felt by companies with large inland populations given P is a consent associated with the health of lower flowing rivers. Severn Trent is disproportionately impacted by these changes as a result of discharges into small rivers, as shown in ***Figure 17***, which takes its values from the PR19 business plans. This shows that consents are rapidly tightening towards the end of AMP7, a trend which will continue into AMP8.

This is set to become a significant cost pressure, both in terms of the costs of implementation, which is covered by enhancement expenditure, and in terms of ongoing opex and capex cost pressures as a result of projects completed in AMP7 which are considered base but not currently accounted for.

Figure 17: Tightening of P-consents over AMP7. Data from PR19 business plan tables.



6.2 Suitability for regression modelling

We have tested $P \leq 0.5mg/l$ in both SWT and WWNP models and found that it is marginally insignificant in WWNP models and insignificant in SWT models. When considered as part of composite wastewater treatment complexity variable, however, it is highly significant.

Table 25 shows SWT models with the addition of $P \leq 0.5mg/l$, and replacing $NH3 \leq 3mg/l$ with a composite variable. We can see that the R-squared has marginally improved, although there are concerns with RESET test failures. The magnitude of the coefficient in SWT4-6 has doubled relative to the % load with NH3 below 3 mg/l. This is expected given the division by 2, but will also be halving the effect of a unit increase in % load with P below 0.5mg/l relative to the models where it is included as a standalone variable. We expect that there will not be much change in the coefficient relative to that on % load with NH3 below 3mg given the ratio of load treated to each consent.

Table 26 and ***Table 27*** show WWNP models with the addition of $P \leq 0.5mg/l$, and replacing $NH3 \leq 3mg/l$ with a composite variable respectively. In these models, nothing really changes in terms of R-squared or the RESET test, again because of the ratios in the proportions of load treated to each consent.

These appear to be the two most pragmatic ways to account for P-consents at PR24. With additional years of data in which P consents are increasing and becoming more ubiquitous across the sector, we do expect that these results will improve.

*Table 25: SWT PR24 specifications with the addition of percentage of load treated with P consents below 0.5mg/l (SWT1-3) and with a wastewater treatment complexity composite variable (SWT4-6) with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	SWT1	SWT2	SWT3	SWT4	SWT5	SWT6
Log Load	0.636*** (0.000)	0.727*** (0.000)	0.739*** (0.000)	0.637*** (0.000)	0.717*** (0.000)	0.782*** (0.000)
% load with NH3 below 3mg/l	0.006*** (0.000)	0.006*** (0.000)	0.006*** (0.000)			
% load with P below 0.5mg/l	0.011 (0.329)	0.012 (0.304)	0.012 (0.284)			
Average % load treated with NH3 below 3mg/l or P below 0.5mg/l				0.012*** (0.000)	0.012*** (0.000)	0.013*** (0.000)
% load treated in size bands 1-3	0.027 (0.280)			0.028 (0.246)		
% load treated at works with PE over 100k		-0.009*** (0.001)			-0.008*** (0.004)	
Log weighted average treatment size			-0.249*** (0.000)			-0.245*** (0.000)
Constant	-3.513*** (0.006)	-4.103*** (0.000)	-2.999*** (0.000)	-3.537*** (0.008)	-3.992*** (0.000)	-2.908*** (0.000)
Observations	110	110	110	110	110	110
R-squared Adjusted	0.856	0.869	0.912	0.855	0.868	0.912
RESET	0.021	0.058	0.52	0.024	0.147	0.726
AIC	-105.093	-107.141	-117.001	-106.524	-108.43	-118.341
Efficiency Range	0.619	0.518	0.336	0.654	0.519	0.314

Table 26: NPWW PR24 specifications with the addition of percentage of load treated with P consents below 0.5mg/l with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.

Variable/Metric	WWNP1	WWNP2	WWNP3	WWNP4	WWNP5	WWNP6	WWNP7	WWNP8
Log Load	0.643*** (0.000)	0.722*** (0.000)	0.690*** (0.000)	0.716*** (0.000)	0.647*** (0.000)	0.727*** (0.000)	0.711*** (0.000)	0.724*** (0.000)
% load with NH3 below 3 mg/l	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)	0.005*** (0.000)
% load with P below 0.5mg/l	0.010 (0.137)	0.010 (0.140)	0.010 (0.136)	0.010 (0.139)	0.009 (0.161)	0.009 (0.171)	0.009 (0.177)	0.009 (0.191)
% load treated in size bands 1-3		0.023* (0.096)				0.023* (0.054)		
% load treated at works with PE over 100k			-0.002 (0.125)				-0.003* (0.079)	
Log weighted average treatment size				-0.098*** (0.010)				-0.098*** (0.002)
Log pumping capacity per sewer length	0.349*** (0.000)	0.361*** (0.000)	0.341*** (0.001)	0.274*** (0.005)	0.340*** (0.000)	0.353*** (0.000)	0.333*** (0.000)	0.261*** (0.000)
Log urban rainfall per length MSOA					0.072** (0.025)	0.073** (0.017)	0.080** (0.017)	0.087** (0.016)
Constant	-2.928*** (0.000)	-4.028*** (0.000)	-3.392*** (0.000)	-2.892*** (0.000)	-2.768*** (0.000)	-3.882*** (0.000)	-3.388*** (0.000)	-2.728*** (0.000)
Observations	110	110	110	110	110	110	110	110
R-squared Adjusted	0.946	0.952	0.948	0.955	0.952	0.958	0.956	0.963
RESET	0.113	0.001	0.158	0.611	0.004	0.001	0.005	0.182
AIC	-193.401	-194.324	-192.243	-196.505	-195.418	-197.34	-195.25	-201.318
Efficiency Range	0.166	0.166	0.171	0.174	0.188	0.127	0.151	0.13

*Table 27: NPWW PR24 specifications a wastewater treatment complexity composite variable with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level.*

Variable/Metric	WWNP1	WWNP2	WWNP3	WWNP4	WWNP5	WWNP6	WWNP7	WWNP8
Log Load	0.636*** (0.000)	0.715*** (0.000)	0.679*** (0.000)	0.708*** (0.000)	0.641*** (0.000)	0.722*** (0.000)	0.699*** (0.000)	0.718*** (0.000)
Average % load treated with NH3 below 3mg/l or P below 0.5mg/l	0.010*** (0.000)	0.010*** (0.000)	0.011*** (0.000)	0.011*** (0.000)	0.011*** (0.000)	0.010*** (0.000)	0.011*** (0.000)	0.011*** (0.000)
% load treated in size bands 1-3		0.023* (0.089)				0.023** (0.046)		
% load treated at works with PE over 100k			-0.002 (0.165)				-0.003* (0.088)	
Log weighted average treatment size				-0.097*** (0.010)				-0.099*** (0.002)
Log pumping capacity per sewer length	0.355*** (0.000)	0.367*** (0.000)	0.345*** (0.001)	0.280*** (0.004)	0.345*** (0.000)	0.359*** (0.000)	0.336*** (0.000)	0.266*** (0.000)
Log urban rainfall per length MSOA					0.074** (0.021)	0.076** (0.013)	0.078** (0.016)	0.085** (0.015)
Constant	-2.852*** (0.000)	-3.950*** (0.000)	-3.269*** (0.000)	-2.804*** (0.000)	-2.701*** (0.000)	-3.812*** (0.000)	-3.258*** (0.000)	-2.649*** (0.000)
Observations	110	110	110	110	110	110	110	110
R-squared Adjusted	0.947	0.952	0.948	0.955	0.952	0.958	0.956	0.963
RESET	0.423	0.33	0.522	0.734	0.104	0.107	0.006	0.186
AIC	-194.533	-195.461	-193.356	-197.676	-196.72	-198.702	-196.625	-202.904
Efficiency Range	0.159	0.172	0.164	0.157	0.18	0.136	0.146	0.111

7. Proposed PR24 modelling suite

Throughout our response, we have highlighted several concerns with the models that we have sought to address on an individual basis. In this chapter, we combine the remedies we have proposed throughout this response, and show what we believe to be the optimum set of models using Ofwat's proposed PR24 suite as a starting point.

