

Understanding the exogenous drivers of wholesale wastewater costs in England & Wales

:vivideconomics ARUP

Contents

Executive Summary

FINDINGS

This report sets out evidence on the drivers of wholesale wastewater costs in England and Wales.

It summarises work undertaken by Arup and Vivid Economics that reviewed the factors associated with water and sewerage companies' operating environments, identified those that affect companies' costs and examined the extent to which these effects can be observed in benchmarking models.

The work brings together engineering and econometric evidence to provide a comprehensive view of the way costs are determined. The project assembles narratives on the causal factors that affect efficient costs, data on how these drivers vary between company regions, and statistical evidence on the relationships between drivers and efficient costs. This integrated approach to wastewater cost assessment adds inter-disciplinary rigour to much of the existing commentary on the topic, which tends to focus on observed statistical relationships within a small set of data.

The study focuses on fourteen topics, having initially considered more than 200 narratives. The study entertained a long-list of factors encompassing all aspects of wastewater service provision. The areas of focus were chosen where there was the greatest scope for more evidence to improve cost assessment.

Evidence gathered by this study has valuable implications for cost assessment at PR19. The study's account of how various drivers affect costs supports a series of recommendations for the conduct of cost assessment at PR19. The recommendations fall into three categories:

- ***Modelling practices*** that make the best use of limited data to estimate how different drivers affect costs
- ***New variables*** that can be added to benchmarking models to reflect critical drivers of company costs
- ***New data*** on company characteristics to support more precise benchmarking of costs.

Three categories of recommendations:
modelling practices, new variables and new data

6x lower

Unit costs of treatment at large works compared to those at smaller works

MODELLING PRACTICES

Benchmarking models should be strictly tailored to engineering narratives and avoid techniques that produce unstable results with small data sets. A critical challenge for cost models is a lack of data: there are only ten comparator companies and published information on costs and drivers is limited in detail and coverage. Though optimal modelling approaches vary with the data available, the study finds:

- Drivers included in models must have strong positive engineering or economic justifications, as limited data precludes the use of an extensive set of drivers. In this context the study recommends adopting the Cobb-Douglas functional form, which allows for the inclusion of more critical drivers than the alternative translog form at a relatively modest cost of reduced flexibility.
- Panel data modelling specifications are not viable with current industry datasets. Analysis shows that, with the small datasets available, panel data model results are highly unstable over time and sensitive to technical choices made by the modeller.
- Estimated relationships between drivers and costs can lie outside plausible ranges, generating questionable predictions. In particular, time trends observed in the data and the effect of regional wage variation on costs cannot be explained using economic narratives of efficient costs. To avoid generating unrealistic projections of costs for future Asset Management Periods (AMPs), this study recommends the use of 'ex-post' adjustments to account for regional wage differences and time fixed effects for temporal variation.

NEW VARIABLES

The inclusion of new drivers accounts for factors for which there is engineering evidence of a relationship with cost. The report identifies several drivers where there is a clear engineering case that they affect costs substantially, but where previous analysis has not adequately accounted for this.

- Discharge permits (also referred to as consents) constrain company choices of treatment technologies, which in turn affect unit costs of wastewater treatment. More complex forms of treatment raise unit costs by around 50 percent.
- Economies of scale mean the unit costs of treatment at large works are around six times lower than those at smaller works.
- Run-offs into combined drainage systems affect the need for network storage assets and pumping costs. Engineering modelling shows that significant differences in network costs between companies can be explained by variation in urban run-offs between company regions.
- Aspects of operating in urban environments, such as slower traffic speeds, space constraints and hard surfaces, raise operating and capital costs.

Econometric assessment shows that, as well as being well motivated by engineering narratives and evidence of inter-company variation, many of these factors are significant in benchmarking models.

NEW DATA

Improved data collection can allow greater precision in cost estimation. The third group of recommendations concerns the evidence that underpins cost assessment. For some variables, notably on permits that relate to treatment quality, the recommendation is to collect comparable sector-wide data. For others, the recommendation is to improve data quality so that their inclusion in econometric models can be tested more robustly. More precise regulatory guidance on how companies should calculate load, in particular, would address concerns on the comparability of reported data: at present company reports vary from estimates based on a single set of demographic data by up to 15 percentage points. On regional wages variation, publicly available datasets based on ONS regions do not reflect localised differences in labour market conditions that affect costs: companies could provide such information during price reviews.

15% variation

in company reporting based on estimates from a single set of demographic data

Individually, each of these recommendations addresses problematic aspects of the cost assessment models used in PR14. The PR14 models omit important engineering drivers, while diagnostic indicators suggest that their results are unstable and that the models are misspecified. As a consequence, the models are unlikely to generate reliable projections of costs in future AMPs, as required. When implemented one-by-one in models similar to those used at PR14, the recommendations set out in Table 1 each cause improvements relative to the PR14 models.

Collectively, the recommendations yield substantial improvements. Models in which all of the feasible recommendations are implemented together outperform the PR14 models across the suite of diagnostic tests used for model assessment:

- ***Engineering narrative:*** the models include a fuller set of drivers than covered by the PR14 models, including variables for drainage, economies of scale and urbanisation. Furthermore, coefficient estimates are consistent with engineering and economic narratives for all variables, whereas this was not true of coefficients on time trends and regional wages in the PR14 models;
- ***Stability and model specification:*** the models perform substantially better on statistical tests for multicollinearity and misspecification than the PR14 models. Variance inflation factor statistics (VIF), which in the PR14 models indicate a high degree of instability, are reduced to acceptable levels; Ramsey RESET test scores, which test for misspecification of functional form, change from failing with a high degree of confidence to passing at a borderline level in some specifications.
- ***Statistical significance:*** there is statistical evidence that all of the factors included in the models affect costs, whereas some drivers including density are not statistically significant in the PR14 models.

Important work remains in modelling and data collection. Further modelling work can build on these promising results, refining aspects of the base cost models such as the measurement of treatment quality and urbanisation, and extending the scope of analysis to cover enhancement costs. This can be reinforced by the collection of more or better quality data in critical areas such as consents and load.

Integrated approach

using engineering narratives to support econometric analysis adds a unifying perspective

CONCLUSIONS

This study has assembled engineering narratives and modelling evidence, new datasets and econometric analysis in support of its recommendations. This integrated approach adds a new and unifying perspective to much of the existing commentary on wastewater cost assessment, which tends to be confined within subject disciplines.

It highlights substantial scope to enrich cost assessment with econometric modelling underpinned by engineering intelligence. The study proposes:

- Modelling practices that are tailored to the small datasets available, including the use of the Cobb-Douglas functional form and avoiding panel data specifications that contribute to instability
- The addition of new drivers that reflect engineering narratives in drainage, economies of scale and urbanisation
- Collection of more data on permits, more reliable data on loads and the use of alternative measures of regional wages.

The study identifies shortcomings of the PR14 models. The PR14 models lack clear supporting engineering narratives on underlying causal processes and omit or misrepresent some of the critical factors assessed by this study. Furthermore, the models have problematic statistical properties including unintuitive or insignificant results, unstable results, and model misspecification. Diagnostic tests show that these statistical issues become more troubling when recent data is added to the models.

It shows that adopting the recommendations in this report alleviates many of the weaknesses of the PR14 models. New models that incorporate a suite of the study's recommendations are better motivated by engineering narratives and outperform the PR14 models on diagnostic statistical tests. This gives good grounds to suppose that these models have superior predictive power to those used in PR14.

More valuable work remains in drawing together new cost assessment models and the collection of more data. Modelling shows that the recommendations of this report go some way to addressing the most acute concerns with the PR14 models. Further work could develop possible models drawing on improved data collection in areas identified as priorities, notably on treatment permits and load. This would allow a suite of models to be developed for use at PR19 that balances performance across a range of criteria.

RECOMMENDATIONS

The table below summarises the overall recommendations from the study.

RECOMMENDATION	RATIONALE	COMPARISON WITH PR14 MODELS
MODELLING DATASETS		
Adopt Cobb-Douglas specification	Cobb-Douglas specification allows the inclusion of more drivers – a key advantage with small dataset. Analysis shows this also reduces instability and does not introduce bias.	Alternative translog specification used, which can accommodate fewer drivers. Though translog is more flexible, there is little engineering or statistical evidence that this flexibility is valuable.
Use only Ordinary Least Squares models	Panel data models cannot be estimated robustly with the small number of data points available. OLS should be used instead.	Panel data models used. Results highly unstable over time and sensitive to technical choices over estimation methods.
Change approach to time trends and regional wages	Use an off-model adjustment for regional wages, which perform poorly when included directly in models. Do not use time trend, which does not reliably predict costs.	Regional wages and time trends included in models.
NEW VARIABLES		
Account for drainage costs using runoff data	Company-level data on urban runoff should be included in models. It is a critical engineering driver that explains variation in drainage costs.	No drainage driver in PR14 models.
Measure economies of scale for individual treatment assets	There is a strong engineering narrative for economies of scale at level of treatment assets.	Company-level variables used for economies of scale, for which there is much less engineering support.
Include capacity in urban areas	Assets in urban areas cost more to operate and maintain. Use a variable reflecting the asset environment to account for this.	Company-wide average density variable employed in PR14 models: this does not align closely with measures of urbanisation.
NEW DATA		
Improve permit (consent) data	Treatment quality is a key determinant of costs. Compile and share a time series of information on permits.	No treatment quality driver included.
Measure load consistently	Issue stricter guidance in reporting guidelines on sources and assumptions, to improve the consistency of load measurement.	Lack of detail in company reports raises concerns that load estimates may not be comparable.
Companies to provide local labour cost data	Statistics based on ONS regions do not offer robust evidence of magnitude of impact on company costs.	Regional wage variable used in PR14 model performs poorly in benchmarking model.

