UUWR_27

PR24 Draft Determination: UUW Representation

Area of representation: Cost and PCDs - Enhancement modelling consultation

This document outlines UUW's response to Ofwat's econometric model consultation.

Reference to draft determination documents:
Expenditure allowances - Enhancement Cost Modelling Appendix
Expenditure allowances

United Utilities

Water for the North West

UUW's response to Ofwat's enhancement modelling consultation

Approach to model development and selection

Question 2.1) Do you agree with our decision to use OLS to estimate our scheme level enhancement models?

Scheme-level modelling is a relatively new approach to enhancement cost assessment, and we welcome Ofwat's willingness to explore the benefits provided by bottom-up datasets. In most cases, Ofwat has used relatively simple linear OLS models to generate cost predictions. These models are transparent and generally easy to interpret. However, simple models may not allow to fully leverage the cost granularity that is potentially available in a scheme-level dataset. We therefore consider that using OLS estimation is appropriate at this stage, but Ofwat should take steps to improve the modelling approach in the future, as described below.

To generate robust results, OLS regressions depend on several key assumptions, including homoskedasticity, normality of errors and no autocorrelation. We observe that in many enhancement models these assumptions do not hold. This is likely because Ofwat's pooled OLS models do not take into account the cross-sectional group dimension of the data (i.e. that each scheme is associated with a specific water company). Intuitively, each company will have company-specific factors which are likely to impact scheme costs. Ignoring this aspect can lead to several issues (e.g. no independence between model errors), undermining model robustness. At a minimum, we would expect Ofwat to always use clustered robust standard errors, as this will help to account for the inter-group error correlations. We also find that Ofwat's enhancement models typically fail the Breusch-Pagan LM test, which indicates that a Random Effects (RE) model would be more appropriate over pooled OLS. Ofwat should therefore look at ways how the group dimension of the data could be taken into account. It is, for instance, possible to run RE models in Stata by only specifying the cross-sectional dimension (and leaving the time dimension unspecified). Triangulating across a diverse model suite would also allow to mitigate the shortcomings of individual regressions.

Therefore, there are clear limitations with the OLS estimation approach as implemented by Ofwat at Draft Determinations (DD). While we consider that it is legitimate to continue using OLS as part of its approach, Ofwat must account for these limitations within the broader cost assessment framework. For example, setting a stringent catch-up challenge on the basis of parsimonious OLS models would be inappropriate - large residuals that result from excessively simplistic models should not so readily be interpreted as inefficiency.

We agree with Ofwat's assessment that top-down company-level models also have clear limitations, including small sample sizes and lack of cost driver granularity. We find that top-down enhancement models often do not perform well (e.g., low R-squared scores, counterintuitive coefficient signs). We therefore consider that using top-down models (whether in a triangulated model suite or standalone) is only justified when top-down assessment produces statically robust outcomes (such as in Sanitary parameters assessment).

Question 2.2) Do you agree with our decision to exclude outliers based on a Cook's distance threshold of 4 / N?

In any dataset, there are likely to be schemes which are atypical, either in terms of cost and cost driver information, or due to scheme-specific engineering and operational factors. Such observations could complicate econometric modelling in certain circumstances. Ofwat has chosen to use Cook's distance metric for outlier identification. This approach is relatively well established in statistical literature, although it is new for cost assessment in the water industry. We consider using Cook's distance to be an acceptable choice, but note that there exist many other statistical tools for outlier identification, which might produce different results. We therefore urge Ofwat to take a holistic approach and examine carefully how outlier exclusion affects the overall cost assessment framework.

We have identified several risks that might be particularly relevant. First, we are concerned about the effects of outlier exclusion on model performance. Ofwat has argued that excluding 'influential' schemes will help to achieve more robust econometric results.[1] However, we observe that in some cases there is actually a deterioration in model quality after removing Cook's distance outliers. For example, after excluding outliers in Phosphorus removal historical PR4 model, the adjusted R-squared decreases from 0.51 to 0.30. This suggests that outlier exclusion might be removing significant amounts of valuable statistical information, thereby creating a biased sample which might not accurately represent the efficient costs of delivering a full range of solutions.

Second, we consider that outlier exclusion can entrench a spurious relationship between cost drivers and forecast scheme costs. This is particularly true for a simplistic model, which may be underspecified. Outliers may simply represent schemes that are more impacted by omitted variables. Excluding those has the effect of inappropriately justifying the simplistic model (as opposed to allowing to formulate a more robust model that reflects all relevant cost drivers). Furthermore, forecast scheme cost data is under management control and will be influenced by both scheme-specific factors and company risk appetite.[2] Consider a scenario where companies submit unrealistically low cost forecasts - this would mean that due to outlier exclusion the model is likely to generate understated efficient cost predictions. We do not consider that this is an appropriate outcome.

Question 2.3) Do you agree with our approach to setting an efficient enhancement expenditure allowance for outlier schemes?

Ofwat has identified two categories of outliers: 'efficient' and 'inefficient'. Efficient outliers are schemes where the model predicted costs exceed the costs requested by the company, while inefficient outliers are those schemes where the opposite is true. If outliers are efficient, the company receives the requested cost allowance. If outliers are inefficient, the company receives model predicted costs, plus a discretionary uplift based on engineering deep dives.

We have strong concerns that Ofwat's approach generates an asymmetry in terms of how outlier allowances are set. In particular, we consider that capping efficient outlier allowances could potentially create perverse incentives. An enhancement programme will have a mix of schemes that are either more or less costly, relative to a model estimate. For example, costs might be unusually low due to site-specific factors or innovative solutions that are only applicable to that specific scheme. As the model establishes the 'average' relationship between costs and cost drivers, we would expect the excess allowance on efficient schemes to be broadly balanced out by deficit allowance on inefficient schemes. However, Ofwat's approach to outlier assessment breaks this balance. The table below illustrates the effects of outlier assessment, using Growth at STWs model as an example. The first row shows how much final allowances have been reduced (relative to model allowances) due to capping of efficient outliers across the industry. The second row shows how much companies have gained (relative to model allowances) through deep dives of inefficient outliers. It can be seen that, in this example, the negative effect of capping is much larger than the gains achieved through deep dives.

Table 1: Aggregate industry effects of outlier assessment in Growth at STWs

Allowance lost due to capping of efficient outliers	**-£269 m**
Allowance gained through deep dives of inefficient outliers	**£174 m**

Source: UUW analysis of "PR24-DD-WW-Growth-at-STWs-2.xlsx" model

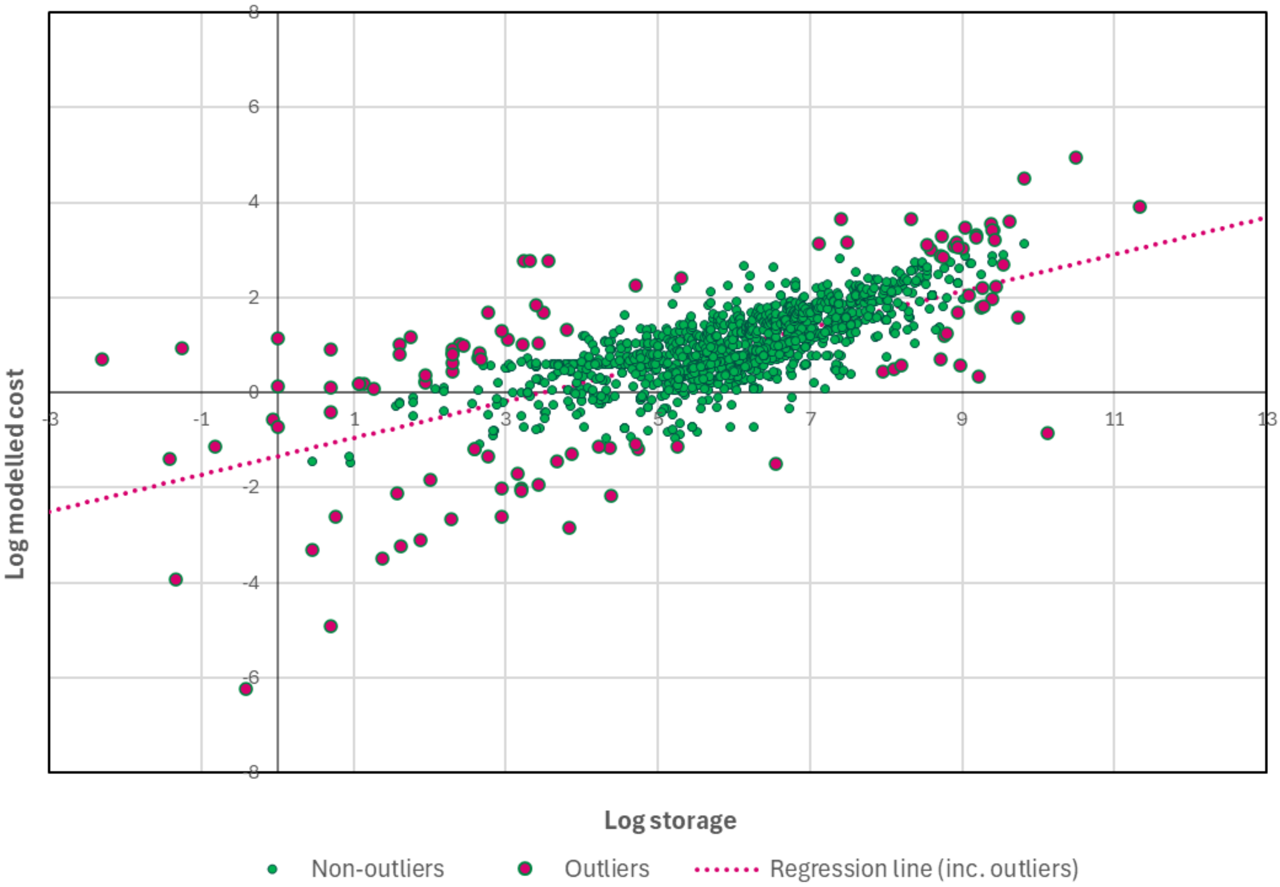
Capping efficient schemes effectively penalises companies for identifying cost effective solutions. Therefore, there is a risk that companies will be reluctant to propose low-cost schemes, as these might be identified as efficient outliers. Ofwat's approach may, as a result, incentivise companies to submit costs which do not reflect the most efficient solution, undermining model accuracy and potentially leading to higher overall costs for customers. We are concerned that this will limit Ofwat's ability to set efficient benchmarks in future price reviews, meaning customers do not benefit from efficiencies going forward. To avoid these risks, we consider that Ofwat should grant the full model predicted allowance for efficient outliers.

[1] Ofwat (2024) *PR24 Draft Determinations: Expenditure allowances - Enhancement cost modelling appendix*, p. 14.

[2] This issue is less relevant for historical datasets, as they are based on outturn data which is not under management control.

In addition, we also do not support Ofwat's decision to generate outlier modelled allowances on the basis of regressions that exclude outliers. We note that Cook's distance statistic does not measure inefficiency, but rather identifies observations that have extreme values relative to the mean of the data. Large and complex schemes, for instance, will almost certainly be identified as outliers, due to their atypical cost and cost driver values. It is wrong to assume that such schemes can be benchmarked against 'normal' (i.e., non-outlier) schemes, as the relationship between costs and cost drivers will likely not be applicable to these large and complex schemes. An example of this is shown in Figure 1. This graph plots storm overflows network log model schemes (by storage and cost), where outlier observations are coloured red. It can be seen that there are many small outlier schemes which are relatively cheap (i.e. below trendline), as well as a group of schemes which are large and relatively expensive (i.e. above trendline). Excluding these schemes from modelling in effect removes valuable statistical information about the full range of schemes (including atypical schemes). We find that in the network log model, excluding outliers from the dataset tends to increase the cost predictions of small outlier schemes, while understating the costs of larger outlier schemes. A more appropriate approach would be to set outlier model allowances using the 'full' model which includes all observations. In the network log model, we find that recalculating model allowances using this method reduces average standardised residuals of outlier schemes from 0.39 to 0.24. This result indicates that using the full dataset leads to less unexplained variation in outlier scheme costs.

Figure 1. Storm overflows network log model scheme scatterplot



Source: UUW analysis

For the above reasons, we consider that outlier assessment could be improved in two key ways: first, by not capping efficient outlier schemes and granting the model allowance; second, by incorporating the full dataset when setting outlier scheme model allowance predictions. The first improvement would remove perverse incentives to avoid submitting low-cost schemes. The second improvement would allow to set a more accurate model prediction for outlier schemes by incorporating all relevant statistical information. We consider that the combination of these two changes will make outlier assessment more symmetrical and preserve incentives for efficiency going forward.

Question 2.4) Do you agree with our decision to apply the PR19 log-bias adjustment to address log-bias (where relevant)?

The log transformation bias is a statistical phenomenon that can lead to inappropriately low predicted cost estimates, and we welcome Ofwat's acknowledgment of this issue. Econometric literature suggests that log-transformation bias is expected to be negative[3] (i.e. predicted costs will be lower than they should be), meaning that a targeted adjustment is both necessary and justified.

Linear OLS models operate on the principle that the sum of residuals is equal to zero. In the context of enhancement modelling this means that the sum of industry forecast costs will always equal the sum of linear model predictions. However, in log-log models, the log-transformation bias artificially depresses the value of the exponentiated predictions. This issue appears to be material: it can be observed, for example, that in the overflows network grey/hybrid log-log model, total industry submitted costs are around 12% higher than total industry model predicted costs (before applying catch-up challenge). This should not be surprising – as scheme-level datasets are often highly skewed, with some schemes having very high costs, this generally results in larger log-transformation bias.