In the table below we set out how we have reviewed the coverage and statistical performance of the PR24 consultation models after inclusion of the remedies that we have described in this document.

In each of the models we show that the coverage of engineering expectations improves relative to both the PR19 models and the proposed PR24 models that were set out in the consultation. In the following sections we also describe how the statistical performance of the models has changed.

Attribute / Test	Basis for test	Blue	Green (Light Green)	Amber	Red
Engineering expectations met	We outlined in January our engineering expectations for each set of models. We stand by the statements we made.	More engineering expectations met than the corresponding PR24 model(s) and PR19 model	More engineering expectations met than the corresponding PR24 model(s)	N/A	N/A
Statistical fit (R2)	R-squared describes the proportion of variance described by the model. We also reviewed the R-squared of unit cost configuration given that scale drivers will dominate the explanatory power of total models, which can give false confidence in how well they explain complexity differences between companies. We have assumed an increase of 0.01 or more to be 'better' here.	Better than both the corresponding PR24 model(s) and PR19 model	Better than the corresponding PR24 model(s)	No change from the corresponding PR24 model(s)	Worse than the corresponding PR24 model(s)
Functional form (RESET)	RESET test considers the need for higher order terms.	N/A	N/A	N/A	Failed RESET test
Predictive Power (AIC)	AIC describes that benefit of adding extra parameters while guarding against over-fitting. We have assumed a decrease of 4 or more to be 'better' here.	Better than both the corresponding PR24 model(s) and PR19 model	Better than the corresponding PR24 model(s)	No change from the corresponding PR24 model(s)	Worse than the corresponding PR24 model(s)
Efficiency range	Efficiency ranges need to be plausible. Poorly specified models will generate implausibly large ranges. A reduction of more than 0.05 is viewed as large here. Less than 5% is considered immaterial.[31,32]	Large reduction relative to the corresponding PR24 model(s) and PR19 model	Large reduction relative to the corresponding PR24 model(s)	No change from the corresponding PR24 model(s)	Large increase relative to the corresponding PR24 model(s)
Number of parameters	To deliver 'sensibly simple' models, we are trying to deliver that maximum number of engineering expectations using the minimum number of parameters	N/A	N/A	N/A	N/A

[31] This also needs to account for companies' changing efficiency performance, and additional engineering fit may reveal some inefficiencies. Our overarching view is that if modelled efficiency can be changed by an additional variable supported by engineering logic then it is a valid change. We are concerned about Ofwat's presuppositions on the nature of inefficiency.

[32] Econometric base cost models for PR24, Ofwat, April 2023, available at: https://www.ofwat.gov.uk/wp-content/uploads/2023/04/Econometric_base_cost_models_for_PR24_final.pdf, p.33.

7.1 Water resources plus models with proposed remedies applied

Table 28 and ***Table 29*** show Ofwat's PR24 WRP consultation modelling suite augmented with the changes we have proposed as part of this consultation. This involves:

- the removal of the squared density terms,
- the addition of WRP APH (***Table 28***) or WT APH (***Table 29***),
- the replacement of weighted average complexity with an alternative that provides better statistical fit, while still accepting that costs increase with more complex treatment processes; and
- the addition of Population per Area density as an additional point of triangulation.

While we consider the models in ***Table 28*** to be the most clearly backed by engineering logic, we note that there are a number of insignificant variables, there is no improvement in R-squared. However, the maximum efficiency range is reduced by 0.228 which constricts the implausible efficiency ranges in the current PR24 WRP suite. There are still concerns with WSX being an outlier that unduly influences the coefficients in these models too.

The models in ***Table 29*** offer significant improvements with R-squared, satisfy the RESET tests, include variables that are significant, are backed by engineering logic, and reduce the maximum efficiency range by 0.43 which is substantial.

*Table 28: A proposed set of WRP models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's corresponding proposed modes are shown in square brackets below the test values for our proposed models where applicable.*

Variable /Metric	PR19 WRP1	PR19 WRP2	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6	WRP7	WRP8
Log properties	1.074*** (0.000)	1.069*** (0.000)	1.057*** (0.000)	1.053*** (0.000)	1.027*** (0.000)	1.025*** (0.000)	1.034*** (0.000)	1.031*** (0.000)	1.022*** (0.000)	1.017*** (0.000)
% water treated in bands 3-6	0.006*** (0.000)		0.004** (0.013)		0.004** (0.023)		0.004** (0.012)		0.004*** (0.008)	
Log weighted average complexity		0.377 (0.123)								
YKY weighted average complexity				0.282* (0.093)		0.263 (0.134)		0.289* (0.083)		0.290* (0.083)
Log weighted average LAD density	-1.614*** (0.000)	-1.412*** (0.005)								
Squared	0.101*** (0.000)	0.087*** (0.009)								
Log LAD from MSOA			-0.166* (0.099)	-0.168* (0.093)						
Log MSOA					-0.134 (0.466)	-0.151 (0.395)				
Log Properties /Length							-0.478 (0.141)	-0.503 (0.106)		
Log Population /Area									-0.190* (0.066)	-0.194* (0.058)
Log WRP APH (2 EEs)			0.116 (0.340)	0.118 (0.265)	0.140 (0.272)	0.139 (0.217)	0.138 (0.236)	0.137 (0.174)	0.128 (0.232)	0.129 (0.166)
Constant	-5.093*** (0.000)	-5.805*** (0.000)	-10.312*** (0.000)	-10.295*** (0.000)	-10.087*** (0.000)	-9.974*** (0.000)	-9.218*** (0.000)	-9.118*** (0.000)	-9.936*** (0.000)	-9.905*** (0.000)
Observations	187	187	187	187	187	187	187	187	187	187
EEs met (of 8)	1	1	4 [1]	4 (+1) [1]	4 [1]	4 (+1) [1]	4 [1]	4 (+1) [1]	4	4 (+1)
R-squared Adjusted	0.917	0.907	0.908 [0.909]	0.907 [0.902]	0.899 [0.901]	0.899 [0.896]	0.910 [0.910]	0.911 [0.905]	0.912	0.912
R-squared Unit	0.389	0.305	0.309 [0.326]	0.306 [0.267]	0.252 [0.260]	0.258 [0.219]	0.337 [0.354]	0.343 [0.305]	0.351	0.351
RESET	0.295	0.228	0.806 [0.436]	0.684 [0.367]	0.764 [0.765]	0.661 [0.729]	0.751 [0.324]	0.664 [0.203]	0.592	0.425
AIC	-23	-20	-20 [-21]	-20 [-19]	-17 [-18]	-17 [-17]	-20 [-20]	-20 [-18]	-21	-21
Efficiency Range	1.49	1.48	1.28 [1.49]	1.27 [1.48]	1.29 [1.50]	1.27 [1.51]	1.25 [1.47]	1.25 [1.46]	1.12	1.09
Change in EE from PR19	-	-	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3	3 (+1)
No. of parameters (inc. constant)	5	5	5 [5]	5 [5]	5 [5]	5 [5]	5 [5]	5 [5]	5	5