Table 1: Summary of report recommendations

SECTION 01

Introduction

1.1 THE PROJECT

Arup and Vivid Economics were commissioned to conduct an independent review of the drivers of wholesale wastewater costs.

The work was sponsored by United Utilities under a mandate to produce an independent report.

Independence was supported by internal governance arrangements and external peer review. An internal steering group of senior staff from Arup and Vivid Economics ensured the project adhered to its terms of reference. A panel of expert peer reviewers provided comments on the work at three points during its development. Melvyn Weeks of the University of Cambridge provided peer review of the consolidated draft report.

External peer reviewers

Dr. Paul Leinster
Cranfield University

Dr. Kieran Conlan
Cascade-Ricardo

Dr. Ralf Martin
Imperial College London

Dr. Thijs Dekker
University of Leeds

Dr. Julia Ortega-Martin
University of Leeds

Dr. Melvyn Weeks
University of Cambridge

The work proceeded in two phases.

1 The first phase of the project, led by Arup, collated a long list of engineering narratives that could explain inter-company variation in wholesale wastewater costs. This is summarised in Appendix C.

2 The second phase of work, presented in this report, examined the potential for drivers associated with some of these narratives to explain costs in a way that builds on the cost assessment models employed in PR14. The study focused on exogenous drivers, which are outside of company managements' control. These are summarised in Appendix A.

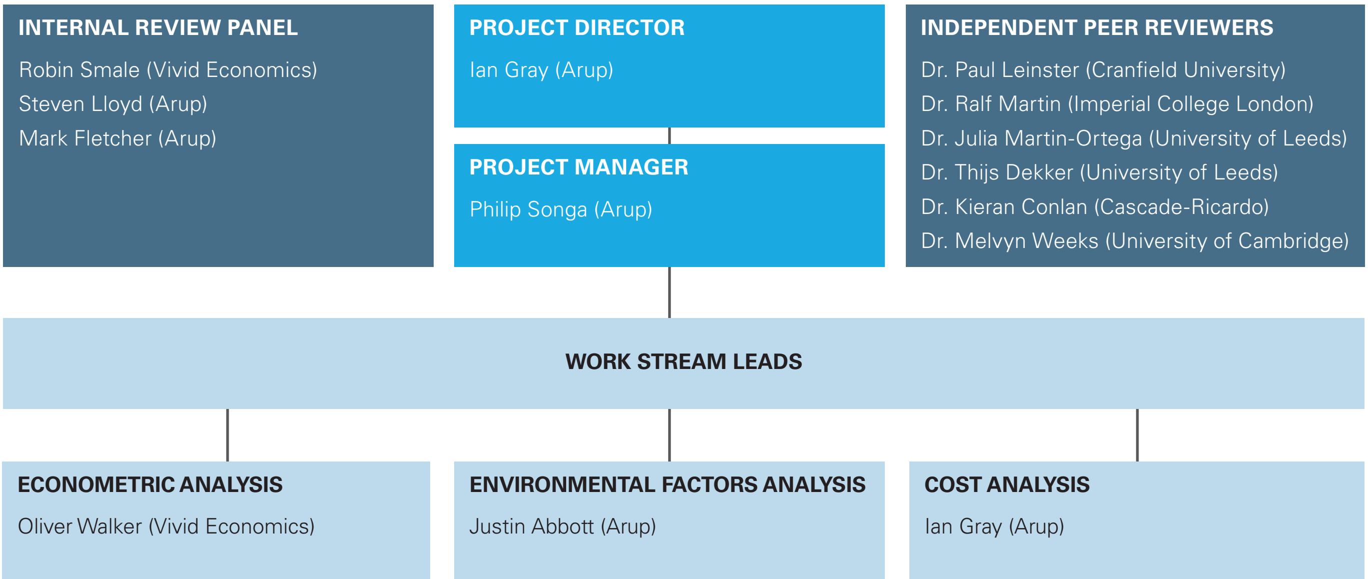


Figure 1: Project delivery and peer review team

1.2 THIS REPORT

The remainder of this report presents the evidential base that supports the executive summary, conclusions and recommendations presented. It is structured as follows:

Section 2

Section 2 presents a summary of the engineering and econometric evidence and summarises the case supporting the recommendations.

Section 3

Section 3 considers issues around model specification, setting out a critique of the PR14 models and analysis that supports the recommended modelling changes.

Section 4

Section 4 sets out new factors that can be included in models, describing supporting engineering narratives and modelling evidence.

Section 5

Section 5 explains current shortcomings in data that further research can address.

Section 6

Section 6 presents new modelling results that indicate the collective impact of the study's recommendations.

Further supporting methodological materials are contained in the appendices.



1.3 METHODOLOGY

This study focuses on fourteen topic areas of investigation. These were developed from more than 200 possible causal narratives encompassing all aspects of wastewater service provision in England and Wales. A process of expert review and filtering using Social, Technological, Environmental, Economic and Political (STEEP) framework and the Arup/Rockefeller Foundation-developed City Resilience Framework, resulted in a shortlist of fifty three factors in the first phase. Further refinements reduced the shortlist to the final fourteen areas of investigation listed in Appendix A.

The areas of focus selected were where there was the greatest scope for more evidence to improve cost assessment. Subsequently, the engineering and econometric evidence was developed using the methodology described below.

1.3.1 COMPILING THE DATA AND EVIDENCE BASE

The study followed a few clear, but often iterative, steps when compiling the evidence base:

- identify the wastewater services and cost driver categories;
- develop narratives for each driver category using engineering and environmental expertise;
- assess the relative significance of each factor across the industry in terms of the services affected, cost impact and evidence of causality: where feasible, identify metrics or proxy metrics to represent the relationship between engineering factors and costs;
- collect and analyse data;
- where feasible, test the narrative with sample data analysis prior to econometric modelling or other detailed analysis;
- where feasible, develop an appropriate time-series dataset for use in econometric modelling;
- carry out econometric assessment and make recommendations.

Figure 2 shows a schematic illustration of methodological approach.

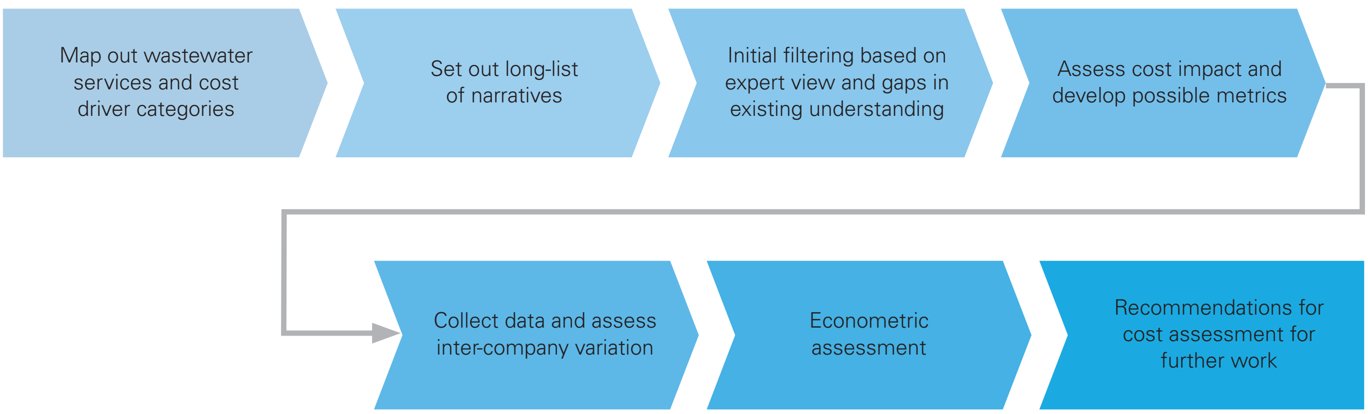


Figure 2: Schematic representation of project methodology

SECTION 02

Summary of findings

Recommendations put forward by this study rest on a consolidated body of engineering and econometric evidence.

This section summarises the study findings and the case supporting the recommendations.