We support Ofwat's approach of applying a uniform upwards adjustment factor so that total industry predicted costs equal total industry submitted costs. Although there exist several alternative adjustment methods (e.g., Alpha factor), we consider that Ofwat's chosen approach is more transparent and intuitive; it also helps to maintain consistency between total industry predicted costs across logarithmic and linear models.

Question 2.5) Do you agree with our decision to set the efficiency benchmark at the company level instead of scheme level?

We support Ofwat's approach of setting the catch-up challenge at an aggregated company level. Scheme-level efficiency scores exhibit considerable variability, with some schemes displaying very low costs, likely due to site-specific factors not accounted for by the models. Consequently, applying a scheme-level catch-up challenge risks setting the benchmark on the basis of these atypically low-cost schemes, which might not accurately represent efficiency at the programme level. We consider that setting a company-level catch-up challenge helps to mitigate this risk by providing a more in-the-round view of enhancement programme efficiency. This approach ensures consistency in how the efficiency challenge is set across both base and enhancement models.

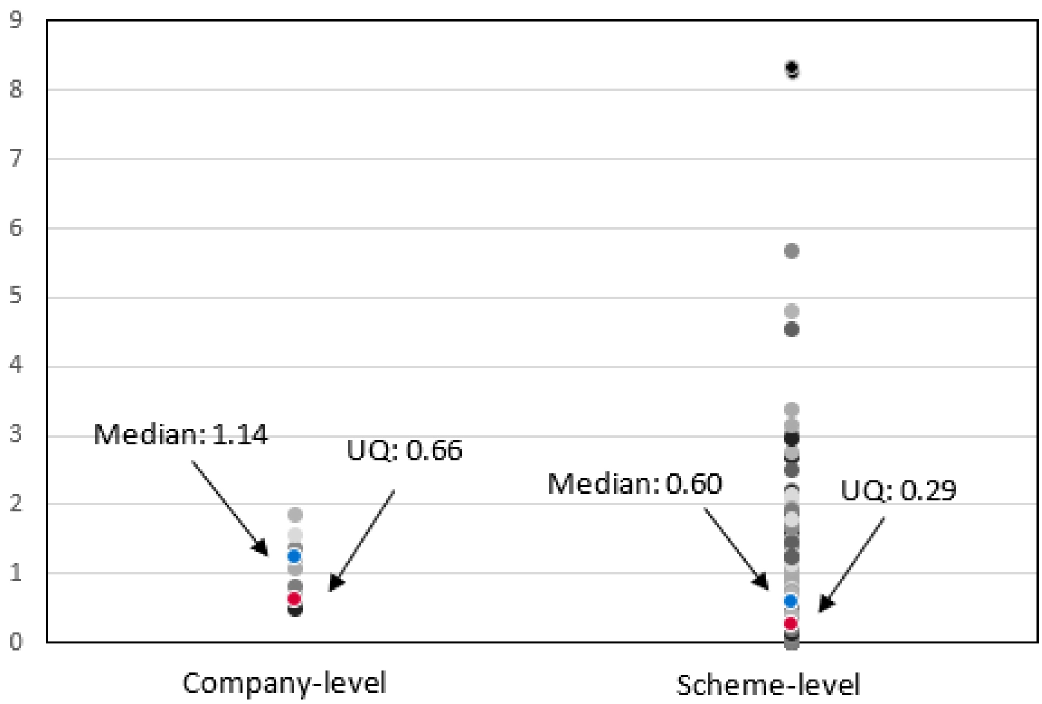
We appreciate Ofwat's concerns that company-level benchmarking sometimes leads to counterintuitive catch-up challenge adjustment factors. Growth at STWs triangulated enhancement model is a case in point: here, the company-level median catch-up challenge is above 1, while the upper-quartile challenge is around 0.66. Accordingly, setting the median challenge would imply that companies should receive more than the triangulated model prediction, whereas the upper quartile challenge would imply a potentially punitive cost reduction.

In such situations, it might be tempting to consider setting a scheme-level efficiency challenge. However, analysis of enhancement models indicates that this approach is unlikely to be appropriate. As mentioned, scheme-level datasets often exhibit significant skewness, with some (non-outlier) schemes having exceptionally low or high costs. This is illustrated in Figure 2, which shows the distribution of scheme-level and company-level efficiency scores from the Growth at STWs triangulated model. As shown, scheme-level scores exhibit much greater variability, with some schemes being extremely efficient or inefficient. In this instance, the scheme-level median challenge is actually more stretching than the company-level upper quartile challenge and would imply an severe

[3] Wooldridge, J. M., 2002. *Introductory Econometrics: A Modern Approach*. 2e ed. South-Western, p. 202.

cost reduction. We observe a similar pattern of scheme-level efficiency challenge being very stringent in other enhancement areas (e.g. overflows). We consider that a scheme-level efficiency challenge would risk creating an expectation that all schemes can be delivered in line with an unusually cheap scheme. This would be an unacceptable outcome.

Figure 2: Growth at STW model efficiency scores



Source: UUW analysis of "PR24-DD-WW-Growth-at-STWs-2.xlsx" model

In addition, we are concerned that Ofwat's decision on whether to set an additional catch-up efficiency challenge (beyond 'average' efficiency implied by the models) is somewhat arbitrary. As discussed, in the Growth at STWs model, the median challenge was deemed insufficiently stretching, while the upper quartile challenge was considered excessively demanding. As a result, Ofwat decided not to impose any additional catch-up challenge – we support this decision. However, in the Bioresources IED models, very stretching company-level catch-up challenge adjustments are still applied, despite the poor model fit (we provide further analysis in section 0). We are unclear about the criteria Ofwat uses to determine whether to impose additional catch-up challenge across different enhancement areas. Models with poor statistical fit often result in a broad range of efficiency scores, even at the company level. Consequently, the discrepancy between predicted and actual costs cannot be solely attributed to company cost inefficiency. As we set out in our Future Ideas Lab paper "Principles of Regulatory Cost Assessment"[4], efficiency targets should be set relative to the quality of the model. This also entails that that when model performance is inadequate, no additional catch-up challenge should be applied.

On a more technical note, we have observed inconsistencies in how percentiles are calculated in Ofwat's Excel formulas. For example, the Growth at STWs spreadsheet uses the "PERCENTILE.EXC" formula to estimate the upper quartile challenge, whereas other enhancement models employ the "PERCENTILE.INC" formula (or its equivalents). These two formulas determine whether the endpoints of the data range are included in the calculation of percentiles, and can drive material differences in catch-up challenge adjustment (i.e. several percentage points). To avoid artificial distortions, it is crucial to use a consistent method for estimating percentiles across different Excel models, as well as other statistical software (e.g., Stata). We believe that exclusively using "PERCENTILE.INC" (and its equivalents) would be preferrable, as this statistic provides more intuitive results by incorporating the full range of available data.

---

[4] United Utilities, 202. *Future Ideas Lab: The Principles of Regulatory Cost Assessment, p. 21.* Available here: https://www.unitedutilities.com/globalassets/documents/pdf/the-principles-of-regulatory-cost-assessment.pdf

Storm overflows

Question 3.1) Do you agree with our approach to assessing grey and grey-hybrid storage storm overflow enhancement costs?

Our full representations on Ofwat's approach to storm overflows can be found in *'UUWR_10_storm overflows'*.

Ofwat has modelled £10.6bn of storm overflow enhancement expenditure using simplistic models. These have been first revealed as part of the draft determination with the only interaction with companies conducted via written queries, responses to which tended to be constrained by a two working day turnaround. This contrasts with the approach taken to base cost assessment, which was informed by engagement with industry cost assessment working groups and consultations in the lead up to both PR19 and PR24.

As a consequence of its approach, Ofwat has not collected appropriately detailed and granular high-quality data on the scheme costs and on potential explanatory variables that could explain the key underlying exogenous cost drivers. Instead, Ofwat's draft determination is reliant upon a severely limited dataset. In our view, it does not enable Ofwat's benchmarking to reasonably reflect fundamental cost drivers including urban/rural location, contaminated land, different solution types and so on. Additional engagement with the industry could have resulted in a more robust, high-quality dataset and constructive input on the underlying engineering rationale, which could have better informed Ofwat's approach.[5]

For an issue of this importance and scale, it is reasonable for us to have expected Ofwat to engage with companies during its model development process. The simplistic nature of the models indicates a lack of engagement with engineering rationale (as set out in section 3 of our main overflows representation document). It also appears to have led to issues with data quality as we discuss in the next section.

There are clear issues with data quality

We have identified clear data quality issues within the dataset Ofwat uses to set storm overflow allowances.

Data appears inconsistent across the industry

Companies appear to have interpreted Ofwat's guidance in different ways. This means that Ofwat's benchmark will not be consistent across companies. For example, we have not included the equivalent storage of Pass Forward Flow solutions in our business plan submission while other companies have[6]. We note that some of the companies that have included the storage avoided by FTFT schemes within their equivalent storage data are those that are not included within Ofwat's FTFT assessment. This suggests that there is a risk that Ofwat is assessing FTFT costs inconsistently across the industry. The implication of this is that relative unit costs may appear misaligned, which risks leading Ofwat to an inappropriate conclusion on relative efficiency.

[5] We also note that Ofwat did not initially publish its storm overflow model calculations at DD. Instead, it published its calculations two weeks after the Draft Determinations were published in response to an inbound query. This left us with only five weeks to robustly examine Ofwat's proposals and prepare our response. Despite this, we have done what we can to present as robust an approach as possible.

[6] In our further submission to OFW-OBQ-UUW-178, we provided Ofwat with our equivalent storage volumes for FTFT solutions and suggested that Ofwat could use these to ensure that it assesses FTFT and storage costs consistently across the industry. We note that Ofwat is proposing to assess FTFT separately and, as such, have not reflected the equivalent storage related to FTFT schemes in our DD submission.

Figure 3: Companies appear to have taken different approaches to reflecting FTFT volumes in data tables

Company	Avoided FTFT storage included?	Included in Ofwat's FTFT assessment?
Anglian	Yes	No
Severn Trent	Yes	No
Welsh Water	Yes	No
Southern	Yes	Yes
South West	Unclear	Yes
Yorkshire	Unclear	Yes
Wessex	Unclear	No
Hafren	Unclear	No
UUW	No	Yes

Source: UUW analysis of other company business plans

The use of forecast data means there is a risk Ofwat is not assessing efficiency but cost forecasting methodology differences

Ofwat's approach to benchmarking storm overflow expenditure is based on scheme level forecast data. This means Ofwat uses the forecast cost and equivalent volumes for each scheme within its benchmark. Ofwat does not include any historical data in its regressions. The only type of historical data Ofwat considers is historical forecast data from PR19 to compare unit costs. However, it isn't clear whether the PR19 schemes have a similar range of exogenous site-specific characteristics to those within AMP8.

We do acknowledge that using forecast AMP8 costs is legitimate. This is because the Environment Act 2021 has changed the regulatory framework surrounding storm overflows and is leading companies to intervene at overflows that were previously considered as non-cost beneficial. As such, forecast costs will better reflect cost pressures within AMP8. However, this does mean that Ofwat must account for the other issues that using forecast costs creates.

The implication of using forecast costs is that the benchmarking model is not based on observable differences in outturn efficiency between companies. Instead, the model reveals the different forecasting approaches that companies have used. Ofwat has applied a similar approach to the cost assessment of other enhancement areas but the risks inherent in this approach are greater when:

(i) the benchmark is set by a company that has adopted a simplistic approach to cost forecasting;

(ii) the investment programmes are inherently complex; and

(iii) companies do not have sufficient experience of a programme of this type and scale.

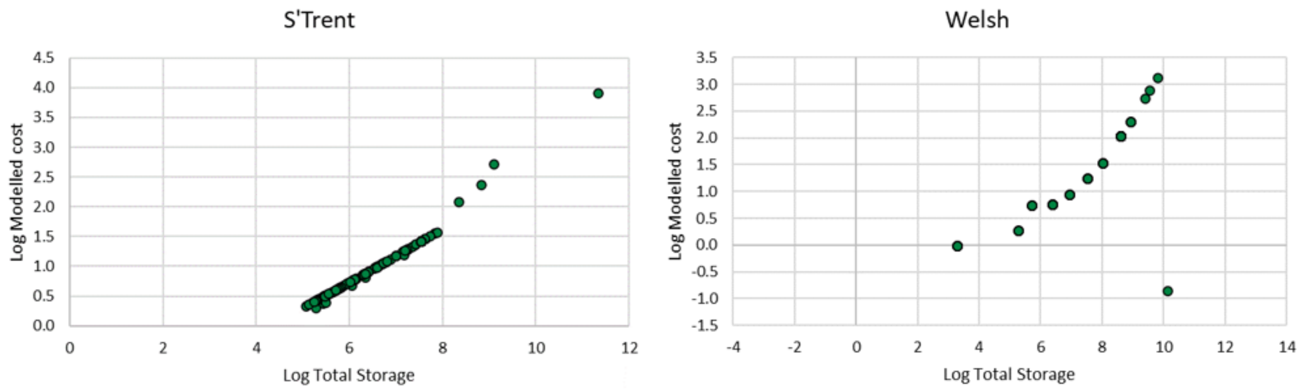
All factors are relevant in the case of storm overflows. For example, Figure 4 shows the range of efficiency scores across the industry, overlaid with Ofwat's triangulated upper quartile challenge. It is clear that, under Ofwat's methodology, Severn Trent is setting the efficiency benchmark.

Figure 4: Severn Trent sets the upper quartile in network models

Source: UUW analysis of "PR24-DD-WW-Storm-overflows-econometric-model.xlsx"

The problems with this become apparent when considering the scheme level data submitted by different companies. Figure 5 below illustrates the cost forecasts for network schemes submitted by Severn Trent Water and Welsh Water on a log-log scale. The scatterplots show that there is almost no variation in costs other than that driven by cubic metres of storage.