Table 29: A proposed set of WRP models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's corresponding proposed models are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 WRP1	PR19 WRP2	WRP1	WRP2	WRP3	WRP4	WRP5	WRP6	WRP7	WRP8
Log properties	1.074*** (0.000)	1.069*** (0.000)	1.044*** (0.000)	1.039*** (0.000)	1.019*** (0.000)	1.017*** (0.000)	1.018*** (0.000)	1.014*** (0.000)	1.005*** (0.000)	1.000*** (0.000)
% water treated in bands 3-6	0.006*** (0.000)		0.004** (0.017)		0.004** (0.033)		0.004** (0.017)		0.005*** (0.008)	
Log weighted average complexity		0.377 (0.123)								
YKY weighted average complexity				0.290 (0.108)		0.280 (0.146)		0.303* (0.093)		0.302 (0.102)
Log weighted average LAD density	-1.614*** (0.000)	-1.412*** (0.005)								
Squared	0.101*** (0.000)	0.087*** (0.009)								
Log LAD from MSOA			-0.166** (0.018)	-0.168** (0.017)						
Log MSOA					-0.175 (0.162)	-0.190 (0.114)				
Log Properties /Length							-0.490** (0.032)	-0.510** (0.019)		
Log Population /Area									-0.182** (0.019)	-0.186** (0.015)
Log WT APH			0.125** (0.033)	0.127** (0.039)	0.126* (0.060)	0.128* (0.068)	0.127** (0.031)	0.128** (0.036)	0.117** (0.041)	0.118* (0.055)
Constant	-5.093*** (0.000)	-5.805*** (0.000)	-9.933*** (0.000)	-9.913*** (0.000)	-9.384*** (0.000)	-9.299*** (0.000)	-8.669*** (0.000)	-8.597*** (0.000)	-9.487*** (0.000)	-9.458*** (0.000)
Observations	187	187	187	187	187	187	187	187	187	187
EEs met (of 8)	1	1	3 [1]	3 (+1) [1]	3 [1]	3 (+1) [1]	3 [1]	3 (+1) [1]	3	3 (+1)
R-squared Adjusted	0.917	0.907	0.927 [0.909]	0.927 [0.902]	0.920 [0.901]	0.920 [0.896]	0.929 [0.910]	0.929 [0.905]	0.928	0.928
R-squared Unit	0.389	0.305	0.466 [0.326]	0.464 [0.267]	0.425 [0.260]	0.432 [0.219]	0.485 [0.354]	0.489 [0.305]	0.479	0.481
RESET	0.295	0.228	0.806 [0.436]	0.684 [0.367]	0.764 [0.765]	0.661 [0.729]	0.751 [0.324]	0.664 [0.203]	0.654	0.437
AIC	-23	-20	-25 [-21]	-24 [-19]	-21 [-18]	-20 [-17]	-24 [-20]	-24 [-18]	-24	-24
Efficiency Range	1.49	1.48	1.04 [1.49]	1.03 [1.48]	1.09 [1.50]	1.08 [1.51]	1.03 [1.47]	1.03 [1.46]	0.96	0.92
Change in EE from PR19	-	-	2 [1]	2 (+1) [1]	2 [1]	2 (+1) [1]	2 [1]	2 (+1) [1]	2	2 (+1)
No. of parameters (inc. constant)	5	5	5	5	5	5	5	5	5	5

7.2 Treated Water Distribution models with proposed remedies applied

Table 30 and ***Table 31*** shows Ofwat's PR24 TWD consultation modelling suite augmented with the changes we have proposed as part of this consultation. This involves:

- allowing boosters per length and TWD APH to feature in the same model,
- removing Thames' observations from the model (***Table 30***) or imposing a linear relationship between density and cost for Weighted Average Density models (***Table 31***); and
- adding Population per Area density as an additional point of triangulation.

While these models show reduced R-squared values against Ofwat's proposed models in some cases, we see some improvements in the Unit R-squared and AIC, and they eradicate the concerns with Thames as an outlier fitting the squared density term where Weighted Average Density drivers are used. The combination of network complexity variables is also backed by engineering logic. There is a RESET test failure in ***Table 31***, but as discussed, the RESET test can be sensitive to outliers.

We have shown efficiency ranges with and without TMS included. A CAC will be required for TMS if either of these options are considered, and the modelled efficiency range for TMS cannot be entirely attributed to inefficiency here – there are likely to be legitimate company specific circumstances that should be allowed for (but outside of models).

Table 30: A proposed set of TWD models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 TWD1	TWD1	TWD2	TWD3	TWD4
Log lengths of main	1.077*** (0.000)	1.049*** (0.000)	1.044*** (0.000)	1.053*** (0.000)	1.089***
Log boosters per length	0.437*** (0.002)	0.366*** (0.001)	0.255** (0.015)	0.406*** (0.002)	0.557*** (0.000)
Log APH TWD		0.272*** (0.000)	0.418*** (0.000)	0.322*** (0.000)	0.202*** (0.000)
Log LAD	-2.946*** (0.000)				
Squared	0.235*** (0.000)				
Log LAD from MSOA		-2.214 (0.210)			
Squared		0.180 (0.170)			
Log MSOA			-9.244** (0.034)		
Squared			0.630** (0.027)		
Log Properties/Length				-11.808*** (0.000)	
Squared				1.520*** (0.000)	
Log Population/Area					-1.541*** (0.005)
Squared					0.166*** (0.000)
Constant	4.722*** (0.002)	1.134 (0.828)	27.174* (0.086)	17.113*** (0.006)	-1.467 (0.265)
Observations	187	176	176	176	176
EEs met (of 6)	2	4 [2]	4 [2]	4 [3]	4 [3]
R-squared Adjusted	0.957	0.959 [0.961]	0.963 [0.965]	0.966 [0.966]	0.967
R-squared Unit	0.620	0.418 [0.651]	0.477 [0.687]	0.520 [0.697]	0.538
RESET	0.101	0.286 [0.460]	0.481 [0.721]	0.441 [.844]	0.684
AIC	-128	-136 [-132]	-142 [-138]	-148 [-140]	-142
Efficiency Range (No TMS)	0.61	0.64 (0.64) [0.59]	0.68 (0.58) [0.61]	0.56 (0.56) 0.54	0.91 (0.51)
Change in EE from PR19	-	2	2	2	2
No. of parameters (inc. constant)	-	2	2	2	2

Table 31: A proposed set of TWD models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 TWD	TWD1	TWD2	TWD3	TWD4
Log lengths of main	1.077*** (0.000)	0.989*** (0.000)	0.969*** (0.000)	1.060*** (0.000)	1.132*** (0.000)
Log boosters per length	0.437*** (0.002)	0.452*** (0.006)	0.455*** (0.002)	0.387*** (0.002)	0.525*** (0.000)
Log APH TWD		0.297*** (0.000)	0.334*** (0.002)	0.339*** (0.000)	0.222*** (0.002)
Log LAD	-2.946*** (0.000)				
Squared	0.235*** (0.000)				
Log LAD from MSOA		0.349*** (0.000)			
Log MSOA			0.691*** (0.000)		
Log Properties/Length				-14.232*** (0.000)	
Squared				1.811*** (0.000)	
Log Population/Area					-2.968*** (0.000)
Squared					0.294*** (0.000)
Constant	4.722*** (0.002)	-7.046*** (0.000)	-9.956 (0.000)	21.933*** (0.000)	1.822 (0.268)
Observations	187	187	187	187	187
EEs met (of 6)	2	4 [2]	4 [2]	4 [3]	4 [3]
R-squared Adjusted	0.957	0.936 [0.961]	0.951 [0.965]	0.971 [0.966]	0.967
R-squared Unit	0.620	0.433 [0.651]	0.565 [0.687]	0.745 [0.697]	0.706
RESET	0.101	0.020 [0.460]	0.086 [0.721]	0.851 [0.844]	0.102
AIC	-128	-115 [-132]	-126 [-138]	-148 [-140]	-142
Efficiency Range (No TMS)	0.61	1.29 (0.81) [0.59]	0.85 (0.76) [0.61]	0.56 (0.56) 0.54	0.49 (0.49)
Change in EE from PR19	-	2	2	2	2
No. of parameters (inc. constant)	-	5	5	6	6

7.3 Wholesale Water models with proposed remedies applied

Table 32 and ***Table 33*** shows Ofwat's PR24 WW consultation modelling suite augmented with the changes we have proposed as part of this consultation. This involves:

- allowing boosters per length and WW APH (***Table 32***) or separated APH values (***Table 33***) to feature in the same model;
- the replacement of weighted average complexity with an alternative that provides better statistical fit, while still accepting that costs increase with more complex treatment processes; and
- the addition of Population per Area density as an additional point of triangulation.