The study began with a survey of possible causal narratives for company costs, spanning the dimensions of wastewater service provision in operating environments relevant to England and Wales. This identified a range of drivers neglected in previous work on wastewater costs, including factors related to drainage, treatment quality, urbanisation and economies of scale. It also highlighted areas where more accurate or relevant industry-wide data could improve cost assessment, including on permits, load and regional wages. The use of these findings in benchmarking models is constrained by a small sample of company data on costs and drivers: econometric analysis shows that relatively parsimonious functional forms and simple estimation techniques are best suited to this context. By adopting changes recommended by the study, models not only more adequately reflect engineering narratives on costs, but also substantially outperform the PR14 models on tests for significance and stability of results and model misspecification.

The study concentrates on base totex costs. The project focuses on the determinants of base operating and capital cost, known collectively as ‘botex’, rather than enhancement spending. Improvement to understanding of botex is particularly valuable both because it accounts for approximately 75 per cent of company spending and because shortcomings in econometric botex models are inherently less straightforward to amend through a special factor process than those of the simpler unit cost models used to estimate enhancement spending.

Botex accounts for approximately 75% of company spending

2.1 UNDERSTANDING THE DRIVERS OF COSTS

The study began with a survey of pertinent causal narratives. The space of possible narratives can be defined in two dimensions. First, there are the services that companies are obliged to provide: foul water conveyance, wastewater treatment, sludge treatment and disposal, and highway and surface water drainage. Second, there are features of company operating environments outside management control that affect the cost of service provision. These can be grouped as: factors affecting quantity of service, factors affecting quality of service, geographical constraints on assets and operations, service provision by others, and features of asset inheritance. A long-list of over 200 possible factors was mapped in this space for consideration (Appendix C).

Research focused on fourteen narratives where the potential to add value was greatest. An initial filtering of narratives tested the potential for improved understanding of the relevant areas to improve cost assessment. This considered, taking account of all company operating environments in England and Wales, whether the narratives described processes that had significant effects on costs, whether they described processes that varied between companies, and the degree of existing comprehension of the process and data available on relevant company conditions. On this basis, the project concentrated on fourteen narratives, encompassing all aspects of companies' activities. A complete list is set out in Appendix A.

Further assessment assembled a research strategy for each narrative. This revealed that some of these factors, such as the effect of 'hidden' subterranean urban rivers on flow volumes, described processes that were not sufficiently measurable or did not materially impact costs, so were not pursued in detail. In other areas, such as the relationship between load and costs, the nature of the relationship was already understood, but valuable insights could be gained through assessing the reliability of the data used in the sector.

The data collection and modelling work focused on the drivers identified as measurable and material. In the areas of drainage, treatment quality, economies of scale, urbanisation, and rural sparsity, the project assembled: detailed engineering evidence of how aspects of a company's environment affect its assets, operations and costs; summary data on how these factors vary between company regions; and modelling evidence on the extent to which factors affect costs. This produced positive recommendations for modelling in the areas of drainage, economies of scale and urbanisation, set out in subsequent paragraphs. For treatment quality, the engineering case for materiality is also set out but more complete and consistent sector-wide data on discharge permits is needed to account for inter-company variation in benchmarking models. For sparsity, though a valid engineering narrative suggests costs are higher in very sparse areas, this could not be substantiated by modelling evidence.

Economies of scale are most pronounced at the level of treatment assets. Economies of scale occur when the unit cost of service provision falls as the volume of service increases. Previous modelling evidence, including that used in the PR14 models, focuses on economies of scale at the level of **companies**, which would reflect managerial or operational efficiencies that can be exploited by larger companies but not by smaller ones. However, for sewerage companies these efficiencies are modest in comparison to economies of scale at the level of **assets**, where the average cost of wastewater treatment declines as the size of a treatment works increases. An engineering assessment shows that unit costs of treatment in works of 1,000 population equivalent (PE) capacity are as much as 6 times greater than costs at larger (25,000 PE) works, while industry data shows disparities between the volume of treatment carried out at small works by different companies (1 – 7 per cent). In econometric models, the percentage of load treated in small works shows a consistently positive, significant relationship with costs. Hence, the study recommends using this variable to account for economies of scale in future models.

Drainage costs can be accounted for using data on urban runoffs. Drainage is a significant component of the wholesale wastewater service, but the relationship between company-level data and costs has not been explored before. A company's activities in drainage service provision depend chiefly on inflows into combined drainage and sewerage networks. In general, the greater the volume of such inflows, the larger network and storage assets need to be, and the greater the amount of pumping. Data on urban runoffs (Figure 03), a proxy for drainage inflows, shows variation between company regions of the order of 50 per cent. Engineering analysis of the drainage processes and industry cost data suggest that costs associated with the drainage service vary notably between companies. This hypothesis is supported by modelling evidence showing that urban runoff does have a positive, significant relationship with costs, controlling for other variables included in the PR14 models. On this basis, the project recommends including urban runoff as a driver in cost benchmarking models.

Consistently positive, significant relationship

between percentage of load treated in small works and costs

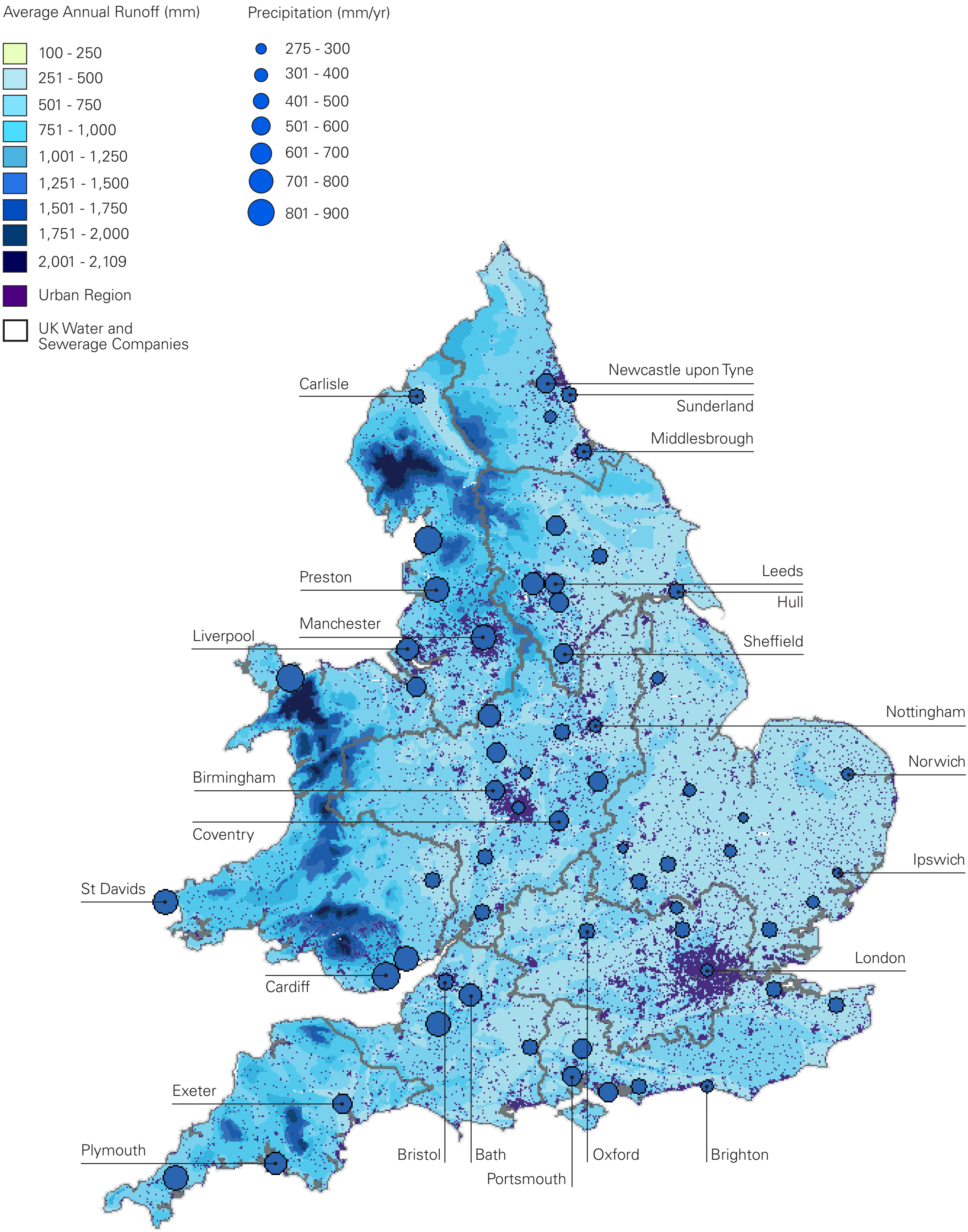


Figure 3: There is considerable variation in runoff between company regions which drives drainage costs.