Figure 5: Some companies' network storage cost estimates are clearly oversimplistic



The most likely explanation of the pattern shown in these charts is that these companies have used highly simplistic top-down methods to forecast costs by scheme with volume as the primary cost driver. While there is some limited non-scale-related variation, it would be clearly non credible to suggest that:

- These companies' cost estimates are materially driven by a range of exogenous cost drivers. It is clear that volume is the primary cost driver; and
- the patterns set out in Figure 5 are realistic forecasts of the costs likely to be incurred at each of the sites.

In turn, these simplistic cost forecasts appear to lead Ofwat to conclude that volume is the only cost driver of relevance. For example, in its response to OFW-IBQ-UUW-008, Ofwat states:

"Some companies have no outliers and therefore we infer that the model includes all costs relating to storage solutions, including a range of ground conditions and site constraints."

We are concerned that Ofwat may not appreciate that its inference in this statement is informed by the cost forecasting methodology employed by individual companies and, crucially, not the underlying engineering rationale. Ofwat appears to suggest that the fact no companies have outliers is evidence that its model is capturing all relevant cost drivers. However, Ofwat may wish to reconsider the direction of causality inherent within its statement. We would consider a more appropriate statement to be that: its model appears to capture all relevant cost drivers because some companies have costed their schemes using a simple unit cost model – as such, mechanistically, these companies will not have any outliers **if Ofwat also aligns its cost model to a simple unit cost model**. As a result, we consider that Ofwat's approach risks being characterised as 'data fitting'.

We are clear that the relationships set out in Figure 5 do not imply that the actual costs for each scheme and site will only be driven by cubic metres of storage. Instead, it means that site-specific characteristics were not considered by these companies when developing their cost forecasts. This does not mean that site-specific factors will not drive differences in efficient costs. Equally, we must stress that this does not mean that Ofwat's model is already reflecting these site-specific factors. To suggest otherwise would appear misguided.

We accept that the simplicity of the benchmark company's approach to cost forecasting may limit the ability of forward-looking benchmarking models to reflect variation in site-specific costs. This is because the benchmark can only capture those factors that companies reflect in their costings. If, for example, some companies only use cubic metres of storage to inform their cost forecasts, any other site-specific explanatory variables are likely to be unsupported by the models. We are clear that this is not because these site-specific factors are not important cost drivers. Instead, it is because the benchmarking model is being influenced by differences in company costing methodology. As we illustrated in Figure 4, the network storage benchmark is demonstrably being set by a company that has taken a very simple approach to site-specific costing.

In any case, we would strongly reject the conclusion – which Ofwat appears to have drawn – that a simple unit cost model is already reflecting differences in site-specific costs in the presence of regional variation in exogenous factors, particularly where the benchmark company has very different concentrations of exogenous regional factors. Ofwat has not presented any evidence that suggests it has appropriately accounted for these factors when considering the reasonableness of its modelling approach.

Ofwat should not assume that inaccuracies will even out 'in-the-round'

Ofwat could take the view that it does not matter if companies have submitted costs based on simple equations because the inherent inaccuracy of this approach will broadly 'even out'. We do not agree with this approach for the following reasons:

- Ofwat has not presented any evidence to support its belief that top-down approaches for costing storage solutions are robust and will result in allowances sufficient to ensure programme delivery. For example, we have not seen evidence of any cross check to historical costs at schemes with similar exogenous site-specific characteristics.
- There is a strong risk that Ofwat is observing differences in forecasting approaches instead of differences in efficiency. We do not consider that this point is adequately recognised by Ofwat, especially given its view that the models capture site-specific cost drivers. Indeed, Ofwat has explicitly stated its view that the gap between UUW's business plan and the benchmark is indicative of 'inefficiency'.
- If most companies rely on a top-down method for developing costs but we use a bottom-up method that is site specific, the industry approach will not be appropriate for our costs in the case where we have a higher share of exogenous regional characteristics that drive higher costs. Section 3 of our main overflows representation document presents clear evidence this is the case.

One of the remedies Ofwat could consider is how to include outturn historical data and we welcome Ofwat's indication that it will consider this for the Final Determination. However, we caution that in doing this it is important that Ofwat recognises that distribution of historical costs is likely to be skewed towards lower cost sites as – consistent with the requirements of the SOAF process used at the time - the early sites tackled were usually those that revealed a high benefit-cost ratio. This will tend to bias historical investment towards relatively lower cost sites. It will also be important to control for site-specific exogenous factors when carrying out comparisons between historical and forecast costs.

We do not consider Ofwat's modelling approach to be robust

For the network schemes, Ofwat uses a log and a linear model with storage volumes as a single explanatory variable. It reasons that the log model captures economies of scale better while the linear model captures high fixed costs for smaller schemes. Ofwat averages the results from both models. For STW[7], Ofwat uses a linear model with cubic meters of storage as the single explanatory variable. This section sets out our views on Ofwat's proposed approach.

Ofwat's models do not adequately explain the variation in costs

Ofwat's models have a relatively low R-squared of around 60 percent. This is particularly low for a total cost model specification, which strongly suggests that there are explanatory variables missing. However, rather than working with the industry to try and understand this unexplained variation, Ofwat has made the very strong assumption that these differences can only be attributed to efficiency.

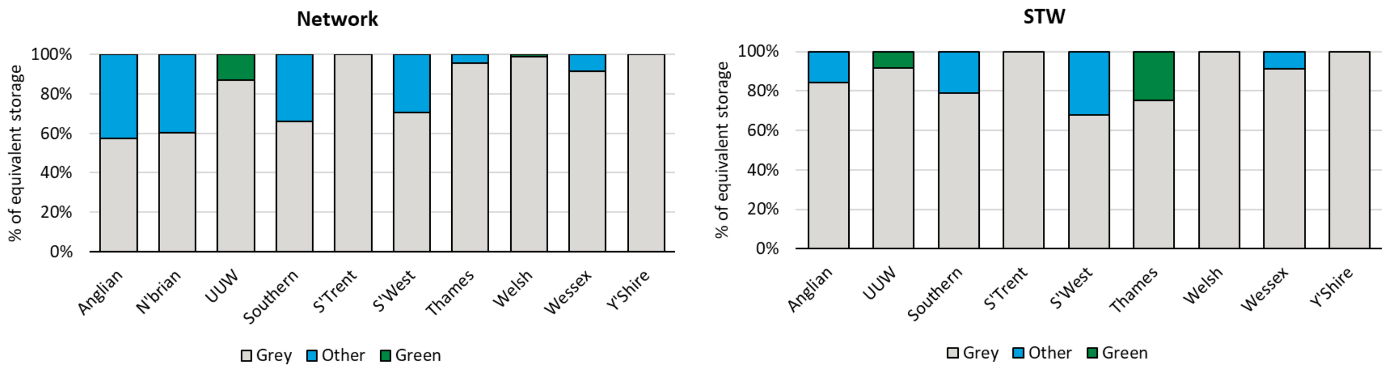
In particular, Ofwat has given insufficient weight to other credible cost drivers that could explain site-specific costs. It would be inappropriate to assume that the influence of these factors evens out 'in-the-round'. Without careful consideration and analysis, Ofwat cannot be confident that its modelling is identifying differences in efficiency between companies as opposed to differences in the distribution of these cost drivers. We present clear evidence of variation in exogenous cost drivers in section 3 of our main overflows representation document. Therefore, we consider there is a risk that a simple model is leading Ofwat to incorrectly assume that UUW's exogenous variation in cost is inefficiency.

Additionally, we can only observe the variation in costs that companies have reflected in their simplistic forecasts. In this context, the low R-squared is particularly surprising but also likely to be understated - the dispersion in costs is likely to be much higher than that implied by top-down cost curves. Using cubic metres of storage only is therefore likely to have even less explanatory power than suggested by the models.

Ofwat's failure to distinguish between grey and hybrid storage will drive worse environmental outcomes

A hybrid solution uses a mix of grey storage and blue-green infrastructure to deliver spill reductions. There are clear environmental and societal benefits from implementing blue-green solutions where possible. However, Ofwat's DD approach uses a single volumetric cost driver, which reflects storage delivered through both grey and hybrid solutions. This implicitly assumes that all companies are delivering the industry average mix of grey and hybrid solutions. Business plan data suggests that this is not the case, as illustrated in Figure 6.

Figure 6: The industry is demonstrably delivering storage through different mixes of grey, green and other storage at hybrid schemes



Source: UUW analysis of Ofwat's dataset

It is worth reflecting on the incentive implications of this methodological decision:

[7] Sewage treatment works (STW), also known as wastewater treatment works (WwTW).

- A company that wishes to implement above average levels of hybrid solutions will be underfunded by the simplistic unit cost. This creates a clear disincentive to identify and implement additional blue-green schemes, even where it is feasible and would drive additional environmental and societal benefits.
- Conversely a company is incentivised to outperform against Ofwat's simplistic unit cost by reducing the number of hybrid schemes.

We consider that this could be characterised as a perverse incentive, which will drive poorer environmental outcomes. We also consider that this incentive is inconsistent with Defra's statement that: *"…The Environment Agency and Ofwat will work to ensure assessment processes promote and incentivise the use of nature-based solution in favour of more carbon intensive alternatives."*[8]

Companies have proposed a wide range of solution types. The optimal mix for each company will be different and will depend on a range of factors, including whether blue-green infrastructure is feasible at a specific site. As such, it appears inappropriate to assume that all companies have equivalent opportunities to deliver blue-green solutions. However, this is the effect of Ofwat's simplistic modelling approach.

Ofwat appeared to justify its approach to assessing hybrid schemes in response to OFW-IBQ-UUW-08. It said:

"As can be seen in table 1 below, prior to removing outliers, the majority of company costs where [sic] reasonably comparable between grey only and hybrid schemes, whereas green only schemes were significantly higher. This led to the decision to combine grey only and hybrid solutions into one model."

We are concerned Ofwat's conclusion that grey only and hybrid schemes are comparable in cost is inappropriate. Ofwat has accepted that green infrastructure is associated with additional costs. Therefore, simple intuition would suggest that a hybrid scheme – which includes both grey and green elements – is more expensive than a grey only scheme.

Additionally, we are concerned that data consistency issues have led Ofwat to an inappropriate conclusion. We have replicated Ofwat's table 1[9] in Table 2 below. This shows that three companies consider that hybrid schemes are cheaper to deliver than grey only schemes. We have highlighted these instances in red. It appears counter intuitive and contrary to engineering rationale for hybrid schemes to cost less than grey schemes. This appears to contradict Ofwat's approach to green-only schemes, which it accepts costs more than grey-only schemes.

We have edited Ofwat's analysis to provide a view on what the median hybrid unit cost would be if companies with counter-intuitive unit costs were removed. This demonstrates a clear increase in the cost of delivering hybrid schemes relative to grey only. If Ofwat considers it is inappropriate for us to remove ANH, NES and SVE from the sample, then it should also explain why it is consistent to expect green schemes to cost more than grey schemes, but hybrid schemes to cost less as these companies suggest.

---

[8] Defra (2021) *Storm Overflows Discharge Reduction Plan*. Available here: https://assets.publishing.service.gov.uk/media/631063778fa8f5448a3836e4/Storm_Overflows_Discharge_Reduction_Plan.pdf

[9] OFW-IBQ-UUW-08

Table 2: Replication of table 1 in Ofwat's response to OFW-IBQ-UUW-008

Company	Total	Only grey	Only green	Hybrid	% difference
ANH	1,758	1,796		1,480	82%
NES	5,201	4,490		2,463	55%
SWW					
SRN	3,749	2,571	22,996	3,316	129%
SVE	2,343	1,956	8,169	1,671	85%
TMS	2,892	2,980			
UUW	4,876	4,300	4,185	7,526	175%
WSH	1,394	1,376	9,108	1,511	110%
WSX	3,080	3,551	1,538		
YKY	4,887	4,607		5,476	119%
Total	3,353	3,070	9,199	3,349	109%
Median	3,080	2,980	8,169	2,463	83%
UUW edit – median excluding companies with counterintuitive hybrid unit costs					
Median excl. ANH, NES, SVE	3,415	3,266	6,647	4,396	**135%**

Source: OFW-IBQ-UUW-008

As such, we do not consider Ofwat's decision to ignore the higher efficient costs associated with hybrid schemes to be legitimate. We would support the reflection of higher costs of a hybrid scheme in its FD approach.

Ofwat's approach to outliers is not robust

Given the limited dispersion of data observed in the sample driven by top-down cost equations, excluding outliers means that Ofwat is removing specific schemes that may contain important information about what drives costs. While these schemes may appear as outliers when only using cubic meters as the single explanatory variable, **they may be explained by other explanatory variables**. For example, companies may have used top-down cost equations for most schemes but for schemes with specific characteristics a more detailed approach may have been used. By removing such outliers Ofwat misses the opportunity to understand what drives costs. It also reduces the information available to the model that could inform a robust benchmark. As we set out in section 5.3 of our main overflows representation document 'UUWR_10_Overflows', this appears to be supported by empirical evidence – supplementing Ofwat's simple model with exogenous variables reduces the number of outliers.

In its response to OFW-IBQ-UUW-008, Ofwat says:

"…The storm overflow model includes a large number of schemes, so the model reflects the characteristics of an average scheme. Some schemes may be more complex and others less complex than the average scheme."

We note that Ofwat's approach to outliers means this not strictly true. The model does not reflect the costs of an average scheme. Instead, it reflects the costs of an average scheme **once outliers have been removed**. Crucially, these schemes are considered outliers by a simple model with one variable – they may not be outliers but instead might be legitimate high-cost schemes and would be recognised as such by a more robust, better specified model. As such, it is not clear to us that Ofwat's model is an appropriate benchmark for high-cost (or low-cost) schemes.