While the models in both these tables offer only modest R-squared improvements, they show significant improvements in Unit R-squared and AIC, reduce the modelled efficiency ranges, and in all cases satisfy the RESET test. The efficiency score ranges are also substantially reduced, our proposed variables are significant throughout, and the engineering logic behind these models have been strengthened.

Table 32: A proposed set of WW models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 WW1	PR19 WW2	WW1	WW2	WW3	WW4	WW5	WW6	WW7	WW8
Log properties	1.071*** (0.000)	1.059*** (0.000)	1.091*** (0.000)	1.079*** (0.000)	1.067*** (0.000)	1.059*** (0.000)	1.072*** (0.000)	1.064*** (0.000)	1.099*** (0.000)	1.089*** (0.000)
% water treated in bands 3-6	0.004*** (0.000)		0.002* (0.091)		0.002 (0.148)		0.002 (0.118)		0.001 (0.264)	
Log weighted average complexity		0.430*** (0.001)								
YKY weighted average complexity				0.206* (0.056)		0.178* (0.090)		0.197* (0.063)		0.159 (0.151)
Log boosters per length	0.335** (0.032)	0.334 (0.018)	0.452*** (0.004)	0.449*** (0.002)	0.504*** (0.001)	0.496*** (0.000)	0.303** (0.013)	00.301** (0.014)	0.383*** (0.009)	0.380*** (0.007)
Log APH WW (3 EEs)			0.319** (0.010)	0.286** (0.014)	0.348*** (0.005)	0.319*** (0.006)	0.365*** (0.003)	0.337*** (0.004)	0.390*** (0.002)	0.359*** (0.002)
Log LAD	-2.094*** (0.000)	-1.832*** (0.000)								
Log LAD sq	0.147*** (0.000)	0.128*** (0.000)								
Log LAD from MSOA			-1.898*** (0.000)	-1.742*** (0.000)						
Log LAD from MSOA sq			0.137*** (0.000)	0.126*** (0.000)						
Log MSOA					-4.670*** (0.000)	-4.350*** (0.001)				
Log MSOA sq					0.305*** (0.000)	0.284*** (0.000)				
Log Properties/Length							-13.740*** (0.000)	-12.954*** (0.000)		
Log Properties/Length sq							1.611*** (0.000)	1.516*** (0.000)		
Log Population/Area									-2.680*** (0.000)	-2.521*** (0.000)
Log Population/Area sq									0.231*** (0.000)	0.217*** (0.000)
Constant	-1.566* (0.074)	-2.590*** (0.001)	-3.647** (0.038)	-3.984** (0.015)	8.127 (0.114)	7.041 (0.161)	18.510*** (0.000)	17.015*** (0.000)	-3.130*** (0.008)	-3.404*** (0.003)
Observations	187	187	187	187	187	187	187	187	187	187
EEs met (of 12)	3	3	6	6 (+1)	6	6 (+1)	6	6 (+1)	6	6 (+1)
R-squared Adjusted	0.970	0.971	0.967 [0.965]	0.969 [0.967]	0.966 [0.963]	0.968 [0.965]	0.971 [0.966]	0.972 [0.968]	0.968	0.970
R-squared Unit	0.532	0.547	0.499 [0.457]	0.523 [0.486]	0.483 [0.401]	0.505 [0.458]	0.549 [0.470]	0.568 [0.501]	0.509	0.531
RESET	0.222	0.122	0.562 [0.838]	0.517 [0.821]	0.661 [0.898]	0.591 [0.937]	0.976 [0.781]	0.943 [0.611]	0.887	0.797
AIC	-161	-162	-163 [-159]	-165 [-161]	-164 [-156]	-166 [-158]	-171 [-160]	-173 [-162]	-167	-169
Efficiency Range	0.61	0.65	0.58 [0.61]	0.59 [0.74]	0.58 [0.72]	0.59 [0.73]	0.52 [0.70]	0.53 [0.70]	0.55	0.57
Change in EE from PR19	-	-	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]
No. of parameters (inc. constant)	-	-	7	7	7	7	7	7	7	7

Table 33: A proposed set of WW models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 WW1	PR19 WW2	WW1	WW2	WW3	WW4	WW5	WW6	WW7	WW8
Log properties	1.071*** (0.000)	1.059*** (0.000)	1.072*** (0.000)	1.058*** (0.000)	1.053*** (0.000)	1.044*** (0.000)	1.055*** (0.000)	1.046*** (0.000)	1.079*** (0.000)	1.068*** (0.000)
% water treated in bands 3-6	0.004*** (0.000)		0.002** (0.037)		0.002 (0.109)		0.002* (0.079)		0.002 (0.156)	
Log weighted average complexity		0.430*** (0.001)								
YKY weighted average complexity				0.238** (0.023)		0.195* (0.064)		0.200** (0.045)		0.174* (0.096)
Log boosters per length	0.335** (0.032)	0.334 (0.018)	0.341*** (0.007)	0.332*** (0.004)	0.408*** (0.001)	0.398*** (0.001)	0.274** (0.011)	0.268** (0.011)	0.373** (0.012)	0.361*** (0.009)
Log APH TWD			0.263*** (0.006)	0.248*** (0.008)	0.276*** (0.007)	0.265*** (0.008)	0.211** (0.036)	0.201** (0.038)	0.211** (0.043)	0.203** (0.045)
Log APH WT			0.044 (0.229)	0.044 (0.210)	0.044 (0.258)	0.042 (0.261)	0.057 (0.118)	0.054 (0.120)	0.05 (0.210)	0.046 (0.209)
Log APH WR + RWD			0.036 (0.413)	0.017 (0.701)	0.052 (0.173)	0.038 (0.307)	0.090** (0.010)	0.077** (0.019)	0.099*** (0.006)	0.084** (0.013)
Log LAD	-2.094*** (0.000)	-1.832*** (0.000)								
Log LAD sq	0.147*** (0.000)	0.128*** (0.000)								
Log LAD from MSOA			-1.856*** (0.000)	-1.653*** (0.000)						
Log LAD from MSOA sq			0.131*** (0.000)	0.117*** (0.000)						
Log MSOA					-4.860*** (0.000)	-4.472*** (0.000)				
Log MSOA sq					0.313*** (0.000)	0.287*** (0.000)				
Log Properties/Length							-12.788*** (0.000)	-11.920*** (0.000)		
Log Properties/Length sq							1.499*** (0.000)	1.394*** (0.000)		
Log Population/Area									-2.352*** (0.000)	-2.178*** (0.000)
Log Population/Area sq									0.203*** (0.000)	0.187*** (0.000)
Constant	-1.566* (0.074)	-2.590*** (0.001)	-3.688*** (0.003)	-4.215*** (0.000)	9.129*** (0.007)	7.734** (0.021)	17.025*** (0.000)	15.316*** (0.000)	-3.346*** (0.023)	-3.732** (0.011)
Observations	187	187	187	187	187	187	187	187	187	187
EEs met (of 14)	3	3	6	6 (+1)	6	6 (+1)	6	6 (+1)	6	6 (+1)
R-squared Adjusted	0.970	0.971	0.972 [0.965]	0.973 [0.967]	0.971 [0.963]	0.972 [0.965]	0.973 [0.966]	0.974 [0.968]	0.969	0.970
R-squared Unit	0.532	0.547	0.564 [0.457]	0.590 [0.486]	0.546 [0.401]	0.568 [0.458]	0.576 [0.470]	0.596 [0.501]	0.516	0.542
RESET	0.222	0.122	0.532 [0.838]	0.546 [0.821]	0.570 [0.898]	0.569 [0.937]	0.922 [0.781]	0.875 [0.611]	0.791	0.643
AIC	-161	-162	-164 [-159]	-167 [-161]	-165 [-156]	-167 [-158]	-168 [-160]	-170 [-162]	-162	-164
Efficiency Range	0.61	0.65	0.58 [0.61]	0.59 [0.74]	0.58 [0.72]	0.59 [0.73]	0.52 [0.70]	0.53 [0.70]	0.55	0.57
Change in EE from PR19	-	-	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]	3 [0]	3 (+1) [0]
No. of parameters (inc. constant)	6	6	9	9	9	9	9	9	9	9

7.4 Sewage Treatment models with proposed remedies applied

Table 34 shows Ofwat's PR24 SWT consultation modelling suite augmented with the changes we have proposed as part of this consultation. This involves:

- Replacing % load treated with NH3 consents below 3mg/l with a composite of it and % load treated with Phosphorous consents below 0.5mg/l.