Urbanisation significantly increases operating and maintenance costs. The provision of services in urban areas can be costlier than elsewhere for a variety of reasons. Access to networks for maintenance is limited by hard surfaces and the need for permissions for lane closures, operations are hampered by slow traffic speeds, while treatment assets may be constricted by land footprints and more stringent conditions on odour. However, quantitative evidence on the relationship between urbanisation and costs remains limited: for example, the network density variable used in the PR14 models measures the number of connections per kilometre of network averaged across both urban and rural parts of a company's region, a poor proxy for the relevant narratives. A more accurate reflection of how much of a company's activities take place in urban environments is the proportion of treatment capacity located in districts classified as urban by the ONS. Though this variable remains an imperfect proxy, it has a positive, significant relationship with costs in econometric models. The study therefore recommends using this variable, or one similar to it, in future models.

Treatment permits affect treatment unit costs, but current sector-wide data is insufficiently complete to value this in benchmarking models. The level of discharge permits drive the choice of treatment technologies, which in turn can have an appreciable effect on unit costs of treatment. For a large works, moving from a 'basic' permit of 20 mg/l BOD5 to a more 'stringent' permit that limits discharges to 3 mg/l NH_3 requires the installation of a nitrifying activated sludge process, a shift that causes unit costs of treatment to increase by around 47 per cent. As Figure 4 highlights, the stringency of permits varies markedly between treatment works and would thus be expected to be an important factor in explaining differences in costs. However, complete data on the permits at different treatment works is not currently reported across the industry. As explained in Section 5, the study recommends this be addressed through more extensive sharing of information across the sector.

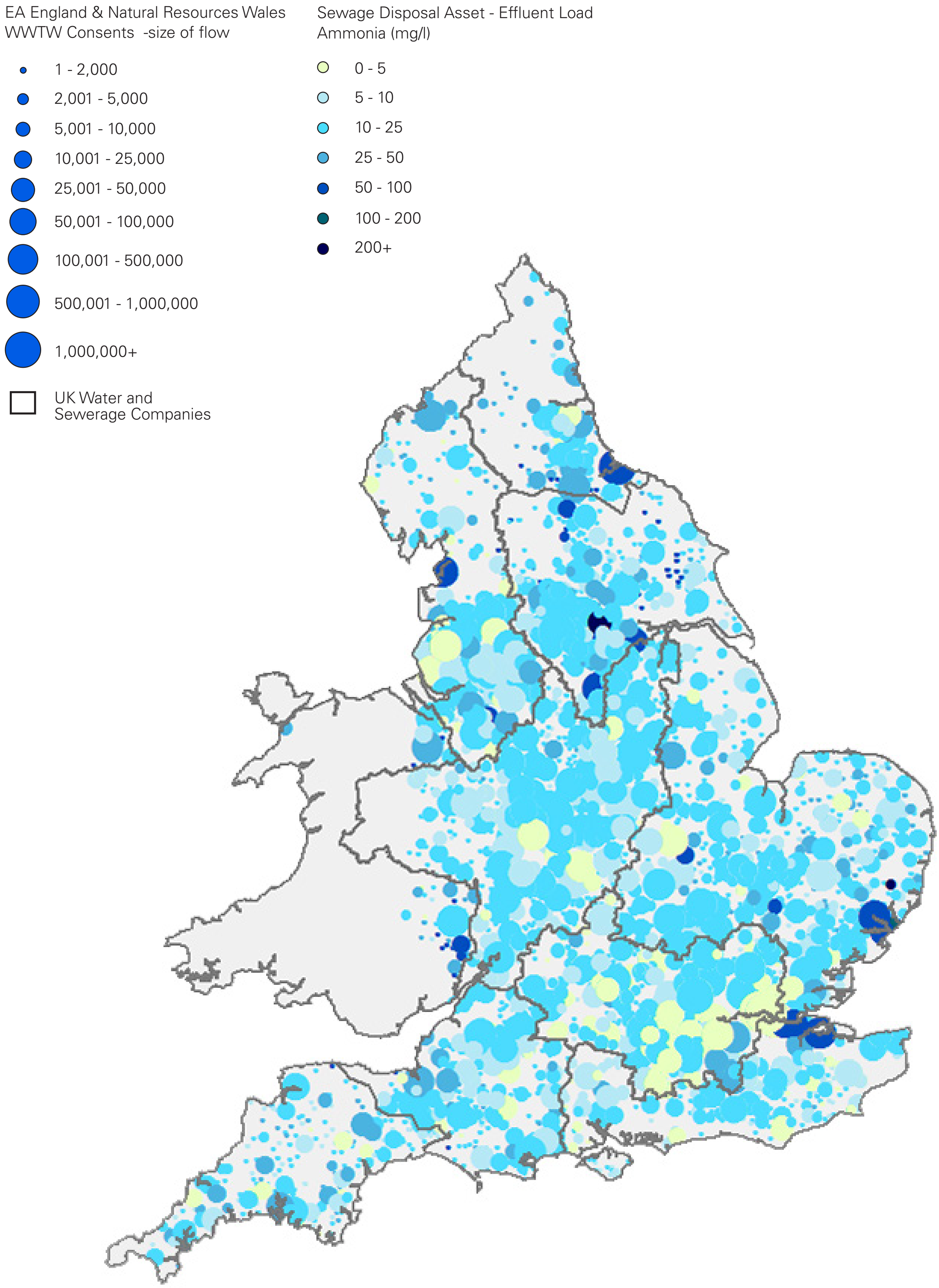


Figure 4: There is considerable variation in the size of and permit (consent) standards applied to treatment works in England and Wales.

2.2 FURTHER DATA COLLECTION PRIORITIES

Accurate cost assessment requires data that reflect causal factors and that are measured accurately and consistently across companies. However, in some critical areas this study revealed gaps in evidence or concerns over data reliability, which constrains both modelling and the wider cost assessment process. There is a need for greater consistency in reporting across companies.

More comprehensive data on permits is needed to account for differences in treatment quality drivers. As explained above, data on this critical driver is currently lacking. In order to measure accurately treatment quality drivers, companies could report comprehensive, time series information on the distribution of permits among treatment works within each size band.

More precise guidance on the data sources and assumptions used to measure load will improve the accuracy and consistency of a key driver in the wastewater models. Load is a critical volumetric driver of costs: in the PR14 models it accounts for more variation in company costs than any other variable. Yet despite its centrality to the cost assessment process, regulatory guidance on how it is calculated provides latitude to companies and the information published by companies does not allow load estimation methods to be verified. To illustrate potential discrepancies between company measures, this study used a consistent methodology to estimate total loads from published information on demographics and trade activity. It found estimated loads differed from reported loads by between -2 and +15 per cent. Stricter guidance on calculation methodologies and improved auditability of company submissions might substantially improve the consistency of this data.

Data collected by this study on regional wage variation is more closely tailored to relevant labour markets. Regional wage metrics used in wastewater cost assessment, both in PR14 and subsequently developed by Ofwat, have been based on pay data across several occupational categories. However, employees who share occupational categories may nonetheless work in quite distant sectors and hence belong to different labour markets: the wages of a manager in the financial services sector do not reflect the pay of a wastewater sector manager, for example. Analysis shows that measures of regional wages based on occupational classifications are skewed by pay in irrelevant labour markets. Alternative indices of regional wages, reflecting pay within only proximate sectors to sewerage, offer an alternative that performs better. However, none of the metrics considered reflect company opportunities to manage labour costs through locational decisions within their regions. The report recommends that companies provide this data during cost assessment, in order to support reasonable ex-post adjustments to cost thresholds.

Estimated loads differed from reported loads by between

-2% and +15%

2.3 MODELLING COSTS WITH SMALL DATASETS

Modelling practices can be adapted to the small sample of data on costs and drivers. The small sample of data points available at the company level represents a formidable barrier to including many drivers in models, with only ten companies and modest variation in summary variables over time. This lack of degrees of freedom means that models that use a long list of variables are unlikely to produce stable results with statistically significant coefficients. However, modelling practices can ameliorate the problem of a small sample size. Such practices include adopting functional forms that allow for the inclusion of more variables and avoiding methods whose performance is most adversely affected by small samples.

The use of the Cobb-Douglas functional form allows for more engineering factors to be incorporated. The PR14 models employ a translog functional form, in which the squares and cross-products of some variables are included as additional explanatory variables. Though this makes the translog a flexible functional form, there is no strong engineering or statistical case that these additional variables have a significant effect on costs. The flexibility it brings is thus of little value relative to adding further variables for which there is such a case. And given the limits imposed by the small sample size, the translog's use of additional explanatory variables makes it less feasible to include other factors for which there is a solid engineering narrative. This study therefore recommends the adoption of the more parsimonious Cobb-Douglas functional form in order to accommodate more engineering factors.

Panel data methods cannot be implemented with the small number of data points available. Panel data methods make use of the fact that companies are observed repeatedly over time and can produce results with appealing statistical properties, particularly in assessing companies' relative efficiency. Three of the five PR14 botex models make use of panel data techniques. Despite their potential advantages, however, panel data models require estimation techniques that can be highly approximate where small samples are involved. This study tested the PR14 panel data models' sensitivity to these approximations by comparing the results generated by several panel data estimation techniques, all of which perform equally well with larger samples. It found pronounced variation between results, including a greater than a factor of two difference in the coefficient on the primary load driver in one estimation. This shows the sample on wastewater companies is so small that any theoretical advantage of panel data models is outweighed by an unavoidable loss of stability in implementing them. On this basis, only Ordinary Least Squares (OLS) models are suitable for benchmarking wastewater companies. These models pool data from all companies across all years, which assumes that there is no heterogeneity in coefficients across time or across companies. Although this may be a restrictive assumption, the small dataset at hand means there is insufficient power to explore coefficient heterogeneity.