Once the outliers are removed, Ofwat's response to evidence of site-specific variation in efficient cost (e.g. contaminated land, rail infrastructure, etc) is to state that these factors are already captured by the models. However, this appears to be a value judgement rather than one based on objective evidence. We consider that the evidence of regional variation of exogenous factors and data issues (presented in our main overflows representation document) demonstrates that these factors are not appropriately reflected in the benchmark.

We are clear that Ofwat has not presented the data that would enable it to be confident that its simple model is appropriately reflecting these exogenous factors and therefore, differences in the cost of efficient delivery. Therefore, in its assessment of outliers, we consider it is important for Ofwat to be clear about the limitation of using simplistic models and reflect this as a mitigating factor in its assessment. Ofwat should recognise and address the obvious point that, by definition, for outliers that have been removed from the modelling the model predictions are irrelevant.

Finally, Ofwat's approach is to remove the same set of outliers from both the log and linear model. While we understand why Ofwat has chosen to do this, as it simplifies the approach for setting the costs for the outliers, the problem is that it undermines the modelling for the rest of the scheme. From a modelling perspective removing outliers identified from the log model from the linear model and vice versa is not appropriate, for the simple reason that an outlier in the log model outlier will not necessarily be, and often will not be, an outlier in the linear model.

The choice of efficiency benchmark is not justified

Ofwat uses the upper-quartile as the industry benchmark for network schemes and the median for STW schemes. The reasons for this choice are not transparent. Ofwat states that the upper quartile is chosen in line with the mid-range of the unit cost benchmarks considered but does not provide a list of these benchmarks. We suspect that Ofwat is referring to PR19 forecast data but it is not clear whether Ofwat has also considered any outturn cost data. It also isn't clear whether Ofwat controlled for the exogenous characteristics of schemes when it made this assessment. If it didn't, then we would question the validity of any subsequent conclusions.

We also consider that the explanatory power of the model is too low to support an upper quartile benchmark. For STW, Ofwat states that the median is appropriate as it is in line with engineering judgement that costs should be lower than network schemes. However, details on the engineering judgement are missing. Ofwat appears to have selected the upper quartile benchmark for network purely because the median would have been higher than the median for STW without any regard for the explanatory power of the model, the relevance of historical outturn costs or the presence of exogenous factors that might lead efficient costs to vary across the industry. As such, we do not consider that the efficiency benchmarks are justified.

Ofwat's approach to FTFT schemes is simplistic

Flow to full treatment (also known as Pass Forward Flow) schemes will be increasingly required in AMP8. This is because the volume of storage required to deliver ten spills is substantial and as such, is not always feasible to build and/or operate. However, we consider Ofwat's approach to FTFT schemes to be simplistic.

We have strong concerns that exogenous factors (set out in section 3 of our main overflows representation document) will also lead to higher FTFT costs. As such, the efficiency challenge that Ofwat is currently applying against our FTFT costs is inappropriately high. We do not consider that Ofwat's resulting benchmark represents a realistic cost forecast.

We note that Ofwat is collecting additional data through the DD submissions to support a more robust approach to FTFT cost assessment. Ofwat should bear in mind the relevant exogenous factors when considering company submissions.

We continue to consider that companies have included equivalent volumes relating to FTFT schemes in their cost driver data. We believe it is important for Ofwat to appropriately investigate whether this is a risk because this will likely lead to inconsistent cost assessment and an unrealistic benchmark.

UUW has developed models that better reflect the underlying engineering rationale that determines costs at storage solutions

UUW has been able to develop a subset of these exogenous factors into variables suitable for use in an econometric model. We did this by cross-referencing Ofwat's dataset with WaterUK's National Storm Overflow Plan dataset. WaterUK's data includes geographical identifiers, meaning we were supplement Ofwat's dataset with additional site-specific variables that are informed by engineering rationale.

We found that these variables performed well and generally improved model fit across a number of different measures. We also found the number of outliers identified by the Cook's Distance test reduced, which is demonstrable evidence that Ofwat's simple model is not recognising underlying engineering rationale that would support a different efficient benchmark.

Full details can be found in *'UUWR_10_Overflows'*.

Phosphorus removal

Question 4.1) Do you agree with our approach to assessing phosphorus removal enhancement costs?

We have identified several issues with Ofwat's approach, which are set out in this section.

Ofwat's approach appears to systematically underfund phosphorus removal

In its DD, Ofwat notes that its approach:

"…Leads to a strong efficiency challenge to business plans of 19% at the sector level."[10]

We are clear that this represents an unrealistic challenge and this representation sets out the reasons for this. However, we consider that the sector-level efficiency challenge quoted by Ofwat masks some company-level issues. In particular, we consider that one low-cost company appears to be masking the level of efficiency that Ofwat is requiring across all other companies in the delivery of a statutory obligation.

There is a risk that Ofwat is understating the true efficiency challenge at an industry level

Table 3 sets out the gap between Ofwat's models and company business plan requests (excluding those schemes that are assessed as outliers). It reveals that most companies are subject to a substantial cost challenge across their phosphorus removal programmes. However, it also reveals that Anglian Water is targeting costs that are substantially below the industry benchmark. This effectively masks the true efficiency challenge that Ofwat is imposing upon the industry.

Table 3: Gap to Ofwat's models by company (outliers excluded)

Company	Cost challenge (£m)	Percentage challenge
ANH	413	85%
HDD	0	-1%
NES	9	43%
SRN	-71	-16%
SVE	-146	-24%
SWB	-10	-9%
TMS	-516	-41%
NWT	-208	-43%
WSH	32	21%
WSX	-422	-45%
YKY	-2	-1%
Total	**-922**	**-19%**
Total excluding ANH	**-1,335**	**-30%**

Source: UUW analysis

[10] Ofwat (2024) *Draft Determinations: expenditure allowances.*

Given this, we consider that it is reasonable to ask whether this cost challenge is realistic. One company is a materially low-cost outlier, while most others face substantial cost challenges. Excluding Anglian, the total industry cost challenge is £1.3bn. We do not consider that this would appear reasonable to an outside observer, particularly when the programme in question is statutory. We consider that there is considerable scope to better align cost assessment to the pressures and challenges that companies are attempting to navigate in AMP8.

This representation sets out evidence to support our case that Ofwat's chosen set of models are less able to explain the costs the industry in general, and UUW in particular, can expect to efficiently incur in the delivery of their AMP8 WINEP.

AMP8 cost pressures are not appropriately reflected in AMP7-based cost forecasts

Our Future Ideas Lab paper highlighted the need for cost assessment to ensure that it was forward-looking and capable of accounting for a different relationship between cost and cost drivers in future periods.[11]

We consider that there is a fundamental disconnect between the AMP7 and AMP8 phosphorus removal programme. Ofwat's proposed approach assumes both AMP7 and AMP8 costs are equally capable of reflecting the costs companies will incur in AMP8. This is because it applies equal weight to forward-looking and backward-looking econometric models. However, this section will provide clear evidence that AMP7 costs will be poor predictors of AMP8 costs. As such, UUW considers that more weight should be placed upon forward-looking, AMP8 costs within cost assessment.

Table 4 sets out the number of schemes included within Ofwat's DD phosphorus removal model, across the entire industry. There is a substantial increase in the number of smaller treatment works. We consider this is clear evidence that the AMP8 programme is different to AMP7.

Table 4: The industry will focus upon much smaller treatment works in AMP8 relative to AMP7

	AMP7	AMP8	Percentage difference
Size band 1	26	93	258%
Size band 2	66	120	82%
Size band 3	215	281	31%
Size band 4	242	258	7%
Size band 5	113	114	1%
Size band 6	123	127	3%
Total	**785**	**993**	**26%**

Source: UUW analysis of Ofwat's phosphorus removal model (PR24-DD-WW-p-removal.xlsx)

In additional to working at smaller sites, the industry is targeting much more stringent permit limits, towards the technically achievable level of 0.25mg/l.

The difference is illustrated in Figure 7, which tabulates the inputs to Ofwat's backward-looking AMP7 and forward-looking AMP8 model by i) consent limit; and ii) treatment works size band. Shades of red indicate a higher number of sites within that category – the colour scale is consistent across both AMP7 and AMP8 data. The higher concentration of red towards the upper left corner in AMP8 illustrates that companies are carrying out more stringent work at very small sites in AMP8 relative to AMP7. There is also a higher number of works targeting the technically achievable limit across all size bands.

[11] UUW (2021) *The Principles of Regulatory Cost Assessment.* Available here: https://www.unitedutilities.com/globalassets/documents/pdf/the-principles-of-regulatory-cost-assessment.pdf

Figure 7: The industry is targeting substantially more phosphorus removal at smaller sites in AMP8 relative to AMP7

AMP7	Size band 1	Size band 2	Size band 3	Size band 4	Size band 5	Size band 6	Total
0.25mg/l and less	3	4	13	46	33	38	**137**
0.25mg/l to 0.5mg/l	4	13	56	100	46	40	**259**
0.5mg/l to 1mg/l	3	27	87	74	17	35	**243**
1mg/l to 2mg/l	9	12	41	12	16	10	**100**
2mg/l to 3mg/l	2	5	12	4	0	0	**23**
3mg/l to 4 mg/l	3	3	4	1	1	0	**12**
4mg/l to 5mg/l	2	2	2	3	0	0	**9**
Above 5mg/l	0	0	0	2	0	0	**2**
Total	**26**	**66**	**215**	**242**	**113**	**123**	**785**

AMP8	Size band 1	Size band 2	Size band 3	Size band 4	Size band 5	Size band 6	Total
0.25mg/l and less	10	32	91	148	74	101	**456**
0.25mg/l to 0.5mg/l	10	21	59	47	15	14	**166**
0.5mg/l to 1mg/l	23	35	63	42	15	10	**188**
1mg/l to 2mg/l	24	21	36	7	10	2	**100**
2mg/l to 3mg/l	10	3	8	2	0	0	**23**
3mg/l to 4 mg/l	9	3	12	5	0	0	**29**
4mg/l to 5mg/l	7	5	12	5	0	0	**29**
Above 5mg/l	0	0	0	2	0	0	**2**
Total	**93**	**120**	**281**	**258**	**114**	**127**	**993**

Source: UUW analysis of Ofwat's dataset

Engineering rationale and evidence suggests that targeting stringent permits at smaller sites will drive higher efficient costs, all else equal. We set out this engineering evidence in the remainder of the section. We then show that backwards-looking AMP7 econometric models appear to have a poorer fit against AMP8 costs, which is evidence that AMP7 models are less capable of reflecting ex-ante engineering rationale.

Why small treatment works are associated with higher capital costs

Sites of less than 2,000 population equivalent (i.e. bands 1-3) are considered small works and tend to be located in rural areas. These small works tend to have relatively simple treatment processes such as trickling filters, which are incapable of delivering performance in line with more stringent permits. We set out some reasons why costs at small works are expected to be higher, all else equal:

- Trickling filter works with a new phosphorus permit of less than 1 mg/l require tertiary solids removal, which is a step-change in complexity from the trickling filter technology currently used. Installing tertiary solids removal adds substantially to the cost of the scheme.
- Small works often treat a high multiple of dry weather flow and so have a high maximum flow relative to the minimum flows treated at site;
- Small works often have higher rates of infiltration due to nature of the catchment which also leads to treatment of high maximum flows. This is exacerbated in the Lake District which experiences higher rainfall and infiltration rates;
- Peaking factors at small works can be high due to short networks;
- Access is often poor, which means additional costs will result from improving transport infrastructure which will allow the delivery of chemicals and tankering of additional sludge generated when dosing ferric;

- Land on site is constrained and so often an extension of the WwTW is required which requires land purchase;
- Power upgrades are frequently required as many of these works have limited or no current power;
- Sites often have descriptive consents only and so require a major upgrade to meet new numerical permits;
- Many of these sites are within the Lake District National Park and so will need addition features such as environmental impact assessments, use of sensitive materials, sensitive landscaping etc;
- Sites are often close to watercourses which means flooding can be prevalent which will need remediation;
- The sites often have no dosing assets and so will require all dosing ancillary equipment such as dosing, storage, blind tanks, delivery areas, sludge storage etc. In addition these sites often do not have potable water which is required for dosing and bringing this to site adds to the costs; and,
- Sites in the Lakes have high tourist populations that must be accounted for in design.

AMP7 models are objectively poorer at reflecting AMP8 cost pressures

As the previous sections have evidenced, the AMP8 WINEP is requiring interventions at small works with stringent permit limits. While there were some examples of such activity during AMP7, we have evidenced that AMP8 represents a fundamental step-change, with stringent permit limits required at a larger number of works, and in particular at the smallest treatment works.

This is important because this means that the historical and forecast datasets are fundamentally different. This difference was highlighted in Figure 7, which illustrate that the AMP8 dataset contains more information that will allow the model to establish a robust relationship between costs and:

(a) stringent permits at the technically achievable level (TAL); and,

(b) stringent permits at the TAL at very small works.