These models do not show improvements in statistical terms, nor do they have any great bearing on the modelled efficiency ranges. However, they enable us to capture the cost pressures present because of tight P-consents without any model concerns.

Table 34: A proposed set of SWT models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	PR19 SWT1	PR19 SWT2	SWT1	SWT2	SWT3
Log Load	0.648*** (0.000)	0.643*** (0.000)	0.637*** (0.000)	0.717*** (0.000)	0.782*** (0.000)
% load with NH3 below 3mg/l	0.006*** (0.000)	0.006*** (0.000)			
Average % load treated with NH3 below 3mg/l or P below 0.5mg/l			0.012*** (0.000)	0.012*** (0.000)	0.013*** (0.000)
% load treated in size bands 1-3	0.028 (0.231)		0.028 (0.246)		
% load treated in size band 6+		-0.008 (0.305)			
% load treated at works with PE over 100k				-0.008*** (0.004)	
Log weighted average treatment size					-0.245*** (0.000)
Constant	-3.664*** (0.004)	-2.890*** (0.001)	-3.537*** (0.008)	-3.992*** (0.000)	-2.908*** (0.000)
Observations	110	110	110	110	110
EEs met (of 4)	1	1	2 [1]	2 [1]	3 [2]
R-squared Adjusted	0.854	0.856	0.855 [0.856]	0.868 [0.869]	0.912 [0.912]
R-squared Unit	0.362	0.367	0.366 [0.373]	0.422 [0.427]	0.612 [0.613]
RESET	0.058	0.163	0.024	0.147	0.726
AIC	-105	-105	-107 [-105]	-108 [-107]	-118 [-117]
Efficiency Range	0.68	0.65	0.65 [0.62]	0.51 [0.52]	0.31 [0.34]
Change in EEs from PR19	-	-	1 [0]	1 [0]	2 [1]
No. of parameters (inc. constant)	4	4	4	4	4

7.5 Network Plus Wastewater models with proposed remedies applied

Table 35 shows Ofwat's PR24 WWNP consultation modelling suite augmented with the changes we have proposed as part of this consultation. This involves:

- Replacing % load treated with NH3 consents below 3mg/l with a composite of it and % load treated with Phosphorous consents below 0.5mg/l.

These models do not show improvements in statistical terms, nor do they have any great bearing on the modelled efficiency ranges. However, they enable us to capture the cost pressures present because of tight P-consents without any model concerns. The RESET test fails only where it also fails in the original suite.

Table 35: A proposed set of WWNP models for use at PR24 with p-values in parentheses. *** significance at 1% level, ** significance at 5% level, * significance at 10% level. Our test values for Ofwat's best corresponding proposed model are shown in square brackets below the test values for our proposed models where applicable.

Variable/Metric	WWNP1	WWNP2	WWNP3	WWNP4	WWNP5	WWNP6	WWNP7	WWNP8
Log Load	0.636*** (0.000)	0.715*** (0.000)	0.679*** (0.000)	0.708*** (0.000)	0.641*** (0.000)	0.722*** (0.000)	0.699*** (0.000)	0.718*** (0.000)
Average % load treated with NH3 below 3mg/l or P below 0.5mg/l	0.010**** (0.000)	0.010*** (0.000)	0.011*** (0.000)	0.011*** (0.000)	0.011*** (0.000)	0.010*** (0.000)	0.011*** (0.000)	0.011*** (0.000)
Pumping capacity per length	0.355*** (0.000)	0.367*** (0.000)	0.345*** (0.001)	0.280*** (0.000)	0.345*** (0.000)	0.359*** (0.000)	0.336*** (0.000)	0.266*** (0.000)
% load treated in size bands 1-3		0.023* (0.089)				0.023** (0.046)		
% load treated at works with PE over 100k			-0.002 (0.165)				-0.003* (0.088)	
Log weighted average treatment size				-0.097*** (0.010)				-0.099*** (0.002)
Urban rainfall per length (MSOA)					0.074** (0.021)	0.076** (0.013)	0.078** (0.016)	0.085** (0.015)
Constant	-2.852*** (0.000)	-3.950*** (0.000)	-3.269*** (0.00)	-2.804*** (0.000)	- 2.701*** (0.000)	-3.812*** (0.000)	-3.258*** (0.000)	-2.649*** (0.000)
Observations	110	110	110	110	110	110	110	110
EEs met (of 7)	2 [1]	2 [1]	2 [1]	3 [2]	3 [2]	3 [2]	3 [2]	4 [3]
R-squared Adjusted	0.947 [0.947]	0.952 [0.952]	0.948 [0.949]	0.955 [0.956]	0.952 [0.953]	0.958 [0.959]	0.956 [0.956]	0.963 [0.964]
R-squared Unit	0.709 [0.710]	0.737 [0.738]	0.718 [0.720]	0.758 [0.758]	0.740 [0.741]	0.773 [0.774]	0.758 [0.761]	0.800 [0.801]
RESET	0.423 [0.572]	0.330 [0.481]	0.522 [0.700]	0.734 [0.901]	0.104 [0.241]	0.107 [0.109]	0.006 [0.009]	0.186 [0.248]
AIC	-195 [-192]	-195 [-193]	-193 [-190]	-198 [-195]	-197 [-194]	-199 [-196]	-197 [-194]	-203 [-200]
Efficiency Range	0.16 [0.15]	0.16 [0.17]	0.17 [0.15]	0.16 [0.14]	0.18 [0.17]	0.14 [0.14]	0.15 [0.13]	0.11 [0.09]
No. of parameters (inc. constant)	4 [4]	4 [4]	4 [4]	4 [4]	5 [5]	5 [5]	5 [5]	5 [5]

8. Appendix 1 Engineering expectations

8.1.1 Water

Table 36: Prior engineering expectations 'EEs' in water (WRP and TWD).

Area	WRP (Water resources and Treatment)	TWD (Water distribution)
Scale	**EE1.** Properties is an uncontentious explanatory variable	**EE1.** There is a logical case to use both length and properties as a scale driver: • Length – Traditional scale driver that describes the size of the asset base. • Properties – The number of physical connections to the network drive cost. Point of failure and additional complexity that drive cost despite potentially having limited impact on the network length. It is also a stronger proxy for volume.
Economies of Scale (population density)	**EE2.** Size of WTWs is the most appropriate explanatory variable, however the inability to differentiate between groundwater (GW) and surface water (SW) mean that population density accounting for GW% is pragmatic. **EE3.** No clear engineering basis for including a non-linear density term in WRP models (i.e. density^2) since opportunities for economies of scale do not diminish at higher densities.	**EE2.** There is a strong theoretical basis for population density as an explanatory variable. TWD costs are driven by density and rurality, therefore population / area is a more appropriate metric because local authority densities (LAD) are weighted by population which will skew to higher density areas, therefore putting less weight on costs incurred in rural areas that the squared term is intended to capture.
Complexity (driven by Geography / Geology)	**EE4.** Ofwat's treatment complexity bands do not appear to be good explanatory variables. **EE5.** Water treatment APH is the most appropriate proxy to differentiate between treatment complexity. **EE6.** Raw water APH will allow for differences in water resource opportunities (and make sure that all pumping from source to tap is accounted for).	Boosters/length and TWD APH provide coherent and complementary explanatory power relating to the complexity of providing potable water to the supply area: **EE3.** Booster/length: Increased pumping assets (driven by geography and population location) increase cost, irrespective of the pumping lift (capital maintenance). **EE4.** TWD APH: Increased pumping lift has a direct impact on network opex costs.
Weather effects	**EE7.** Weather is a proxy for peak demand & asset use intensity.	**EE5.** Weather is a proxy for peak demand & asset use intensity.
AMP effects	**EE8.** Capex costs vary materially according to the AMP cycle.	**EE6.** Capex costs vary materially according to the AMP cycle.