Models that include time trends and measures of regional wages do not produce reliable results, so alternative approaches to these factors should be adopted. Analysis of models similar to those used at PR14 shows that coefficient estimates for these variables are problematic, though for different reasons.

- The time trend, which is positive and significant across models, is difficult to support as there is no compelling reason to expect costs to trend upwards in real terms in future AMPs when causal factors have been controlled for. More likely, the increasing trend observed in the sample reflects omitted variables, such as treatment quality. An alternative to the use of the time trend is the use of time fixed effects, which treat temporal fluctuations more flexibly without causing notable loss of statistical power.
- The regional wage, by contrast, has coefficients in models that are usually insignificant and often of sign or magnitude that is inconsistent with economic and engineering narrative. An off-model adjustment would better control for regional variation.

2.4 PERFORMANCE RELATIVE TO PR14 MODELS

The adoption of this study's recommendations substantially addresses the PR14 models' lack of predictive power. Though the PR14 models explain nearly all of the variation within the sample data that they use, they are not grounded on causal engineering narratives and the model results lack robustness. As a consequence, the model results will not be reliable for projections of relationships between cost factors and costs.

Models used at PR14 suffer from four principal shortcomings.

1 **PR14 models are not founded on an account of the engineering factors that cause variation in costs.** This is the most fundamental shortcoming of the PR14 models, which this study addresses directly. As explained above, a consequence of this is that some of the most important dimensions of company operating environments are omitted from the models: most notably treatment permit standards, drivers of drainage costs, and the effects of operating in dense urban environments. A lack of engineering narratives also inhibits effective scrutiny of source data, since the lack of a clear articulation of why a variable is included makes it difficult to assess whether the variable has been measured in the right way. Furthermore, it is a barrier to model assessment, allowing the misinterpretation of correlations in the models as representing causal relationships.

2 **The models are not stable to small changes in the sample data.** This was illustrated most starkly by extending the sample data series by three years using the October 2016 industry data collected by Ofwat, an exercise that caused the values and significance of many of the coefficients to change significantly as illustrated in Table 2. Part of the reason for this lack of stability is multicollinearity, meaning strong correlation between the independent variables, reflected in Variance Inflation Factor (VIF) statistic scores of the order of 10,000 compared to an acceptable level of no more than 20.

3 **Many of the coefficients in the PR14 models are insignificant or have signs or magnitudes contrary to expectation.** A third shortcoming of the models concerns their results, where two problems were observed.

- In some cases, coefficient values lie outside the possible range of causal relationships between costs and explanatory factors. For example, in some models regional wages have a coefficient of more than one, which implies that a one percentage point increase in wage costs leads to an increase in base costs of more than one per cent – impossible for an efficient company. These kinds of model results cannot be used to generate reliable predictions of future costs.
- In other cases, coefficient values are not significantly different from zero. Given the lack of a foundational account of the causes of costs, this can mean there is neither a strong theoretical case for including the variable in the model nor any statistical evidence that the associated variables have explanatory power over costs. An example of such a case is the density variable, whose coefficients are often insignificant and have signs that vary across the models. Using coefficient values such as these to predict costs lacks justification but can nonetheless have an appreciable impact on company allowances.

	NETWORK		T&S OLS		BOTEX OLS	
	PR14	PR14+	PR14	PR14+	PR14	PR14+
Log length	11.3	3.9				
Log load			17.0	11.7	14.7	8.5
Log density	50.4	-21.2	78.6	60.5	59.1	21.7

Table 2: Selection of coefficient estimates from PR14 models

Note: T&S denotes models in which the response variable was total treatment and sludge costs, which are botex minus network costs. PR14+ denotes the addition of three new years of data from the October datashare.

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

4 **The models fail statistical tests for model misspecification.** Consistent with the observation that a number of key engineering factors are not reflected in the models, all of the PR14 models fail tests for omitted variables. This gives grounds to suspect that, as well as being inaccurate, the model results generate predictions of costs that are biased.

Individually, this study's recommendations address these issues. The drivers recommended for inclusion in benchmarking models are all motivated by engineering narratives and, when added to specifications similar to the PR14 models, have significant explanatory power over costs consistent with expectations. Modelling practices endorsed by the study either reduce the instability of results or eliminate likely sources of bias. The adoption of the Cobb-Douglas functional form, in particular, leads to a steep reduction in multicollinearity.

Models that adopt the recommendations of this study collectively show notable improvements over PR14 models. New models, in which feasible modelling changes are adopted, allow the collective impacts of the study's recommended changes to be gauged. These results suggest valuable gains in the quality of cost assessment modelling can be attained at PR19. New models include a wide range of engineering factors that this study shows are important, and produce coefficient values that are consistent with expectations. The models also show improvement on tests for misspecification and multicollinearity. However, to realise these gains, more work is needed to develop and test new modelling specifications, including the collective assessment of suites or models, and to collect more data in areas of priority.

SECTION 03

Model specification

Model specification choices can address challenges posed by small samples.

A major difficulty in benchmarking modelling for sewerage companies is the small and closely related sample of data points on costs and drivers available. This section explores how modelling choices can go some way to alleviating this problem. It finds:

- the Cobb-Douglas functional form allows for more engineering variables than the alternative translog without a great a loss of statistical power;
- panel data techniques are not viable with small datasets available for wastewater companies;
- the use of fixed time effects rather than a linear trend allows the predictive validity of the models to be improved with only a modest loss of statistical power.

All of these findings support changes in modelling practice from PR14.

This section begins by setting out the challenge that small samples of data present for models of base costs and criteria for model assessment in this context, before evaluating the PR14 models against the criteria. It then moves on to show how the report recommendations improve model performance relative to the approach adopted at PR14.

3.1 BASE COST MODELLING WITH SMALL DATASETS

A limited dataset of closely related variables presents an acute challenge for modelling base costs. With only ten companies operating in the wastewater sector, there are few data points from which to draw conclusions. Adding data taken over time helps increase the quantity of information from which to draw conclusions, but at PR14 models could use only a seven year panel, during which there is very little variation in key drivers over time. Furthermore, relevant drivers for which there is data are closely related to one another, meaning the explanatory power of any individual driver is reduced. This can lead to a conflict between internal and external validity: models that fit the sample data well are liable to reflect spurious relationships in the small sample, with the consequence that the model does not perform well in projecting future costs.

This study uses four criteria to assess the external validity of models under such conditions. These are:

1. Engineering and economic motivation of key drivers and functional form. Unless models track plausible causal narratives of cost determination, statistical relationships observed in small samples are much less credible for use outside the sample.
2. Coefficient values and significance. Coefficient values should be consistent with engineering and economic theory in order to generate reliable predictions of costs, while insignificant coefficient values are indicative of model over-fitting.
3. Model stability, reflected by robustness to changes in sample data and tests for multicollinearity. Relationships between drivers and costs estimated by unstable model are less likely to hold outside the sample dataset.
4. Performance on tests for model misspecification. Models that appear misspecified within the sample are less likely to be well specified outside the sample.

3.2 PR14 MODEL ASSESSMENT

The PR14 models considered in this study are a set of five econometric models that estimate the relationship between base capital and operating costs ('botex') and various explanatory factors. Table 3 summarises the specifications used at PR14, with the model coefficients resulting from each row receiving equal weight in the calculation of base cost thresholds. Note that rows two and three each use both Ordinary Least Squares (OLS) and Generalised Least Squares (GLS) for estimation, for a total of five models.

The remainder of this section considers the models' performance against each of the four criteria set out in Section 3.1 in turn.

MODEL	ESTIMATION METHOD	RESPONSE VARIABLE	EXPLANATORY FACTORS (LOGGED)
PR14 Network	GLS	Network costs	Length, density, length^2, density^2, length*density, time, regional wage
PR14 Treatment and Sludge	OLS, GLS	Treatment and Sludge costs	Density, load, density^2, load^2, load*density, per cent load in bands 1-3, year, wage
PR14 Botex	OLS, GLS	Botex costs (T&S+Network)	Density, load, density^2, load^2, load*density, per cent load in bands 1-3, year, wage

Table 3: PR14 model specifications

3.2.1 ENGINEERING AND ECONOMIC MOTIVATION

Prior expectations on the critical engineering drivers are unclear. Some of the variables included lack clear engineering narratives justifying their inclusion, while some important drivers are not included. The absence from the PR14 models of various critical engineering factors is discussed in Section 4, while concerns on the adequacy of data to represent factors are set out in Section 5.