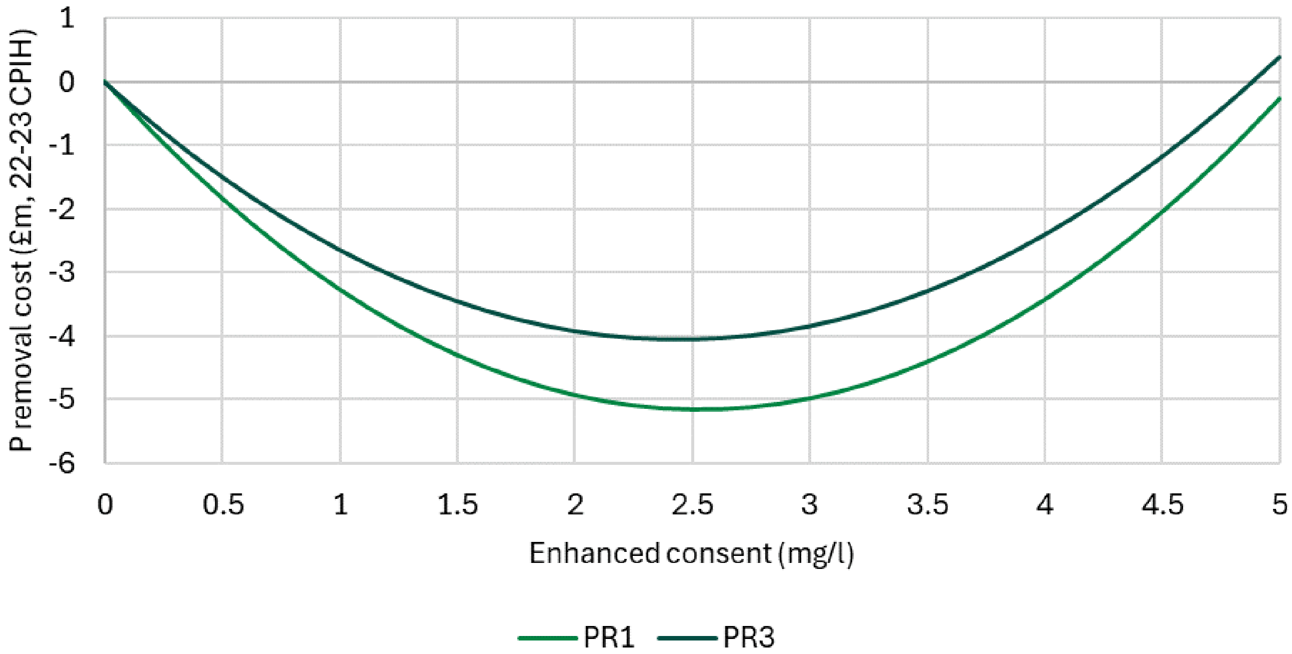
The lack of similar information in the historical dataset evidenced in Figure 7 means that backwards-looking models will struggle to form a similarly robust relationship between cost and these cost drivers. This means they will tend to under-represent the costs companies will incur in the delivery of their AMP8 programmes.

We consider that there is clear evidence of this within the DD model suite.

The backwards-looking models suggest that tighter permits have less of an effect on costs, relative to forwards-looking models. Figure 8 shows how Ofwat’s models PR1 and PR3 models expect P removal costs to vary as the enhanced consent becomes increasingly stringent. PR1 uses forward-looking (AMP8) data while PR3 uses backward-looking (AMP7) data. These models use a squared enhanced permit variable to reflect how costs are expected to exponentially increase as the permit gets increasingly stringent.

It’s clear that PR1 expects permit stringency to have a consistently larger impact on costs – PR1’s curve is below that of PR3 at all points. Importantly, this means that PR1’s slope is steeper as the permit approaches zero. This shows that PR1 expects permit stringency to have a more pronounced effect on costs than indicated by PR3. We consider this suggests that PR3 is understating the costs associated with achieving permits at the TAL. As evidenced in Figure 7, PR3 has less data available to it to help it establish the relationship between cost and permit stringency as permit stringency approaches the TAL.

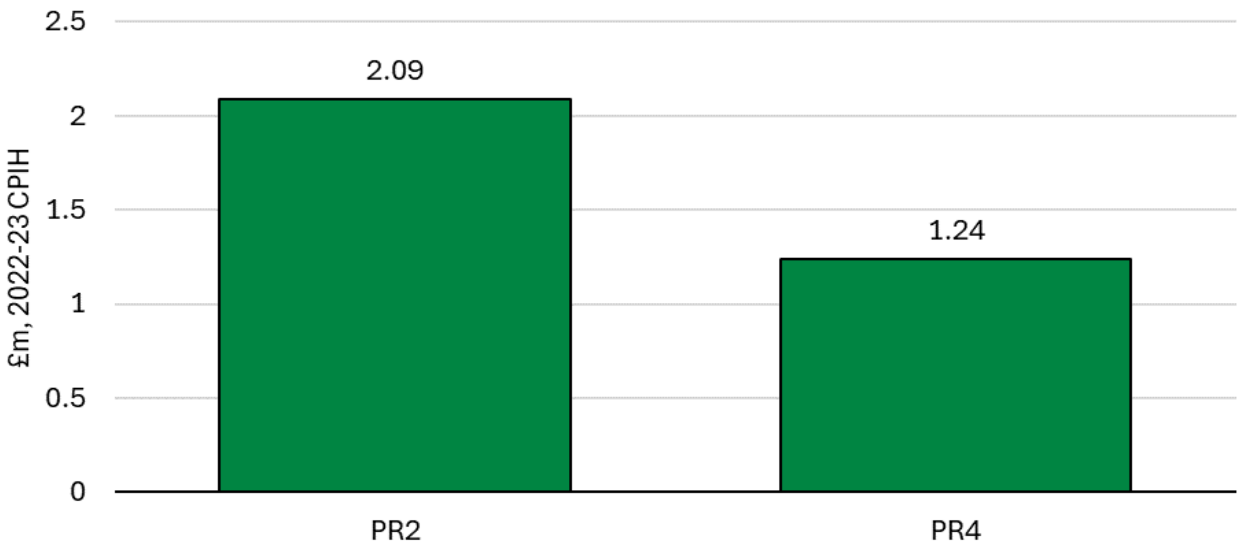
Figure 8: The relationship between P removal costs and the stringency of the enhanced consent within Ofwat's PR1 and PR3 models



Source: UUW analysis of Ofwat dataset.

We also find a similar difference between PR2 and PR4 models. These models use a dummy variable to reflect sites where the permit is at or below the TAL. Again, there is a clear difference between the forward-looking and backward-looking model predictions. The forward-looking model suggests a much stronger relationship between a TAL permit and associated scheme costs – the coefficient is 69% higher in PR2 relative to PR4. This is illustrated in Figure 9.

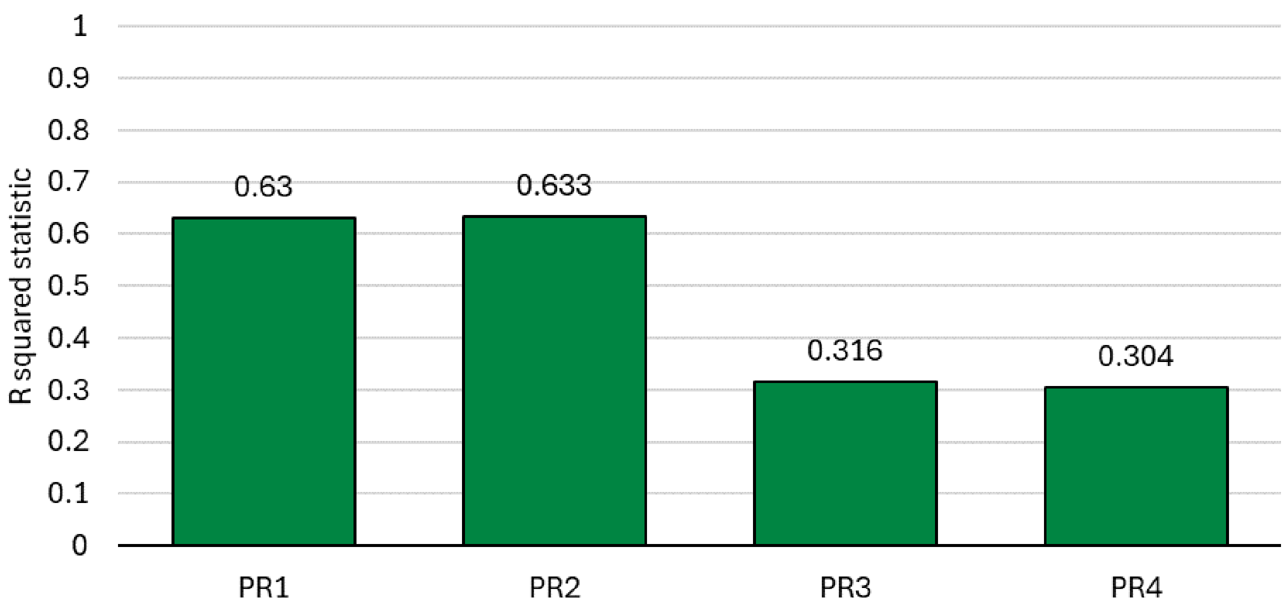
Figure 9: The relationship between P removal costs and the Technically Achievable Limit (TAL) dummy within Ofwat's PR2 and PR4 models



Source: UUW analysis of Ofwat dataset.

Finally, there is a clear reduction in the model fit (denoted by R-squared) for the backwards-looking models. This is illustrated in Figure 10. This suggests that the cost drivers chosen to explain AMP8 costs (e.g. a squared term to capture the exponential increase in cost as permit stringency increases) are less adept at explaining AMP7 costs.

Figure 10: Model fit in Ofwat's backwards-looking models (PR3 and PR4) is objectively poorer



Source: UUW analysis of Ofwat dataset.

We note that the R-squared values in Figure 10 relate to each specific model, meaning that the R-squared value for PR3 and PR4 sets out the explanatory power of the model relating to historical costs. However, we are primarily interested in the power of the model to explain AMP8 costs. As such, we have also calculated the R-squared between forecast costs and the historical models. This can be interpreted as informing to what extent historical P removal models PR3 and PR4 explain variation in forecast costs. This is set out in Table 5. It is clear that the explanatory power of models PR3 and PR4 is low for AMP8 costs.

Table 5: Explanatory power of models RP3 and RP4 relating to AMP8 costs

	RP3	RP4
R squared between forecast costs and modelled costs	0.233	0.229

Source: UUW analysis

These findings are important because they demonstrate that the more powerful treatment complexity coefficient in forecast AMP8 models is not due to companies submitting inefficient cost forecasts. Instead, **it reveals that the variables Ofwat has selected to reflect AMP8 cost pressures are objectively poorer predictors of AMP7 costs**. As highlighted in Figure 7, the AMP7 and AMP8 programmes are fundamentally different. As such, it is not surprising that AMP7 data is less able to support the relationship between cost and cost driver that engineering rationale suggests will exist in AMP8.

We consider that this is compelling evidence to suggest that backwards-looking models are poorer at predicting the costs of the AMP8 P removal programme, in particular the costs of meeting permits at the TAL at all works, but also particularly at small works.

We do not consider it is plausible that companies have collectively influenced the cost benchmark

In its DD, Ofwat considers several potential reasons that may explain why AMP8 costs are higher than AMP7:

- Companies may have different risk appetites;
- Companies have submitted higher business plan cost forecasts;
- PR24 WINEP/NEP programme is much larger than at PR19;
- Data reporting issues; and
- Prevalence of tighter permits.

In our view, Ofwat has not appropriately considered the key factor driving differences in cost estimates between the forward-looking and backward-looking approaches – the inability of AMP7 cost data to reflect efficient cost variation in delivering stringent consents at small treatment works (as discussed in the previous section).

However, we do want to address Ofwat's concern that companies may have submitted higher cost forecasts to:

"…Attempt to obtain a higher allowance under the assumption we will use these costs to set efficient cost allowances."[12]

We would like to make clear that UUW's cost estimates are motivated solely by UUW's view of the efficient level of cost required to deliver our obligations. We reject any statement that suggests our cost estimates are the result of any attempt to influence the process by which Ofwat identifies the efficient cost benchmark.

We note that for Ofwat's statement to be true, **the following conditions would need to hold simultaneously at an industry level**:

- **The BPA provides weak ex ante incentives.** We note that Ofwat's methodology creates powerful incentives for companies to submit efficient costs, through the Business Plan Assessment (BPA) process. Ofwat does not recognise or acknowledge this incentive when it suggests companies may have submitted inefficient phosphorus removal costs. The implication of Ofwat's statement is that the incentives created by the BPA process are insufficient to offset the incentive for companies to attempt to influence cost assessment with inefficient cost forecasts. Ofwat may wish to consider the implications of this statement further, particularly given the widespread use of forecast costs in PR24 cost assessment. In any case, we are clear that the BPA represents a strong incentive to provide efficient cost forecasts.
- **A majority of companies are complicit in submitting inefficient cost forecasts.** Ofwat's methodology compares the costs of different companies to identify an efficient benchmark. All other companies are then expected to deliver in line with the benchmark. This process is explicitly designed to protect customers from inefficient costs. Countering the incentive structure of the benchmarking framework would require a substantial number of companies to independently consider it a worthwhile strategy to submit inefficient costs. However, the *'prisoner's dilemma'* nature of the process creates a clear disincentive against this type of action. We have not seen any evidence that would lead us to suspect a majority of companies are actively targeting a poorer BPA score in return for taking the risk their inefficient cost forecasts are rewarded through cost assessment.
- **Ofwat's models have captured all sources of regional and scheme-level variation in costs.** Ofwat notes that it tested company dummy variables in its model suite. It took the fact that these dummy variables are higher in its forward-looking models as evidence that submitted AMP8 cost forecasts are inefficient. However, Ofwat doesn't acknowledge other reasons that might cause company-specific effects in AMP8 to be higher. For example, companies that have had small capital programmes historically may have to rapidly scale up their capital delivery departments to deliver a much larger AMP8 programme. This would lead to a higher fixed cost estimate for that company, which would be reflected in the dummy variable. Another example could be the effect of regional exogenous factors – UUW's phosphorus removal programme includes a larger number of works in the Lake District, relative to AMP7. These small, rural sites are associated with higher fixed costs and as such, the UUW-specific dummy would likely be higher when estimated using AMP8 data. Ofwat has not presented any evidence to suggest that relative efficiency is the only driver of scheme-level variation its models do not capture. Therefore, we do not consider it is appropriate for Ofwat to cite its dummy variable tests as evidence of relative inefficiency in AMP8 delivery.