8.1.2 Waste

Table 37: Prior engineering expectations in wastewater (SWT and SWC).

Area	SWT (Waste Treatment)	SWC (Waste Network)
Scale	**EE1.** Load is an uncontentious explanatory variable.	**EE1.** Properties provides a way of moving away from PDAS data which is subject to much uncertainty.
Economies of Scale (population density)	**EE2.** We consider that the best variable is a weighted average scale based on industry-wide costs incurred at sewage works within each band.	**EE2.** There is a strong theoretical basis for population density as an explanatory variable. Population / length is a more appropriate metric because Local Authority (LAD) densities are weighted by population which will skew to higher density areas, therefore putting less weight on costs incurred in rural areas that the squared term is intended to capture.

	Opportunities for economies of scale are much reduced at band 6 than they are at bands 1-3. There is a significant difference in cost per unit of load between bands 1-3 and 4+, but much less differentiation between band 6 and bands 4/5. As a result if a weighted average scale driver is not included, we have a preference for a band 1-3 driver over a band 6 driver.	
Complexity (driven by Geography / Geology)	**EE3.** Wastewater treatment complexity is driven by the consents that need to be delivered. This was allowed for by using an explanatory variable of NH_3 load <3mg/l. However, as consents tighten across multiple parameters, costs increase further. Therefore, additional or composite explanatory variables are required, particularly to account for tight P consents which are very expensive to treat to.	**EE3.** The specification of the network complexity driver is important: • Whilst pumping capacity / length has logical appeal, the data quality is poor and therefore may be providing spurious explanatory power. • Rising mains and combined sewers provide legitimate alternative complexity drivers.
Weather effects	**EE4.** Weather less likely to be a material driver of cost than with WRP, TWD and SWC.	**EE4.** Weather describes cost – High intensity rainfall in urban areas should increase sewerage costs (more pumping, managing sewer flooding / CSO spill events).
AMP effects	**EE5.** AMP years are logical but over fitting issues are an increased risk with wastewater models due to the reduced number of companies in the data panel.	**EE5.** AMP years are logical but over fitting issues are an increased risk with wastewater models due to the reduced number of companies in the data panel.

8.1.3 Bioresources

Table 38: Bioresources engineering expectations

Cost driver	Basis
Scale	**EE1.** TDS (Tonnes of Dry Solids received) is an uncontentious scale driver. We have specified models on a unit cost basis therefore inclusion of a scale driver will convert to a description of (dis)economies of scale at a company level.
Economies of scale in sludge collection	**EE2.** Sludge treatment requires the development of complex assets that are discrete from the sewage treatment flow path. It is not economic to construct sludge treatment assets at each sewage works. **EE3.** Sludge treatment facilities have been constructed at the large sewage works as they have on onsite supply of sludge and can benefit from economies of scale of treatment assets. As the size of sewage treatment works get smaller so these economies of scale reduce and eventually to the point where onsite treatment is not economic. At this point treatment of sludge at a regional hub is required with associated intersiting costs. Intersiting costs are a function of both: ○ the need to treat at regional hubs (i.e. sewage treatment works size) ○ the transport requirements to the sludge treatment hub (i.e. the distance travelled and the method of transport) **EE4.** Economies of scale could be directly accounted for using information about the size of sewage works and intersiting 'work' done. **EE5.** Population Density can be considered as a proxy for opportunities for economies of scale. However, it will not be able to distinguish between the various cost drivers related to economies of scale.
Sewage treatment complexity	**EE6.** The characteristics of the sludge being treated will drive costs. Sludges generated from sewage treatment works with tight ammonia will have a higher % of secondary sludge which is more complex to treat. Tight phosphorus consents will have more inert sludge because of the dosing requirements of treatment. Inert sludge has lower renewable energy potential.
Disposal complexity	**EE7.** The way in which sludge is disposed of will drive costs. Disposal to land (either farmland or land reclamation) is significantly more economic that alternative thermal processes (pyrolysis, gasification, incineration). This is because it is highly inefficient to combust material with elevated water content. **EE8.** The opportunities to dispose to land are impacted by the availability of land (i.e. rurality of surrounding areas and sludge to land regulations) and the challenges of transporting it to the disposal site. **EE9.** Disposal complexity can be directly accounted for, or population density may form a weak proxy.
Sludge treatment complexity	**EE10.** Sludge treatment processes impact on the quality of the treated product and the opportunities for renewable energy. They are an important component in managing landbank risk. Advanced anaerobic digestion (ADD) requires significantly more complex assets which in turn have more stringent maintenance requirements. This generates an improved product (reduced pathogens and % dry solids) which is more attractive to landowners and cheaper to transport. It also leads to increased renewable

energy yields. In summary, both cost pressures and opportunities are elevated. This leads to a complex overall cost driver picture.

8.1.4 Retail

Table 39: Retail prior expectations

Cost driver	Basis
Scale	**EE1.** Households served is an uncontentious scale driver. We have specified models on a unit cost basis therefore inclusion of a scale driver will convert to a description of (dis)economies of scale at a company level. As the number of households served by a company increases, we would expect increasing economies of scale as the size of assets and purchasing power increases. We would anticipate that the cost to bill a given customer and respond to routine queries should be covered by the scale driver.
Ability to pay / Deprivation	**EE2.** Customers who are struggling to pay are more likely make contact to query billing and seek support. This is also likely to be a major driver of bad debt costs. **EE3.** The size of the water bill is a strong and direct driver of retail traffic (more queries, more requirements for support, greater likelihood of debt costs) **EE4.** There are a wide range of external potential deprivation metrics that can be used which will for a proxy for ability to pay. **EE5.** Population density is also likely to form a wider proxy, as more urban areas typically have more deprivation issues.
Meter reading	**EE6.** The volume of meters to read will drive cost. However, this will be complicated by the metering technology installed (AMI technologies and ability to remotely read) and the attributes of the supply area / metered population (traffic congestion, proximity of houses to the road). **EE7.** Population density is also likely to form a wider proxy, for metering costs but is likely to be non-linear (costs at extremes of density and rurality).
Population transience	**EE8.** Where there are significant changes in the population served, account changes increase and there is a higher likelihood of debt management / bad debt as customers leave without closing accounts. **EE9.** Population density may form a wider proxy, there is likely to be larger transience in more urban areas, particularly in student populations.

9. Appendix 2: Coverage of engineering expectations in Ofwat's consultation models

The models in the proposed PR24 suite have been assessed relative to the PR19 models.

EE = Engineering expectation. Cells coloured green where satisfied. In Wholesale water and Network plus waste models, cells coloured light green when satisfied for one of TWD/WRP or SWC/SWT. Cells also coloured light green where the engineering expectation is partially satisfied (in these cases, the expectation has not been classified as met).