Of the variables that are included in the models, some have a clear reason for their inclusion but others lack an explanation. Load and network length variables have clear narratives about how they affect costs: it makes sense that company costs increase when more wastewater is treated or more sewers must be operated and maintained. The proportion of load treated in small works and regional wages can also be justified in principle, as discussed in Section 4.3 on economies of scale and Section 5.3 on regional wage. However, density and higher order versions of load and length lack a clear engineering narrative to explain their presence. The latter is detailed in Section 3.3 on functional form.

Coefficients are often not significant

and contrary to engineering advice

Density is associated with both economies of scale and urbanisation, so expectations on its sign are not clear. On one hand, the number of connections per kilometre of sewer main may be associated with treatment by large works, discussed in Section 4.3 on economies of scale. This narrative would suggest that increased density should decrease costs, resulting in a negative sign on the coefficient. On the other hand, network density may be associated with assets located in urban areas, which evidence detailed in Section 4.4 on urbanisation suggests increases costs. Such a narrative would suggest a positive coefficient for density, at least for some ranges of density.

Density is a weak proxy for either external factor. Because economies of scale make sense mostly at the level of an individual treatment works rather than at a company level, density measured as a company-wide average may be a relatively crude metric to capture this effect. Similarly, the effects of urbanisation are relevant only to a portion of a company's network: a company-wide average computed for the whole network is therefore a poor metric.

This study recommends replacing density with other variables. Section 4.3 on economies of scale and Section 4.4 on urbanisation set these recommendations out in more detail.

3.2.2 VALUE AND SIGNIFICANCE OF MODEL COEFFICIENTS

Coefficients of many variables are often not significant and have signs contrary to those expected based on engineering evidence.

The sign and significance of variables was examined by replicating the PR14 models and then extending them with more recent data. The extended versions of the models, which used Ofwat's 2016 industry-wide datashare and are denoted PR14+ in tables throughout the report, allow for an additional check on performance. Given that the PR14+ models use larger samples of observations, their performance is expected to be superior to the PR14 models. Where this is not the case, this is indicative of the original models lacking validity outside the original sample used at PR14.

The magnitude of coefficients varies widely between different PR14 models. Table 4 shows the regression results for the PR14 OLS models for treatment and sludge and botex and for the single network model, as well as their results when the time series is extended. The amount of variation in the coefficients on load and length is notable, while density varies substantially in significance as well as sign and magnitude. Higher order terms such as length squared and density squared are not significant in many specifications.

Performance deteriorates considerably using the most recently available data. Moving from PR14 to PR14+, the variation in magnitude and significance of coefficients between models increases. Length, for example, becomes insignificant and a third the size in the network model using additional data. The network model's density variable changes sign, though it is statistically indistinguishable from zero in both PR14 and PR14+. Even using only the most stable models reported in Table 4, cost thresholds appear to change by as much as 10 per cent for one company by the addition of new data.

	NETWORK		T&S OLS		BOTEX OLS	
	PR14	PR14+	PR14	PR14+	PR14	PR14+
Log length	11.3	3.9				
Log load			17.0	11.7	14.7	8.5
Log density	50.4	-21.2	78.6	60.5	59.1	21.7
Log length²	0.1	0.1				
Log load²			0.1	0.2	0.1	0.1
Log density²	-2.4	4.1	-2.5	-1.9	-1.2	1.4
Log load x log density	-2.8	-1.3	-4.5	-3.4	-3.8	-2.6
Year	0.02	0.02	0.02	0.0	0.02	0.02
Log wage	0.66	-0.23	1.1	0.8	0.8	0.1
Log % bands 1-3					0.1	0.1
Constant	-170.1	23.2	-281.2	-206.8	-224.4	-103.3
R^2			0.97	0.95	0.98	0.97

Table 4: PR14 model regression coefficients

Note: T&S denotes models in which the response variable was total treatment and sludge costs, which are botex minus network costs.

Key:
- Significant at 1% (dark blue): Log length Network PR14+ … see shading
- Significant at 5%
- Significant at 10%

3.2.3 MODEL STABILITY

Models suffer from instability, reflecting overfitting and multicollinearity.

The large effect that the addition of recent data has on the sign, significance and magnitude of coefficients suggests that PR14 models are unstable and sensitive to specification. In addition to the variation across models discussed in 3.2.2, Table 4 also illustrates the inconsistency of variables when estimating the same model over a larger dataset.

Further evidence of multicollinearity between variables reinforces the conclusion that PR14 models are unstable. The extent to which drivers are closely related to one another, known as multicollinearity, is a problem for models because it is symptomatic of a lack of independent variation that is required for robust estimation. In models that are highly collinear, small changes in data or the variables included can result in large changes in coefficient estimates. Table 5 shows the statistical test results for the PR14 regressions using both the original and extended datasets. The variance inflation factor scores, which measure multicollinearity, are in the thousands for all of the PR14 models. Though measuring multicollinearity is somewhat imprecise and high values do not by themselves discount regression results, scores above 20 are signs of multicollinearity that raise concern.

	T&S OLS		BOTEX OLS	
	PR14	PR14+	PR14	PR14+
R^2	0.97	0.95	0.98	0.97
Variance inflation factor - maximum	30638	21180	31608	21347
Variance inflation factor – median	12229	9274	8150	6118
Ramsey RESET	Fail	Fail	Fail	Fail

Table 5: Test results for PR14 models using the original and extended datasets

Note: See Appendix B for more information on the tests.

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

3.2.4 MODEL MISSPECIFICATION

As noted above, there is substantial engineering evidence that key variables are omitted from PR14.

There is further statistical evidence that PR14 models suffer from misspecification. Variables that are important drivers of cost, but which are not included in a model can damage model performance. Not only is the explanatory power of a model reduced if there is an omitted variable, but its omission can bias the other coefficient estimates to the extent that the omitted variable is correlated with the included variables. Table 5 shows results for the Ramsey regression equation specification error test (RESET), which tests for model misspecification, of which omitted variables are a possible explanation. Though, as Appendix B explains, it is an imperfect measure, the PR14 models fail this test at a $p < 0.001$ level before and after recent data is added.

3.3 TRANSLOG AND COBB-DOUGLAS FUNCTIONAL FORMS

The choice between the translog and Cobb-Douglas functional form affects the number of degrees of freedom available. The translog form includes, as separate drivers, the square and cross-products of the primary drivers used to explain costs. The inclusion of these 'higher order' or 'translog' terms means the translog is a more flexible generalisation of a Cobb-Douglas form, which does not include higher order terms. The added flexibility is of benefit if there is reason to believe that there is a non-constant elasticity between the primary drivers and costs. However, the use of the translog terms reduces statistical power, which can contribute to problems of omitted variables and instability in the current setting where the dataset is small.

Higher order terms lack a convincing engineering explanation for their inclusion – this supports a general preference for Cobb-Douglas specifications. Though the flexibility afforded by the translog formulation may be valuable in modelling costs in many industries, the importance of discrete treatment and network assets in wastewater service provision reduces the motivation for its use here. Evidence reviewed by this report suggests that economies of scale are most strongly connected to assets, rather than company size. This means there is little rationale for a quadratic relationship between economies of scale and company size, as posited by the translog functional form. Furthermore, as Section 4 explains, the loss of degrees of freedom entailed by the use of the translog is very costly in this setting.

In PR14 models, translog terms are significant in only some of the models. The PR14 models rely on three primary drivers: load, length, and density. Significance of the translog terms can be assessed jointly or individually:

- A joint F-test can be used to test whether the translog terms, taken together, explain significant variation in costs. If Cobb-Douglas assumptions are too restrictive, one would expect the translog terms to be jointly significant. Though the translog terms are jointly significant in all of the PR14 models using the original data, this is only true in three of the five models when the dataset is extended to include more recent years.
- Table 6 shows the coefficients of the individual translog terms for all of the PR14 models. Unlike the joint F-test discussed above, this allows individual coefficients to be examined for sign and significance. Load squared and load interacted with density seem to perform adequately, though they are somewhat inconsistent in significance between models. Density squared and the translog variables in the network model seem to perform particularly poorly.

Given the lack of engineering motivation and the high cost of degrees of freedom, the inclusion of translog terms in models of wastewater costs does not appear to be justified on this inconclusive statistical evidence.