As stated above, all these factors must be true if Ofwat's suspicion of inefficient cost forecasts is appropriate. We would ask Ofwat to consider how likely this is. If Ofwat considers this is likely, then we look forward to engaging with Ofwat following PR24 as part of the Future Ideas Lab process to consider what improvements can be made to the regulatory framework to mitigate such concerns in the future.

[12] Ofwat (2024) *Expenditure allowances – enhancement cost modelling appendix.*

As a result of the arguments set out in this section, we consider it would be legitimate to drop the AMP7 models entirely. However, we recognise the merits of assessing using outturn data. While we disagree that there is risk that companies are choosing to submit inefficient costs to influence benchmarking, we do consider that outturn costs can provide a legitimate alternative view on efficient delivery. On balance, we would place more weight on AMP8 data because AMP7 data include less evidence on delivery to stringent consents and in particular at small sites, as evidenced in this section. **Therefore, we consider a pragmatic compromise would place 75 percent weight on AMP8 models and 25 percent weight on AMP7 models.**

Question 4.2) Do you agree with our approach to addressing the implementation issues associated with modelling phosphorus removal enhancement costs?

General issues with design of modelling

We have observed the modelling methodology for this area is hard to follow back to source data due to reliance on the use of pivot tables in the 'CWW19 input-CWW3' tab, which predominantly feeds the 'reconciliation adj factor' tab. It can be difficult to clearly ascertain where data has come from and what each section of the table is representing. This adds unnecessary complexity in assessing the validity of this approach. We encourage Ofwat to develop a more simple spreadsheet design with clear signposting to allow easier interrogation of this section of the model.

We consider that data should have been captured at a scheme level for 2023-24 for Phosphorus removal schemes in CWW19. Capturing data in this way at scheme level would have ensured alignment with the consolidated totex figures in CWW3/12/17, removing unnecessary complexity in the adjustments Ofwat is subsequently required to make.

Issues with the adjustment factor

We observe that Ofwat has made an adjustment to exclude costs to be incurred after 2029-30 from efficient allowances. We understand Ofwat's approach in using full scheme data, including post 2029-30 costs, to assess scheme efficiency. However, Ofwat does not set out how or when these costs will be reflected within cost assessment. We consider that Ofwat must clarify its approach to setting allowances for post 2029-30 costs, relating to AMP8 schemes.

We also note that Ofwat makes an adjustment to provide an allowance for schemes dropped from the modelling dataset due to data quality issues. We disagree with this methodology in certain scenarios. Schemes dropped from the dataset due to company data issues, particularly those which are specifically related to scheme complexity, e.g. permit level, should be penalised rather than effectively rewarded with an uplift proportional to the company modelled allowance. This approach appears inconsistent with how Ofwat have assessed similar situations in other cost areas – such as Growth at STWs. In this enhancement area, Ofwat has dropped schemes from the modelling dataset due to data quality issues, which simply receive no allowance.

There are various schemes in this area which demonstrate a counter-intuitive phosphorus permit level change. These schemes are implicitly awarded an allowance broadly in line with the requested totex, with only the general company efficiency across all P schemes applied. We do not consider this to be appropriate, as the assumption these schemes are as efficient as the company's overall programme is speculative. This mechanism means some companies are effectively rewarded for poor data, rather than penalised for it.

We would encourage Ofwat to assess material schemes of this nature via a deep dive process rather than allowing a simple uplift based on the average efficiency of all other schemes.

Growth at sewage treatment works

Question 5.1) Do you agree with our approach to identifying overlap with base costs so that customers do not pay for non-compliance with existing permits?

We agree that customers should not pay twice to achieve compliance with existing permits. Ofwat's proposed base overlap scenario evaluation framework[13] appears to rely on appropriate criteria to identify overlap risk. Ofwat states that 16 schemes have been identified as having a risk of base overlap.[14] However, within Ofwat's models and the accompanying documentation, we have been unable to identify which specific schemes (and for what reason) have been excluded. We were therefore unable to assess whether the base overlap assessment has been implemented appropriately.

Question 5.2) Do you agree with the models we have selected to explain differences in efficient growth at STWs enhancement costs?

Ofwat's proposed modelling suite for Growth at STWs produces sensible results at a programme level. However, we consider that the approach to setting efficient allowances could be further improved, as described below.

Ofwat's DD models contain two scale variables ("Added Process Capacity in PE" and "Expected Change in PE") which are triangulated across two models. Both models contain the "change in DWF permit" variable – this variable is also highly correlated with the FFT permit variable, meaning that it functions as a proxy for both DWF and FFT permit changes.

However, we are concerned that Ofwat's models include only one variable (ammonia) to measure effluent quality permits. The ammonia permit dummy variable can be interpreted as a proxy for sanitary parameter tightness at the STW. However, there is no equivalent variable included that would measure tightness of nutrient permit level. We consider that sanitary parameters and nutrient permits are distinct cost drivers – indeed, this is reflected in Ofwat's enhancement cost assessment framework where they are assessed separately. This means that the inclusion of only one effluent quality variable will likely lead to biased model predictions.

Ofwat claims that there is high correlation between different measures of treatment complexity. However, we find little evidence of this being the case. As can be seen in Table 6, there is only strong correlation between BOD and suspended solids (0.82), in addition to the aforementioned correlation between FFT and DWF. The picture remains broadly the same when these variables are converted to dummies. Therefore, there is only weak association between other complexity metrics, indicating that each is measuring a distinct aspect of effluent quality.

Table 6: Growth at STWs variable pairwise correlation matrix

	DWF	FFT	BOD	Suspended solids	Phosphorus	Ammonia
DWF	1.00					
FFT	0.96	1.00				
BOD	0.11	0.01	1.00			
Suspended solids	-0.03	0.01	0.82	1.00		
Phosphorus	-0.11	-0.11	0.17	0.23	1.00	
Ammonia	0.16	0.28	0.49	0.51	0.24	1.00

Source: UUW analysis of "PR24-DD-WW-Growth-at-STWs-2.xlsx" model data

[13] Ofwat (2024) *PR24 Draft Determinations: Expenditure allowances - Enhancement cost modelling appendix*, p. 54-55.

[14] Ibid.

Based on the above we consider that model quality could be improved by including an expected phosphorus permit (<=0.5 mg/l) dummy.[15] Tight phosphorus permits is a major driver of treatment process costs which are largely distinct from sanitary parameters-related expenditure. Phosphorus removal mostly relies on solid separation (e.g. using settlement tanks), while ammonia removal utilities biological processes (e.g. trickling filters). This means that tighter ammonia and phosphorus permits can lead to distinct and non-overlapping cost pressures. We find that the phosphorus permit dummy is statistically significant and of the correct sign, as shown in Table 7. We propose including the phosphorus dummy separately from ammonia, and triangulating across the 4 models to set the final allowance.

Table 7: Growth at STWs models (excluding outliers) with UUW's proposed additional models

Explanatory variable	Ofwat's DD model specification[16]		UUW additional model specification	
Added Process Capacity in PE	187.08***		183.98***	
Change in DWF permit	6661.48***	6793.74***	6734.32***	6676.41***
Change in PE		913.91***		1063.30***
Ammonia dummy (<3 mg/l)	6209211***	5505801***		
Phosphorus dummy (<= 0.5 mg/l)			2872280***	2837803***
Constant	3052513***	3078328***	3195270***	3112482***
Adjusted R-squared	0.45	0.43	0.42	0.41
Observations	200	200	200	200

Source: UUW analysis

In line with Ofwat's guidance, costs have been proportionally allocated across relevant tables to separately represent expenditure associated with new WINEP requirements and growth demands. As a result, the inclusion of the new dummy variable will not cause a double-count in allowances across phosphorus removal schemes and Growth at STWs schemes. We therefore consider that the inclusion of the phosphorus dummy is a valid way to recognise additional expenditure related to nutrient removal requirements.

Question 5.3) Do you agree with our approach to adjusting modelled allowances to account for costs incurred outside of the 2025-30 period?

To account for costs incurred outside of AMP8, Ofwat used an approach where the share of 2025-30 costs within company forecast totex is applied to modelled allowances. This is a relatively intuitive and transparent approach.

However, we are concerned that there is potential misalignment between cost and cost driver data at the econometric modelling stage. Specifically, Ofwat's dependent cost variable includes post-AMP8 expenditure, while the independent variable "Change in PE" only accounts for growth in the period 2025-30. UUW's submitted cost figures, for example, relate to expenditure to meet the long-term design standard, rather than being based exclusively on AMP8 population growth estimates. Furthermore, we observe that there is almost £400m of expenditure associated with "After 2029-30" capex across the industry. Therefore, the fact that the "Change in PE" only captures AMP8 population growth means that a large share of costs will likely not be accurately explained by Ofwat's models.

[15] Engineering rationale suggests that there is a step-change in costs when 0.5 mg/l limit is reached. This threshold is also consistent with the approach taken when assessing phosphorus removal expenditure at PR19.

[16] Ofwat's model specification coefficients are slightly different from those published at DD; this is because the inclusion of the phosphorus dummy results in the identification of one additional outlier scheme.

We find that adjusting the PE variable to include post 2029-30 data leads to substantial changes in model coefficients, as shown in Table 8. It can be seen that in Ofwat's original DD models there is a large difference in coefficient magnitude between "Added Process Capacity in PE" and "Change in PE" variables (202.9 vs 1032.6) – this is counterintuitive, given that both variables should be measuring the same underlying growth pressures. However, when the PE growth period is extended, the coefficients of these two variables become much more closely aligned (268.8 vs 384.6), indicating that both variables are now proxying similar population growth factors.

Table 8: Growth at STWs models where PE includes after 2029-30 growth (after excluding outliers)

Explanatory variable	Ofwat's DD model[17]		Models where "Change in PE" includes after 2029-30 growth[18]	
Added Process Capacity in PE	202.9***		268.8***	
Change in DWF permit	6461.6***	6573.6***	5711.4***	6041.0***
Change in PE		1032.6***		384.6***
Ammonia dummy (<3 mg/l)	6075942***	5284737***	6831026***	6539570***
Constant	3249715***	3261917***	3250951***	3326323***
Adjusted R-squared	0.40	0.39	0.38	0.35
Observations	201	201	198	198

Source: UUW analysis

As set out in our '*UUWR_42_Wastewater supply and demand*' document, we acknowledge that forecasting post-AMP8 PE growth is more uncertain and that companies may have adopted inconsistent forecasting approaches. However, we consider that more robust estimates of PE growth could be collected before FD, allowing to achieve full alignment between costs and cost drivers. We also expect Ofwat to clarify how the post-AMP8 costs are going to be assessed and renumerated. If, for instance, at PR29 Ofwat uses a different modelling approach to evaluate these post-AMP8 costs, there is a risk that the overall scheme allowance might become incoherent across the two time periods.

Question 5.4) Do you agree with our approach to adjusting allowance to account for past under-delivery?

We fully support the principle that customers should not pay twice for enhancement. However, we have concerns regarding Ofwat's approach to assessing the extent of this past under-delivery. Ofwat's approach to constructing the adjustment for past under-delivery involves comparing historical spend on growth at STWs with company totex requests from their business plans. This approach effectively assumes that any underspend on growth at STWs compared to business plan forecasts is due to under-delivery. We consider that this assumption is largely unsubstantiated.

As acknowledged by Ofwat, the totex outcomes approach gives companies the flexibility to structure their investment programmes in such a way that corresponds best with company-specific needs and emerging risks.[19] This implies that companies should not be penalised if they decide to restructure their investment decisions in a way that somewhat deviates from the original business plan. Population growth is inherently uncertain, meaning that we should expect deviations from business plan totex forecasts. Indeed, the logic of ex-ante price controls dictates that it is up to companies themselves to manage these risks. If, for instance, population growth is higher

[17] Ofwat's model specification coefficients are slightly different from those published at DD - this is because the inclusion of the phosphorus dummy results in the identification of one additional outlier scheme.

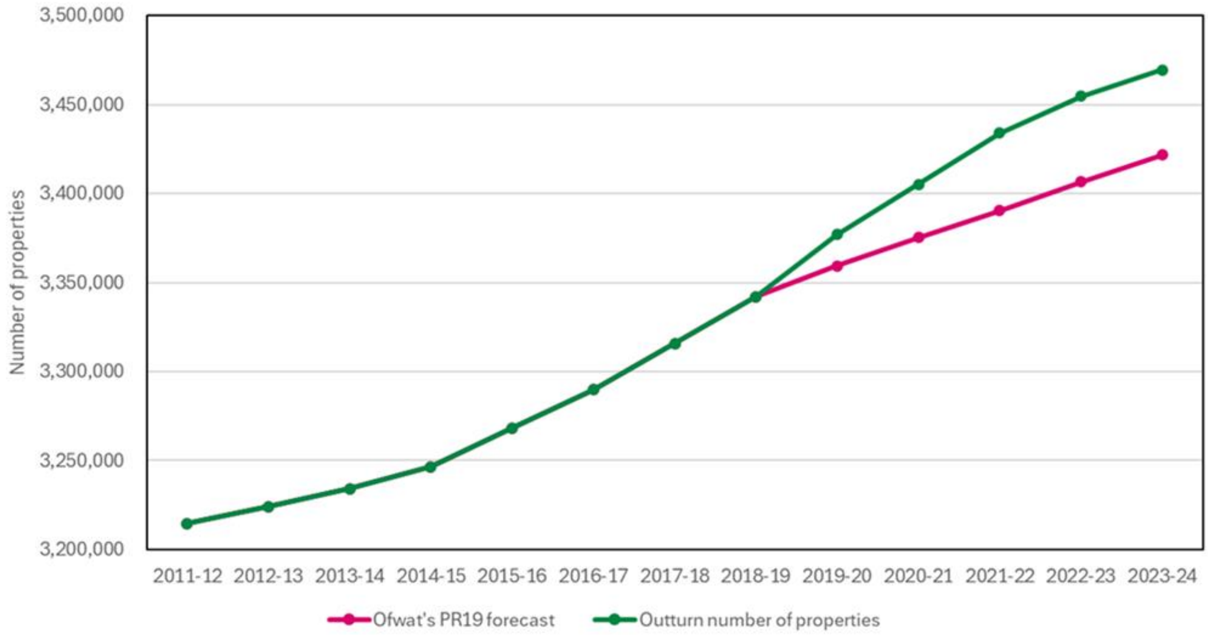
[18] Number of observations in revised models is lower than in Ofwat's DD models - this is because the inclusion of after 2029-30 PE growth results in the identification of additional outlier schemes.