The statistical tests are coloured against PR19 model performance on the PR24 dataset. Red shows a deterioration in performance, Amber shows no change, Green shows an improvement in performance with the same thresholds applied as in ***Table Chapter 7***. Performance is relative to the model with the same point of triangulation (e.g % water treated in bands 3-6 vs WAC in WRP).

9.1 Water Resources Plus

The Water Resources Plus models have shown no improvement in the engineering logic being captured. For most models, there has also been no improvement in performance, and in some cases a substantial decline.

Table 40: Coverage of engineering expectations and statistical performance of Ofwat's PR24 WRP consultation models.

Model	PR19 WRP1	PR19 WRP2	PR24 WRP1	PR24 WRP2	PR24 WRP3	PR24 WRP4	PR24 WRP5	PR24 WRP6
Scale (EE1)	Props	Props	Props	Props	Props	Props	Props	Props
Density	LAD	LAD	LAD from MSOA	LAD from MSOA	MSOA	MSOA	Properties /Length	Properties /Length
GW effects (EE2)								
Density 2 (EE3)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complexity bands (EE4)	Bands 3-6	Weighted	Bands 3-6	Weighted	Bands 3-6	Weighted	Bands 3-6	Weighted
Treatment Pumping (EE5)								
Raw water pumping (EE6)								
Weather (EE7)								
EEs met (of 8)	1	1	1	1	1	1	1	1
R-squared	0.917	0.907	0.909	0.902	0.901	0.896	0.910	0.905
R-squared Unit	0.389	0.305	0.326	0.267	0.260	0.219	0.354	0.305
RESET	0.438	0.324	0.435	0.369	0.768	0.735	0.326	0.205
AIC	-23	-20	-21	-19	-18	-17	-20	-18
Efficiency Range	1.49	1.48	1.49	1.48	1.50	1.51	1.47	1.46
Change in EE from PR19	-	-	0	0	0	0	0	0
No. of parameters (inc. constant)	5	5	5	5	5	5	5	5

9.2 Treated Water Distribution

The Treated Water Distribution models have shown limited improvement in the engineering logic being captured (TWD3 and TWD6 have improved with the uniform measure of density). Half of the models show no improvement in performance, while the other half do. Notably, the models that include APH and/or capture additional engineering logic have shown an improvement in performance. TWD1 also fails the RESET test at the 10% level.

Table 41: Coverage of engineering expectations and statistical performance of Ofwat's PR24 TWD consultation models.

Model	PR19 TWD	PR24 TWD1	PR24 TWD2	PR24 TWD3	PR24 TWD4	PR24 TWD5	PR24 TWD6
Scale (EE1)	Length	Length	Length	Length	Length	Length	Length
Density (EE2)	LAD	LAD from MSOA	MSOA	Properties /Length	LAD from MSOA	MSOA	Properties /Length
Density 2	yes	yes	yes	yes	yes	yes	yes
Boosters/ length (EE3)	yes	yes	yes	yes			
TWD pumping (EE4)					yes	yes	yes
Weather (EE5)							
EEs met (of 6)	2	2	2	3	2	2	3
R-squared	0.957	0.955	0.952	0.958	0.961	0.965	0.966
R-squared Unit	0.620	0.606	0.586	0.636	0.651	0.687	0.697
RESET	0.101	0.089	0.12	0.491	0.460	0.721	0.844
AIC	-128	-126	-127	-134	-131	-138	-140
Efficiency Range	0.61	0.60	0.67	0.64	0.59	0.61	0.54
Change in EE from PR19	-	0	0	1	0	0	1
No. of parameters (inc. constant)	5	5	5	5	5	5	5

9.3 Wholesale Water

The Wholesale Water models have shown no improvement in the engineering logic being captured. For most models, there has also been no improvement in performance by R-squared, but a substantial decline in performance with respect to the other metrics that we have considered. Some models also failed the RESET test at the 10% level.

Table 42: Coverage of engineering expectations and statistical performance of Ofwat's PR24 WW consultation models (WW1-WW5).

Model	PR19 WW1	PR19 WW2	PR24 WW1	PR24 WW2	PR24 WW3	PR24 WW4	PR24 WW5
Scale (WRPEE1/TWDEE1)	Props	Props	Props	Props	Props	Props	Props
Density	LAD	LAD	LAD from MSOA	LAD from MSOA	MSOA	MSOA	Properties /Length
GW effects (WRPEE2)							
Density 2 (WRPEE3/TWDEE2)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complexity bands (WRPEE4)	Bands 3-6	Weighted	Bands 3-6	Weighted	Bands 3-6	Weighted	Bands 3-6
Boosters/length (TWDEE3)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Treatment Pumping (WRPEE5)							
Raw water pumping (WRPEE6)							
TWD pumping (TWDEE4)							
Weather (WRPEE7/TWDEE5)							
EEs met (of 14)	4	4	4	4	4	4	4
R-squared	0.970	0.971	0.965	0.976	0.963	0.965	0.965
R-squared Unit	0.532	0.547	0.457	0.486	0.422	0.458	0.465
RESET	0.222	0.122	0.164	0.075	0.178	0.075	0.205
AIC	-161	-162	-156	-159	-156	-158	-157
Efficiency Range	0.61	0.65	0.73	0.74	0.79	0.79	0.70
Change in EE from PR19	0	0	0	0	0	0	0
No. of parameters (inc. constant)	6	6	6	6	6	6	6

Table 43: Coverage of engineering expectations and statistical performance of Ofwat's PR24 WW consultation models (WW6-WW12).

Model	PR24 WW6	PR24 WW7	PR24 WW8	PR24 WW9	PR24 WW10	PR24 WW11	PR24 WW12
Scale (WRPEE1/TWDEE1)	Props	Props	Props	Props	Props	Props	Props
Density	Properties /Length	LAD from MSOA	LAD from MSOA	MSOA	MSOA	Properties /Length	LAD from MSOA
GW effects (WRPEE2)							
Density 2 (WRPEE3/TWDEE2)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Complexity bands (WRPEE4)	Weighted	Bands 3-6	Weighted	Bands 3-6	Weighted	Bands 3-6	Weighted
Boosters/length (TWDEE3)	Yes						
Treatment Pumping (WRPEE5)							
Raw water pumping (WRPEE6)							
TWD pumping (TWDEE4)		Yes	Yes	Yes	Yes	Yes	Yes
Weather (WRPEE7/TWDEE5)							
EEs met (of 14)	4	4	4	4	4	4	4
R-squared	0.968	0.965	0.965	0.961	0.962	0.966	0.967
R-squared Unit	0.501	0.456	0.464	0.401	0.416	0.47	0.485
RESET	0.072	0.838	0.821	0.895	0.935	0.781	0.614
AIC	-160	-159	-161	-156	-158	-160	-162
Efficiency Range	0.70	0.74	0.74	0.72	0.73	0.71	0.71
Change in EE from PR19	0	0	0	0	0	0	0
No. of parameters (inc. constant)	6	6	6	6	6	6	6

9.4 Sewage Collection

The Sewage Collection models have shown limited improvement in the engineering logic being captured (SWC4, SWC5, and SWC6 have improved with the inclusion of Urban Rainfall). SWC1 was used at PR19.

There has been some improvement in statistical performance in these models, however, with the new MSOA-based density measures performing better than the LAD measure used at PR19. Notably, the models capturing additional engineering logic (rainfall) appear to be more performant by any metric that those in which it is absent.

Table 44: Coverage of engineering expectations and statistical performance of Ofwat's PR24 SWC consultation models.