	NETWORK		T&S GLS		T&S OLS		BOTEX GLS		BOTEX OLS	
	PR14+	CD	PR14+	CD	PR14+	CD	PR14+	CD	PR14+	CD
Log length	3.9	0.8								
Log density	-21.2	1.0	33.9	0.2	60.3	-0.1	1.1	0.8	21.7	0.9
Log load			6.9	0.7	11.7	0.8	3.8	0.9	8.5	1.0
Log length^2	0.1									
Log density^2	4.1		-1.0		-1.9		1.8		1.4	
Log load^2			0.1		0.2		0.1		0.1	
Log length x log density	-1.3									
Log load x log density			-2.0		-3.5		-1.3		-2.6	
Log % in bands 1-3							0.2	0.3	0.1	0.3
Time	0.01	0.02	0.01	0.01	0.00	-0.00	0.02	0.03	0.02	0.03
Log wage	-0.2	0.0	1.1	1.3	0.8	0.4	0.3	0.4	0.1	0.4
Constant	23.2	-8.0	-119.8	-8.0	-206.8	-5.2	-29.4	-10.5	-103.3	-11.7

Table 6: Regression results for alternative functional forms in the PR14 models

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	T&S OLS		BOTEX OLS	
	PR14+	CD	PR14+	CD
R^2	0.95	0.90	0.97	0.94
Variance inflation factor - maximum	21180	2.1	21347	5.3
Variance inflation factor – median	9274	1.5	6118	2.1
Ramsey RESET	Fail	Fail	Fail	Fail

Table 7: Test results for translog and Cobb-Douglas PR14 models

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

The translog terms contribute to model instability in PR14 specifications. As expected, primary drivers, especially length and density, are sensitive to the inclusion of translog terms, even changing sign in the case of density. More surprisingly, regional wage, which one would expect to be largely unrelated to the inclusion of translog terms, appears to be sensitive to the functional form. This sensitivity is additional evidence of model instability.

The adoption of the Cobb-Douglas form reduces multicollinearity in PR14 models. As Table 7 illustrates, the adoption of the Cobb-Douglas form results in a dramatic reduction of the Variance Inflation Factor (VIF) relative to the PR14 models, indicating an alleviation of the multicollinearity problem. The removal of higher-order terms would be expected to reduce collinearity, though the contribution of the translog terms in this case is notable.

The use of Cobb-Douglas does not in itself resolve problems of misspecification in the PR14 models. Tests continue to show evidence of misspecification in the RESET tests, though removing the translog terms allows for more explanatory factors to be included in regressions that may improve test performance, which represents a worthwhile trade-off. Though R^2 scores remain very high, it is not surprising that they decrease somewhat because there are fewer explanatory variables in the Cobb-Douglas form.

3.4 REGRESSION ESTIMATION METHODS

There is a theoretical motivation for the use of GLS estimation in base cost models. Because all of the observations are drawn from ten companies over time, there is reason to believe that the error terms are correlated with one another or have non-constant variance, conditions which mean that OLS will not produce the most efficient results. Generalised Least Squares (GLS), of which the panel data methods used at PR14 are an example, corrects for these issues, and also allows residuals to be decomposed into idiosyncratic error and a company-specific error, which in some circumstances can be advantageous when estimating efficiency.

However, GLS results can be highly sensitive to estimation techniques in settings where the sample size is small. The implementation of GLS requires a weighting matrix that corrects for the correlations and non-constant variance of error terms. It is not known in this case how exactly the errors are correlated, so a weighting matrix must be estimated using Feasible GLS (FGLS) techniques. FGLS methods rely heavily on sufficiently large sample sizes, as otherwise different weighting matrices can produce starkly varying results.

To test whether sample sizes are large enough to use FGLS, alternative FGLS methods can be used that should produce similar results. There are a number of different statistical techniques used to implement FGLS, including a Fuller-Battese transformation, which is the method used in PR14, FGLS in Stata, and a maximum likelihood estimator (MLE) approach. All three of these techniques should obtain similar results in sufficiently large samples. Table 8 shows the results from the same regression using the same data with each different method.

The results vary substantially, offering strong evidence that the sample size is insufficiently large to be able to rely on GLS in spite of its theoretical appeal. Despite being based on identical datasets and implementing notionally identical regression techniques, there is wide variation between the coefficient estimates produced by different techniques. The coefficient on load, the most important parameter for setting costs, differs by more than 20 per cent, for example.

The appeal of GLS techniques is further reduced by the lack of variation over time. As Appendix B sets out, most of the variation in the dataset is attributable to differences between companies rather than over time. This implies that repeated observations of companies offer little additional information on relative efficiency.

	FULLER-BATTESE	STATA FGLS	MLE
Log load	14.23	12.78	15.37
Log load^2	0.07	0.05	0.08
Log density	55.67	52.0	59.29
Log density^2	-1.25	-1.49	-1.18
Log load x log density	-3.51	-3.07	-3.85
Year	-0.00	-0.00	-0.00
Log wage	1.03	1.03	1.05
Constant	-213.1	-196.0	-228.5

Table 8: Regression results for three treatment and sludge GLS methods using the same data

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

3.5 TREATMENT OF TEMPORAL EFFECTS

Temporal variation in company costs can be captured either by a trend or through the use of year fixed effects.

Positive trends in costs over time are difficult to support in theory. A negative trend in costs over time could be justified as reflecting productivity gains if these were expected to continue in the future. However, there is little reason to believe that inflation-adjusted costs should consistently increase over time: positive real price effects, if observed in the historical data, would not generally be expected to persist going forward, while productivity losses should not be built into regulatory forecasts.

Wastewater data shows a positive relationship between company costs and time. The more likely explanation for the positive temporal coefficients in PR14 models, shown in Table 8 is the presence of omitted variables, such as treatment quality, that increase costs over time.

The use of year fixed effects is an alternative, more flexible specification, though adopting this reduces degrees of freedom. Year fixed effects allow each year to have an independent effect on costs rather than forcing the relationship to be linear. However, this adds n-1 explanatory variables, where n is the number of years in the sample, and thus reduces the statistical power of any regression.

This study tested the statistical performance of models with year fixed effects against the linear time trends in PR14 models. If the fixed effects illustrate a relatively linear pattern, an assumption of linearity would be a sensible way of conserving scarce degrees of freedom. Conversely, if the fixed effects regressions retain sufficient explanatory power to retain significant coefficients, the use of fixed effects would be more appealing. Table 9 shows regression results for PR14 models using a linear trend and fixed effects.

Time fixed effect coefficients illustrate that a linear trend does not appear to be a reasonable assumption. The coefficients do not follow a linear pattern: the 2008-2010 years appear particularly high in some specifications, but there is not a strongly discernible upward trend in other periods.

Time fixed effects also appear not to excessively decrease the statistical power of the models. Not only do the coefficients on primary drivers generally attain a similar sign and significance using fixed effects compared to the linear trend, drivers occasionally are significant in the fixed effect models that are not in the linear models. This result implies that the regressions retain sufficient statistical power to usefully distinguish coefficients from zero.

The use of time fixed effects would require careful consideration on how results should be projected into the future. An advantage to the linear time trend was that projection was extremely simple. Because fixed effects allow for each year to have a new coefficient, future projections will require assumptions about how year dummy coefficients project forwards in time. However, such assumptions have been adopted in regulatory settings before, for example by Ofgem in RIIO GD1 and ED1[1].

[1] RIIO-ED1 is Ofgem's electricity distribution price control which reflects the RIIO (Revenue = Incentives + Innovation + Outputs) model for electricity network regulation. GD1 is the equivalent price control for gas distribution.

	NETWORK		T&S GLS		T&S OLS		BOTEX GLS		BOTEX OLS	
	PR14+	Year FE	PR14+	Year FE	PR14+	Year FE	PR14+	Year FE	PR14+	Year FE
Log length	3.9	3.8								
Log density	-21.2	-23.4	33.9	21.6	60.3	26.5	1.1	-0.2	21.7	12.7
Log load			6.9	7.6	11.7	8.9	3.8	4.1	8.5	7.9
Log length²	0.1	0.1								
Log density²	4.1	4.4	-1.0	1.4	-1.9	1.5	1.8	2.1	1.4	2.4
Log load²			0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.7
Log length x Log density	-1.3	-1.3								
Log load x Log density			-2.0	-2.6	-3.5	-3.1	-1.3	-1.4	-2.6	-2.6
Log % in bands 1-3							0.2	0.2	0.1	0.1
Year	0.02		0.0		0.0		0.0		0.0	
Log wage	-0.2	-0.3	1.1	0.0	0.8	-0.2	0.3	0.2	0.1	-0.1
2006/07		0.0		0.0		0.0		0.0		0.0
2007/08		0.0		0.1		0.1		0.1		0.1
2008/09		0.1		0.1		0.1		0.1		0.1
2009/10		0.1		0.2		0.2		0.1		0.2
2010/11		0.1		0.1		0.1		0.1		0.1
2011/12		0.1		0.0		0.0		0.2		0.1
2012/13		0.1		0.0		0.0		0.2		0.2
2013/14		0.2		0.0		-0.0		0.2		0.2
2014/15		0.1		-0.0		-0.0		0.2		0.2
2015/16		0.2		-0.0		-0.0		0.2		0.2
Constant	23.2	28.9	-119.8	-96.1	-206.8	-114.5	-29.4	-27.8	-103.3	-79.1

Table 9: Regression results for PR14 models comparing linear time trend with fixed effects

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

3.6 RECOMMENDATIONS

Modelling specification choices can improve model performance for PR19 relative to PR14. Figure 5 summarises the recommendations based on the evidence presented in this section.