[19] Ofwat (2024) *PR24 Draft Determinations: Expenditure allowances - Enhancement cost modelling appendix*, p.61.

than anticipated and leads to totex overspend, companies would be expected to absorb this cost pressure (after cost sharing). However, Ofwat's approach to account for past under-delivery appears to introduce an asymmetry by clawing back money in cases where companies have spent less than initially forecasted.

It should also be pointed out that Growth at STWs was assessed as part of base costs at PR19, meaning that there was no specific cost allowance associated with this activity. Ofwat claims that there was only a small cost challenge to botex allowances at PR19 (around 2%),[20] implying that it is appropriate to use company cost requests as a proxy for Growth at STWs implicit allowance. However, for UUW there was a much larger 7% wastewater network+ base cost gap relative to the business plan request. In fact, Ofwat imposed a targeted negative adjustment (£37.7 m) to UUW's wastewater network+ base allowances to account for the fact that population growth was forecasted to be lower in AMP7, compared to previous periods.[21] Therefore, Ofwat's approach to past under-delivery appears to involve a double-count. We also note that the population forecasts that Ofwat based this negative adjustment on have not materialised – growth has been significantly higher in UUW's region than Ofwat expected at PR19, as shown in Figure 11.

Figure 11: UUW's total number of properties (forecast vs outturn)



Source: UUW analysis

Ofwat has also not provided any specific engineering or operational evidence that would demonstrate actual under-delivery (e.g., permits not being met). This creates a risk that Ofwat's current approach will punish companies for identifying efficient solutions to growth pressures. We acknowledge that Ofwat has applied cost sharing rates and an additional 50% reduction to the adjustment. However, due to the reasons explained above, we consider that this approach is still too punitive. UUW's position is that a negative adjustment should be made only in those cases where there is clear evidence that underspending has resulted in demonstrable STW capacity shortages.

Bioresources industrial emissions directive (IED)

Question 6.1) Do you agree with our approach to setting efficient IED secondary containment, tank covering and other IED cost allowances?

Bioresources IED enhancement expenditure has been disaggregated into secondary containment, tank covering and 'other IED' cost categories. We have concerns in relation to the approach taken to all three enhancement

[20] Ibid.

[21] Ofwat (2019) *PR19 final determinations: United Utilities final determination,* p. 38

areas. Ofwat has developed highly parsimonious models for the first two cost categories, with only one explanatory variable in each model. These models do not perform well from a statistical point of view. As set out in our IED representation (*'UUWR13.2'*), differences in programme scope and asset configurations are likely to lead to significant variations in efficient costs between companies. We therefore consider it important for Ofwat to continue exploring different model specifications (and potentially improve data consistency, although we recognise this may not be feasible given the timescales involved). We provide more detailed comments on each enhancement area below.

Secondary containment

Ofwat uses bund wall length as the sole explanatory variable to predict secondary containment costs. We agree that wall length is an important cost driver – as observed by Ofwat, wall length individually explains the largest share of costs (compared to other available cost drivers). However, we disagree that wall length is the only relevant factor. We consider that wall height is another important variable, which should be incorporated into the cost model. Height is important as the bunding solution can comprise long, low-height walls or short, high-height walls to provide containment for an equivalent volume. Therefore, the overall cost driver is the total surface area of the wall (length x height) and not just the wall length. The optimisation of wall length and height are primarily driven by a combination of several site-specific factors. This includes the number and size of tanks, layout of existing works, as well as interactions with other assets within the containment solution. Ofwat's current approach to cost assessment penalises companies for building short, high walls, even where this is the most efficient solution to deliver the required containment.

Based on the above, it might be claimed that there could be an inverse relationship between wall height and wall length. For instance, if the wall is built out further away from the tank (thus increasing its length), then lower wall height might be necessary to achieve effective containment. In this scenario, wall length would act as an inverse proxy for height, meaning that no further model modifications are necessary. However, we find no evidence of this being the case. At the scheme level, there is almost no correlation (0.2) between wall length and height; we find that there is no clear relationship at the company level as well. As such, there is little evidence of collinearity between the variables.

There are two main ways in which height could be incorporated into the model. The first method involves multiplying wall length and wall height variables and using this interaction as the sole explanatory variable. This method in effect produces the implied wall 'surface area' estimate. The second method could be implemented by including length and height as two separate variables. We prefer the first approach as it produces better statistical outcomes and is simpler, leaving additional degrees of freedom.

The new variable is statistically significant and increases R-squared to 0.33, as shown in Table 9Table 1. Meanwhile the range of company-level efficiency scores decreases from 3.56 to 2.07, indicating that there is less unexplained variation in costs. The new model results in a median efficiency challenge of 0.9 – we consider that applying this factor would provide appropriate stretch for the industry.

Table 9: Secondary containment models (after excluding outliers)

Explanatory variable	Ofwat's DD model	UUW proposed model
Wall length	**0.005*** **	
Wall length x height		**0.0048*** **
Constant	**0.789**	**1.665*** **
Adjusted R-squared	0.201	0.329
Observations	90	92

Source: UUW analysis

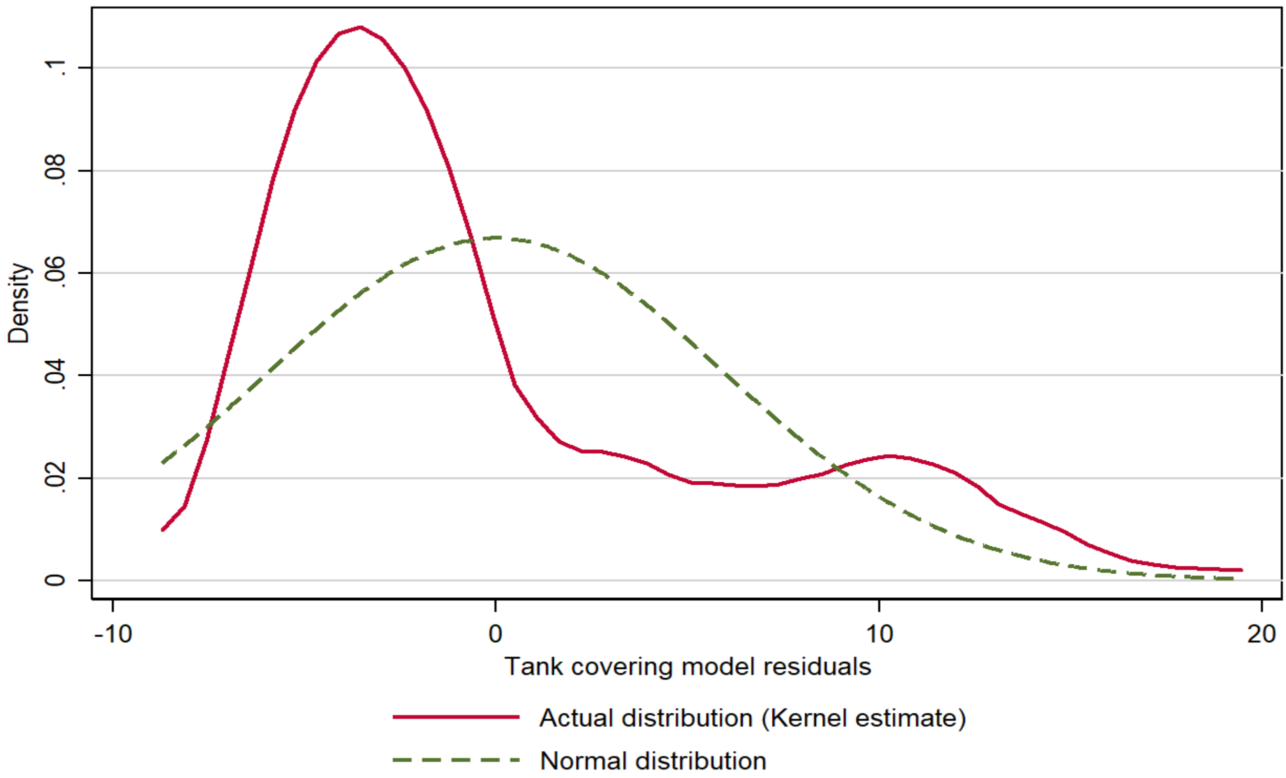
Given these improved modelling results, we support the inclusion of wall height in the secondary containment modelling approach for FD.

Tank Covering

Ofwat's DD tank covering model uses surface area provided as the sole explanatory variable. This simple model is unable to explain most cost variations across different schemes, as indicated by the very low R-squared score of 0.078. However, we have been unable to find a more appropriate model specification, based on the provided cost driver data. To improve model performance, it might be necessary for Ofwat to collect additional data. A range of tank types (e.g. raw sludge tanks, centrate tanks, primary digesters and secondary digesters) will be covered in order to meet IED compliance, reflecting a range of assets and digestion processes. Therefore, it is expected that each tank type would have a different scope of works, and therefore a different unit rate for covering.

We are also concerned that Ofwat has used its highly parsimonious model to set a stringent catch-up challenge (0.37). In a well-specified OLS regression, we would expect residuals to follow a normal distribution, with a mean of zero. However, when the probability density distribution of actual residuals from Ofwat's model is plotted (Figure 12), we find that the residuals are left-skewed. In other words, there are many schemes where model predicted costs are above company requested costs. Unsurprisingly, the model also fails the statistical test for normality. This residual skewness (which is most probably due to model misspecification) in turn results in a very stringent catch-up challenge.[22]

Figure 12: Distribution of tank covering model residuals (outliers excluded)



Source: UUW analysis

However, it could be argued that model residuals are merely capturing company-specific (in)efficiency, rather than statistical randomness. To test whether this is the case, we employed stochastic frontier analysis (SFA), which allows to decompose the error term into technical inefficiency and statistical noise. SFA allows to generate the lambda value. A lambda value >1 would indicate that most of the error is driven by company-specific (in)efficiency. However, we find that Ofwat's tank covering model specification yields a lambda of 0.011, indicating that the vast majority of the error term is driven by statistical noise. The very low lambda value is likely due to the fact that Ofwat's model does not take into account the cross-sectional (i.e., company) dimension of the data.

[22] We note that in our revised secondary containment model residuals more closely follow a normal distribution, and thus imply a more plausible catch-up challenge.

We recognise that Ofwat capped model allowances for some companies to prevent them from receiving more than they have requested. However, the catch-up challenge is still based on the non-capped model estimates, meaning that companies' final allowances are still indirectly influenced by these highly inaccurate predictions. Based on the above evidence, we consider it inappropriate to apply an additional catch-up challenge to tank covering model predictions. Therefore, unless a better model specification is found, the catch-up challenge should be assumed to be equal to 1.

'Other IED' costs

Ofwat benchmarked 'other IED' costs based on the quantity of sludge produced (TDS). This generates counterintuitive results, with some companies receiving a higher allowance than they have requested for 'other IED' costs. This is likely due to differences in programme scope across companies: we observe that significant expenditure in the 'other' category is isolated to a relatively small number of sites. The cost allowance should reflect the site-specific scope and not be benchmarked on TDS.

We consider that are more appropriate approach would be to apply a high-level efficiency to submitted 'other IED' costs. This efficiency factor could be derived by estimating the cost gap (between predicted and submitted costs) from secondary containment and tank covering models. This approach effectively assumes that company-specific inefficiency is equivalent across secondary containment/tank covering and 'other IED' costs. Ofwat has already implemented a similar methodology to assess Flow to Full Treatment costs as part of its overflows enhancement modelling.

Question 7.1) Do you agree with our approach to assessing supply interconnector enhancement costs?

We do not provide representations in this area.

Sanitary parameters

Question 8.1) Do you agree with our approach to assessing sanitary parameters enhancement costs?

We are generally supportive of Ofwat's approach to apply a top-down company level econometric modelling approach in this area. Like Ofwat, we concluded through our internal analysis this is a suitable approach in this area due to the difficulty of building a robust model that incorporates both main cost driver data (population equivalent) and other key sanitary parameters (BOD, ammonia and suspended solids). We have made the following observations regarding what factors Ofwat should carefully consider when attempting to develop scheme-level sanitary parameters models.

We have observed that there is large variation in the scheme-level dataset across schemes and companies. This variance in complexity of scheme will make it difficult to develop a model that accurately reflects costs at a site level. Some companies have much smaller, less complex schemes than others in the industry. For example, Welsh Water have lots of very small sanitary schemes with low PE, whereas UU has much larger schemes, on average. For example, the average PE of a Welsh Water scheme is only 1,799 compared to a UUW average PE of 111,102 (excluding Davyhulme our average is still 64,347). These schemes are clearly fundamentally different in size and scope. Ofwat will need to carefully consider whether prospective models are capable of controlling for these fundamental differences.