Model	PR19 SWC1	PR19 SWC2	PR24 SWC1	PR24 SWC2	PR24 SWC3	PR24 SWC4	PR24 SWC5	PR24 SWC6
Scale (EE1)	Sewer length	Sewer length	Sewer length	Sewer length	Sewer length	Sewer length	Sewer length	Sewer length
Economies of Scale (population density) (EE2)	Properties /Length	LAD	Properties /Length	LAD from MSOA	MSOA	Properties /Length	Lad from MSOA	MSOA
Density 2								
Complexity (EE3)	Pumping capacity	Pumping capacity	Pumping capacity	Pumping capacity	Pumping capacity	Pumping capacity	Pumping capacity	Pumping capacity
Weather (EE4)						Yes	Yes	Yes
EEs met (of 5)	1	1	1	1	1	2	2	2
R-squared	0.917	0.884	0.917	0.889	0.889	0.919	0.909	0.908
R-squared Unit	0.621	0.476	0.621	0.498	0.503	0.632	0.585	0.582
RESET	0.356	0.308	0.356	0.308	0.256	0.172	0.346	0.323
AIC	-120	-112	-120	-113	-114	-123	-119	-120
Efficiency Range	0.22	0.36	0.22	0.34	0.37	0.24	0.28	0.26
Change in EE from PR19	-	-	0	0	0	1	1	1
No. of parameters (inc. constant)	3	3	3	3	3	3	3	3

9.5 Sewage Treatment

The Sewage Treatment models have shown limited improvement in the engineering logic being captured (SWT3 has improved with the addition of weighted average treatment works size). SWT1 was used at PR19.

There has been some improvement in statistical performance in these models, however, with the new economies of scale variables appearing to be more performant than those used at PR19. Notably, the model capturing additional engineering logic (weighted average treatment works size) appears to be the most performant model.

Table 45: Coverage of engineering expectations and statistical performance of Ofwat's PR24 SWT consultation models.

Model	Ofwat PR19 (SWT1)	Ofwat PR19 (SWT2)	PR24 SWT1	PR24 SWT2	PR24 SWT3
Scale (EE1)	Load	Load	Load	Load	Load
Economies of Scale (population density) (EE2)	Size band 1-3	Size band 6	Size band 1-3	% over 100k	Weighted
Complexity (EE3)	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3
EE met (of 4)	1	1	1	1	2
R-Squared	0.854	0.855	0.854	0.869	0.911
R-squared unit	0.362	0.367	0.362	0.420	0.609
RESET	0.059	0.157	0.056	0.272	0.849
AIC	-105	-107	-105	-106	-116
Efficiency Range	0.68	0.65	0.68	0.54	0.33
Change in EE from PR19	-	-	0	0	1
No. of parameters (inc. constant)	3	3	3	3	3

9.6 Wastewater Network Plus

There were no Network Plus models at PR19 from which to draw comparisons with the PR24 suite. Here, we compare models to the worst performing in the suite (coloured red for each test metric). The best model by each metric is coloured blue.

We can see that the two models that include weighted average treatment works scale are the best in their triangulation state (with or without rainfall), and the models that include rainfall perform better than the corresponding model that does not.

The worst performing models by most metrics that do not include and do include rainfall are WWNP1 and WWNP5 respectively. Neither of these account for any form of economies of scale at treatment works.

We note that WWNP7 fails the RESET test.

Table 46: Coverage of engineering expectations and statistical performance of Ofwat's PR24 WWNP consultation models.

Models	PR24 WWNP1	PR24 WWNP2	PR24 WWNP3	PR24 WWNP4	PR24 WWNP5	PR24 WWNP6	PR24 WWNP7	PR24 WWNP8
Scale (SWCEE1/SWTEE1)	Load	Load	Load	Load	Load	Load	Load	Load
Density/Economies of Scale (SWCEE2/SWTEE2)		Size bands 1-3	% over 100k	Weighted Scale		Size bands 1-3	% over 100k	Weighted Scale
Density 2								
Treatment Complexity (SWTEE3)	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3	Ammonia <3
Network Complexity (SWCEE3)	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length	Pumping capacity per length
Weather (SWCEE4)					Yes	Yes	Yes	Yes
EEs met (of 7)	1	1	1	2	2	2	2	3
R-squared	0.947	0.952	0.949	0.956	0.953	0.959	0.956	0.964
R-squared Unit	0.71	0.738	0.72	0.758	0.741	0.774	0.761	0.801
RESET	0.572	0.478	0.700	0.901	0.241	0.109	0.009	0.248
AIC	-192	-193	-190	-195	-194	-197	-195	-202
Efficiency Range	0.15	0.17	0.15	0.14	0.17	0.14	0.13	0.09
No. of parameters (inc. constant)	3	4	4	4	4	5	5	5

9.8 Bioresources

For ease of comparison, we transform all the models to a unit cost model.

The bioresources models have seen a decrease in the number of engineering expectations captured since PR19, and while some models have seen an improvement in performance, there has predominantly been a deterioration in performance.

These models explain very little of the variation between companies. Particularly when we consider that more cost is now being captured in the dependent variable, it seems strange that the models have become more parsimonious.

We note that the models that capture the most engineering logic appear to be the most performant.

We note that BR5, BR6, BR8, and BR9 all fail the RESET test.

Table 47: Coverage of engineering expectations and statistical performance of Ofwat's PR24 BR consultation models.

Unit cost model	PR19 BR1	PR19 BR1	BR1	BR2	BR3	BR4	BR5	BR6	BR7	BR8	BR9	BR10
Scale	Sludge Prod.	Sludge Prod.	Sludge Prod.	Sludge Prod.	Sludge Prod.	Sludge Prod.	Sludge Prod.	Sludge Prod.				
Economies of Scale	PctBands 13	STWs / Prop	PctBands 13	PctBands 13	STWs / Prop	PctBands 13			PctBands 13			STWs / Prop
Density	Prop / Length		LAD from MSOA	MSOA			LAD from MSOA	MSOA		LAD from MSOA	MSOA	
EEs met (of 5)	3	2	3	3	2	2	2	2	1	1	1	1
R-squared (unit)	0.257	0.144	0.294	0.275	0.144	0.28	0.123	0.107	0.239	0.124	0.108	0.133
RESET	0.179	0.280	0.766	0.609	0.280	0.403	0.000	0.020	0.508	0.000	0.005	0.445
AIC	30	32	30	31	32	29	32	33	28	31	31	30
Efficiency Range	0.92	0.87	0.75	0.83	0.87	0.85	0.84	0.89	0.90	0.86	0.90	0.92
Change in EE from PR19	-	-	0	0	-1	-1	-1	-1	-2	-2	-2	-2
No. of parameters (inc. constant)	4	3	4	4	3	3	3	3	2	2	2	2

9.10 Retail Total Cost

The R-squared values reported here are all unit cost.

While there has been a reduction in the number of engineering expectation captures, we accept that here this is for legitimate reasons (variables not performing well).

There has been an increase in the performance of the models in this area since PR19, although we note that the models that include economies of scale do perform better than those that don't.

Table 48: Coverage of engineering expectations and statistical performance of Ofwat's PR24 RTC consultation models.

Unit cost model	PR19 RTC1	PR19 RTC2	PR19 RTC3	RTC1	RTC2	RTC3	RTC4	RTC5	RTC6
Scale		Households	Households	Households	Households	Households			
Meter penetration	Yes	Yes	Yes						
Deprivation	% Default	% Default	Income Score	% Default	Partial Insight	Income Score	% Default	Partial Insight	Income Score
Transience			Yes						
Bill Size	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covid Dummies				Yes	Yes	Yes	Yes	Yes	Yes
EEs met (of 6)	3	3	5	4	4	4	3	3	3
R-squared (unit)	0.613	0.636	0.617	0.697	0.669	0.648	0.65	0.645	0.638
RESET	0	0.006	0.006	0.103	0.054	0.128	0.092	0.023	0.176
AIC	-153.193	-155.88	-151.952	-181.94	-175.513	-175.968	-176.031	-171.92	-174.131
Efficiency Range	0.49	0.50	0.52	0.53	0.54	0.57	0.49	0.47	0.56
Changes from PR19	-	-	-	-1	-1	-1	-2	-2	-2
No. of parameters	5	6	8	6	6	6	5	5	5