Ofwat will also need to carefully consider whether its proposed models reflect whether existing on-site assets could contribute to cheaper delivery for some companies. This is important because existing infrastructure at some sites will makes delivering AMP8 solutions cheaper. Unless this can be appropriately reflected in the model, this could make these schemes appear unrealistically efficient. As these schemes are incomparable to solutions requiring full infrastructure upgrades, and are much cheaper to deliver, they would also have the effect of artificially compressing the modelled allowance to unachievable levels for schemes which require infrastructure upgrades to deliver new permit requirements.

We have also observed there is a lack of datapoints for sanitary parameters relative to other areas assessed by scheme level data such as phosphorus removal and overflows. This makes it difficult to develop a robust scheme level model. There are only 287 schemes for sanitary parameters across the industry, compared to 1,879 for overflows and 993 for P removal. Of these 287 Sanitary Parameter schemes, 90 are Welsh Water schemes which are predominantly low cost, low PE sites. These sites are hard to compare to the majority of our solutions which are generally much larger sites. The variability in scheme sizes between companies mean a consolidated company level model is more appropriate to accurately reflect costs.

We note Ofwat's intention to attempt to develop a scheme-level sanitary parameters model for FD. However, Ofwat will need to carefully consider the factors set out above when doing this. We are not confident that a robust sanitary parameters model can be developed without an extensive period of industry engagement on methodology and data consistency. We would be concerned if Ofwat presents an entirely new model at FD without giving companies the opportunity to examine and comment on its validity. We would be especially concerned if this updated model resulted in worse outcomes for companies.

We consider that design PE best aligns with engineering rationale

PE is a major driver of costs at sanitary sites. Ofwat has used average PE served over the modelling period when estimating cost models. We consider that design PE more accurately reflects the design specification of sanitary parameter solutions and therefore, the associated costs. When we design our schemes, we consider future PE requirements carefully to ensure the solution remains robust to future growth in PE served by the STW. This usually results in a higher design PE than PE served, which more accurately reflects scheme size and cost. We consider that Ofwat should use design PE for both Sanitary Parameters and P removal.

Question 8.2) Do you agree with our approach to addressing the implementation issues associated with modelling sanitary parameters enhancement costs?

We have observed the methodology for this area is very hard to follow back to source data due to reliance on the use of pivot tables in the 'Scheme data input – CWW3' tab, which predominantly feeds the 'reconciliation adj factor' tab. It can be difficult to clearly ascertain where data has come from and what each section of the table is representing. This adds unnecessary complexity in assessing the validity of this approach. We encourage Ofwat to develop a more simple spreadsheet design with clear signposting to allow easier interrogation of this section of the model.

We consider that data should have been collated at a scheme level for 2023-24 for Sanitary parameters schemes. Capturing data in this way would have made it much easier to align scheme level data with the consolidated totex figures in CWW3/12/17, removing unnecessary complexity in the adjustments Ofwat are required to make.

We observe that Ofwat has made an adjustment to exclude costs to be incurred after 2029-30 from efficient allowances. We understand Ofwat's approach in using full scheme data, including post 2029-30 costs, to assess scheme efficiency. However, Ofwat does not set out how or when these costs will be reflected within cost assessment. We consider that Ofwat must clarify its approach to setting allowances for post 2029-30 costs, relating to AMP8 schemes.

We note that Ofwat also makes an adjustment to provide an allowance for schemes dropped from the modelling dataset due to data quality issues. We disagree with this methodology in certain areas. Schemes dropped from the dataset due to company data issues should be penalised rather than effectively rewarded with an uplift proportional to the company modelled allowance. This approach appears inconsistent with how Ofwat has assessed similar situations in other cost areas.

Metering

Question 9.1) Do you agree with our approach to assessing new meter installation and meter upgrade costs?

We consider that there are several issues with Ofwat's approach that have led to an inappropriate cost allowance. Whist there are a number of potential issues, we consider that two factors are most material. Specifically, they are the limited consideration of important operational context (such as meter fit locations), and an over-weighting of cost data from a single group of associated companies.

Ofwat's cost models use a single cost driver, missing important additional factors.

This approach is unable to recognise the substantial differences in the unit cost of installing different types of meters, or that companies have significant differences in the mix of meter installations included in their programme.

In the case of UUW, the unit costs of different meter install types can vary by over 300%. Factors that influence these costs include meter installation location, household vs non-household, and whether the meter is part of a proactive fitting programme initiated by the company or reactive fitting as a result of a customer request or meter failure.

***Table 10: UUW examples of new meter install unit costs – 2022/23 price base*[23]**

Fit type and location	£/meter
Household – Company initiated – External existing box	£106.36
Household - FMO - External - External existing box	£122.79
Household - FMO - Internal	£202.70
Household - Company initiated - New box	£424.36
Household - FMO - External – New box	£496.83

Source: UUW analysis

Crucially, companies have a limited ability to influence the profile of meter fitting activity type included in their meter fitting programme. For example, the take up of higher cost, Free Meter Options by customers is, to a material degree, driven by customer choice rather than company control. Similarly, the choice of fitting inside or outside a property is strongly influenced by the age and mix of property types in a given location.

The inability of Ofwat's cost model designs to account for these important cost drivers is likely to be a major part of the very wide range of efficiency challenges applied by Ofwat, with enhancement claims adjusted by up to c.+/-60%.

We have not sought to generate alternative cost models in this instance, as we have limited access to information on the make-up of other companies metering programmes, or commercially sensitive details of the maturity of their commercial arrangements. However, a review of companies' business plan submissions indicates that there are material differences across companies in the relative proportion of different meter fitting types. For example, we have identified that Severn Trent plan to fit 45% of new household meters in existing boundary boxes[24], whilst in the North West we expect no more than 7% of meters can be fitted in this way[25] (35,000 of 500,000 new household meters).

We consider that through the query process Ofwat have acquired, or could acquire, detailed information on each companies' unit cost and projected activity levels[26] across a range of different metering types. This information could allow more advanced and accurate cost model development.

Ofwat's models don't appropriately correct for correlation between Severn Trent and Hafren Dyfrdwy data points

Severn Trent and Hafren Dyfrdwy are associated companies. Both companies have the same parent company and share many of the same support functions. It is likely both companies share a similar approach to commercial strategy. This is highlighted by the similarities between the two companies' unit cost for fitting new smart meters.

[23] Quoted rates reflect the lowest of latest submitted bids into the ongoing smart meter tender process and include costs associated with meter purchase, installation, read communication, and internal costs and overheads.

[24] Severn Trent PR24 Business Plan submission "Meeting-our-future-water-needs", p. 56.

[25] UUW response to Ofwat query Ref – OFW-OBQ-UUW-045.

[26] Ofwat query Ref – OFW-OBQ-UUW-045.

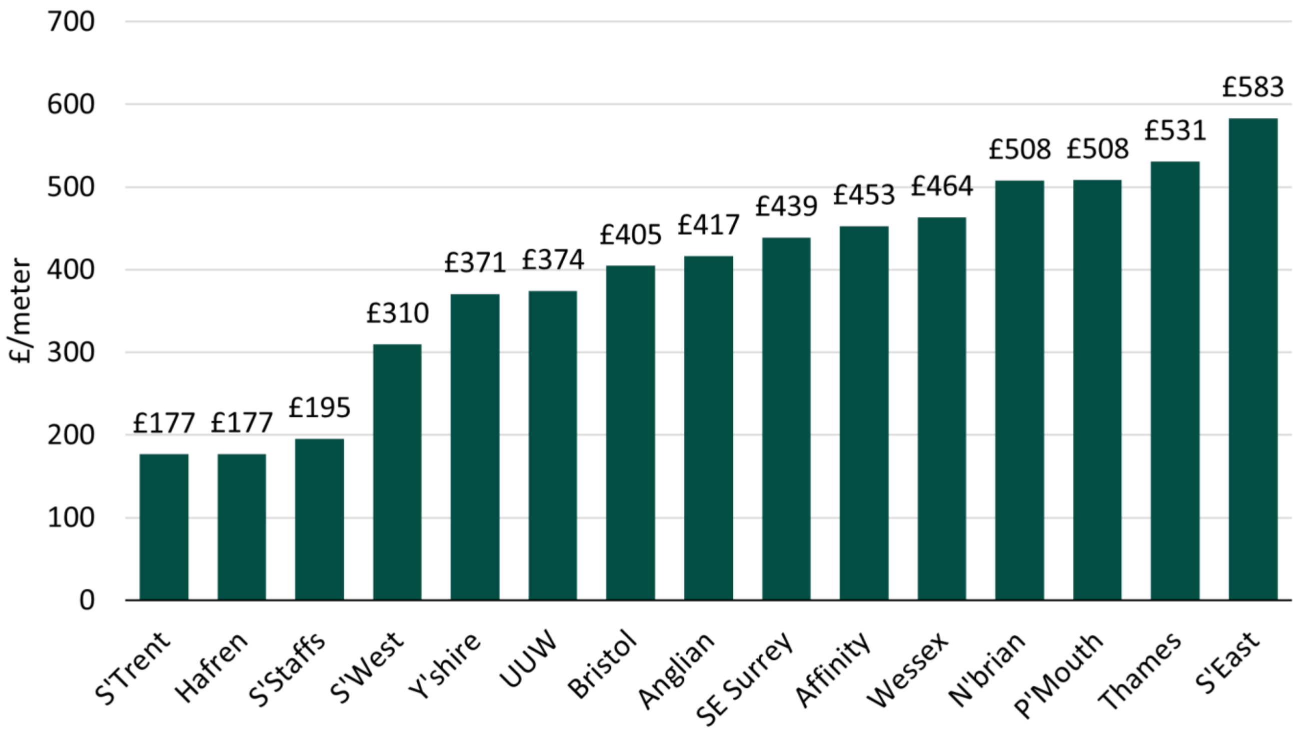
Ofwat's current cost modelling approach treats Severn Trent and Hafren Dyfrdwy as entirely independent. This might be appropriate in some circumstances, for example:

- **Where the two companies have entirely separate management and procurement structures.** In this circumstance, it would be appropriate to reflect both companies separately within the benchmarking assessment. This is because this allows each company to add into the assessment additional information about the relationship between cost and cost driver, as separate companies are likely to implement individual procurement strategies. As such, the sample size increases, and the resulting analysis is more robust.
- **Where the two companies operate in materially different environments subject to unique regional factors and the service of interest is affected by these circumstances.** In the case where two companies share common ownership, it may still be justifiable to include each separate company if it is thought this adds additional information into the benchmarking analysis. For example, Bristol Water and South West Water are owned by the same parent company but are included as separate observations within botex benchmarking. This is justified because each of these companies serves a unique region and as such the inclusion of each company separately adds additional information, making the resulting benchmark more robust. The same is true for the separate inclusion of Severn Trent and Hafren Dyfrdwy in water botex benchmarking.

However, in the case of smart metering activity, we do not consider these conditions to hold. There is clear evidence that Severn Trent and Hafren Dyfrdwy have the same procurement strategy, each company has projected the same unit cost to fit a new AMI meter. This suggests that the costs submitted within each company's business plan have been derived from the same commercial tendering process. As such, the inclusion of the two companies separately effectively doubles the weight of a single procurement strategy.

Figure 13 below shows the extent to which these two companies may be influencing the benchmark position. There is clearly a risk that treating these companies as sperate entities could be skewing modelled cost projections.

Figure 13: New meter household installations – PR24 data tables £/Hh



Source: Data taken from PR24 data tables, January resubmission

Given this, caution should be exercised when considering cost model construction. There is a high likelihood that the two companies should in practise be treated as a single entity for the purposes of cost assessment.

Otherwise, there is a risk of giving outsized influence to what is, for all intents and purposes, a single highly influential company within this particular cost assessment.

Taken together, these two material issues have resulted in a cost model design that is unlikely to accurately reflect the true efficient costs associated with delivering 920,000 AMI enabled meters in the UUW region.

Question 9.2) Do you agree with our decision to assess smart infrastructure costs within the meter installation and meter upgrades models?

Ofwat's approach increases the simplicity of cost assessment. However, we note that while simplicity may be desirable from a regulatory perspective, it does not mean the subsequent outcome will be accurate. We consider that this approach risks creating allocation problems across the industry. As such, Ofwat should take additional precautions to ensure that it does not set an unachievable benchmark e.g. by choosing a benchmark that is proportionate to this methodological simplification.

Lead

Question 10.1) Do you agree that the number of lead communication pipes replaced or relined is the key factor that explains differences in efficient costs?

We acknowledge that the number of lead communication pipes replaced or relined is a key factor that explains costs in this area. However, there are additional explanatory factors Ofwat should consider to further develop a robust model that more accurately represents other key drivers of cost.

We consider that pipe length should be a key consideration for Ofwat in their models. Some companies may have a limited pipe length per replacement, as their strategy is focussed predominantly on simple jobs selected at the company's discretion, which can often have a short average pipe length.

Our strategy for lead communication pipes in AMP8 ensures that we are non-discriminatory in what pipes we replace. Our replacements are customer-driven which means that we have little control over the mix of jobs that we will complete. Based on our AMP7 data, we anticipate a wide range of replacements. This can lead to a wide range of costs for each of our lead replacements.

We recognise that concerns over data robustness may reduce confidence in the use of other cost drivers. However, we consider that Ofwat should reflect these concerns as a mitigating factor within its general approach e.g. through the continued use of median unit costs and a catch-up challenge that is proportionate to the omission of key cost drivers from the model.

Question 10.2) Do you agree with our approach to triangulating between the median unit cost and an econometric model?

We are broadly supportive of an approach that triangulates between different models. This mitigates the risk that any one method unduly benefits or penalises a particular company.

United Utilities Water Limited
Haweswater House
Lingley Mere Business Park
Lingley Green Avenue
Great Sankey
Warrington
WA5 3LP

unitedutilities.com

United Utilities

Water for the North West