TYPE	RECOMMENDATION
Include additional variable	Replace density, which lacks clear engineering motivation, with measures of urbanisation and economies of scale, discussed in detail in Section 4.
Change modelling approach	Adopt a Cobb-Douglas specification, which reduces multicollinearity, contributes to model stability, and enables the addition of other engineering variables.
Change modelling approach	Use only Ordinary Least Squares when estimating models. Generalised Least Squares approaches have too few data points to be viable.
Change modelling approach	Relax the linear trend assumption used in PR14, which is difficult to justify both theoretically and empirically. Time fixed effects are one solution, but careful consideration of how to project it forward is required.

Figure 5: Summary of recommendations.

SECTION 04

New drivers

4.1 ENGINEERING BACKGROUND

The costs of wastewater service provision depend on the assets that must be constructed, operated and maintained to provide these services.

Wastewater companies provide services to their customers comprising: the transfer of foul water from premises to wastewater treatment works and treatment of wastewater before its return to the environment. Wastewater companies are also responsible for sludge treatment and disposal, and highway and surface water drainage. The assets required to provide all of these services are affected by a range of factors reflecting the human and physical geography of company regions, often enforced by regulation. These in turn determine the efficient costs.

In order to assess the exogenous factors that drive engineering costs of wastewater service provision, it is therefore necessary to look at both assets and services. Table 10 shows the service categories into which wastewater assets were grouped for this study, and the main drivers of cost in each category.

At PR14, the econometric models included explanatory variables that accounted for a subset of the drivers set out in Table 10. Table 11 shows the variables that were included in PR14 botex models and identifies possible gaps. These gaps informed priorities for collecting new data and testing the significance of new drivers.

New drivers motivated by clear engineering narratives can be added to models to directly address the omission of such factors in PR14. Of the 14 factors identified and outlined in Appendix A, three were sufficiently measurable and material to be fully tested as new econometric variables. These were drainage, economies of scale, and urbanisation. This section examines the engineering and econometric case for each of these drivers in turn.

SERVICE	ASSOCIATED ASSETS	MAIN DRIVERS OF BASE COST
Surface water and highways drainage	Domestic sewers; surface water drains; highway drains; manholes; 3rd party assets	Flow volumes conveyed (peak or average); CSO spill volume and frequency; length and size of network; asset age and configuration
Foul water conveyance	Main sewers; combined sewers; pumps; CSOs; storage tanks	Foul water volumes (directly linked to populations served - both permanent and transient); location of population centres; asset age and configuration.
Wastewater treatment and disposal	Treatment works (configurations dependent on processes); storage tanks; pumps; buildings; meters; outfalls; screens; manholes	Treatment processes, influenced by regulation (permits); environmental sensitivity and effluent source (industrial versus domestic); asset age and configuration
Sludge management	Sewage treatment works; sludge treatment centres (STCs); land bank; transfer pipelines; tankers/trucks	Operational costs are influenced by level of treatment required, volumes of sludge produced, location of STCs and available disposal options

Table 10: Wastewater assets associated with wastewater service cost.

SERVICE	PR14 MODEL VARIABLE(S)	GAPS (POSSIBLE VARIABLES)
Surface water and highways drainage	Network length; network density	Combined network percentage; runoff volume; cumulative storage volumes; number of CSOs; CSO spill volume; CSO spill frequency; assets at risk of flooding (cellars, gardens)
Foul water conveyance	Network length; network density	Urbanisation; population sparsity
Wastewater treatment and disposal	Load received; percent load in bands 1-3; density	Load removed (BOD, NH_3, P per person); urbanisation; population sparsity; asset-level economies of scale
Sludge management	None	Available land bank; load; quality; volume produced; volume treated; volume disposed; transport; density of WaSC area

Table 11: Engineering services and assets PR14 gap analysis.

4.2 DRAINAGE

This section presents engineering and econometric evidence justifying the inclusion of a drainage driver to explain costs. It shows that:

- drainage services are a major component of sewerage company activities, but there is no driver associated with this service in the PR14 models;
- in order to provide these services, companies with combined networks need to invest in and maintain greater network storage capacity and incur pumping costs when operating these assets;
- variation in drainage requirements leads to highly material variation in network costs. Sample industry data and modelling show that base costs vary significantly between wastewater companies, allied with volumes of flows received and the size of networks required;
- representing drainage requirements by urban runoff shows very significant variation between company regions. It is reasonable to expect associated cost differences;
- the base cost assessment models, in which runoff variables have explanatory power, support this narrative. This is particularly evident when runoff is interacted with the extent of combined sewer networks.

4.2.1 ENGINEERING ASSESSMENT

Drainage services provided by companies are not included in PR14 models. Although models do include volumetric drivers of sewage load, highway and surface water drainage services are not. Their variation between companies is not captured by any variables, despite evidence suggesting that the variation between companies is significant. This section explores the effect of drainage volumes on company costs and demonstrates the variation in company drainage volumes.

Impacts on business activities and costs

Higher rainwater runoff volumes in combined networks result in increased costs. They require capital investment in order to maintain and replace larger and more extensive pipework required to convey the flows, as well as the bigger storage volumes required to prevent larger and more frequent untreated spills from CSOs. The associated increase in pumping requirements resulting from increased average flows in sewers raises operating costs; the maintenance costs associated with the network, storage and pumps also increase.

The scale and type of drainage services required dictate drainage network costs. The volumes of flow into the network, the proportion of combined networks, the balance between urban and rural networks and the sensitivity of receiving waters into which networks spill all influence business activities and costs.

Urban runoff volume

is influenced by proportion of urban area and amount of rainfall received

The relationship between drainage and costs is generally supported by hydraulic modelling evidence. This indicates base costs vary by 13 to 17 per cent when dry weather flow (DWF) volumes pumped are varied by up to 25 per cent. The analysis assumed a baseline amount of DWF pumping (75 per cent) as a starting point, which was varied to simulate the additional flows due to drainage in the system. The costs of maintaining additional storage assets to accommodate drainage flows were not included in the modelling exercise but, from a separate simulation, would be expected to be substantial given the level of capex required to build them.

For the existing combined drainage systems to perform to the required levels of service, companies in areas of high runoff will have already invested in larger infrastructure - such as storage tanks and conveyance capacity – in order to meet equivalent service levels to companies in areas with low urban runoff. This is supported by analysis of base capex and operating costs (botex) reported by companies in the Ofwat 2016 datashare, which broadly aligns with the results of the modelling. The main drivers appear to be the size of network, allied to volumes of flow received, and the length of network per population, allied to urbanised or rural areas.

Rainfall and runoff analysis

Storm water flows (runoff) into combined networks are largely driven by rainfall on the impermeable surfaces of urban areas. Annual average rainfall received in a company's area is therefore an important consideration in determining the design capacity of combined sewer systems. However, peak flow often occurs in the rainier winter months of November, December and January, so winter average rainfall was collected as an alternative variable. Both annual average and winter rainfall vary significantly across England and Wales, with more rainfall received in the west than in the eastern and central areas.

Annual and winter average urban runoff metrics reveal the variance of storm inflows across companies. A measure of annual urban runoff volume (m^3) was produced by multiplying effective runoff (per km^2) by the total urban area in each company area. Analogous metrics were constructed for winter runoff, using a different approach and dataset to annual runoff. Further methodology details are noted in Appendix A.

Findings

Urban runoff volume is highest for companies with the highest proportion of urban area, but is also influenced by where the greatest amount of rainfall is received. There is a five-fold difference in annual urban runoff volume between Thames Water and South West Water (Table 12). Despite having one-and-a-half times the urban area of Welsh Water, Southern Water received about the same volume of annual urban runoff due as a result of receiving less annual rainfall. A similar comparison can be made for Anglian Water and Yorkshire Water.

The high frequency and long duration of winter rainfall events, which can force combined systems to run at peak capacity more frequently, makes them particularly important for base maintenance and replacement costs. Furthermore, peak flows last longer in the network, increasing pumping and treatment operational costs. The findings for winter runoff were broadly similar to those observed for annual runoff as shown in Table 13.

Monthly annual average and 3-month winter average runoff time series were generated for use in the econometric models discussed in 4.2.2. More details can be found in Appendix A.

WASTEWATER COMPANY	ANNUAL URBAN VOLUME (million m^3)	PERCENTAGE URBAN AREA (ONS 2011) VS TOTAL WASC AREA	AVERAGE ANNUAL URBAN RUNOFF (mm/yr)
Anglian	621	7.2%	392
Northumbrian	310	7.2%	529
Severn Trent	1,086	11.6%	503
South West	251	4.3%	676
Southern	525	11.6%	514
Thames	1,230	22%	459
United Utilities	1,155	12.5%	729
Welsh	565	3.8%	827
Wessex	381	7.9%	552
Yorkshire	618	9.7%	516

Table 12: Annual urban runoff volumes based on long-term standard annual average rainfall, 1941-1970

WASTEWATER COMPANY	WINTER VOLUME (million m^3)	URBAN AREA (ONS 2011) (km^2)	WINTER RUNOFF (mm/yr)
Anglian	183	1,632	114
Northumbrian	105	623	169
Severn Trent	276	2,148	132
South West	82	386	205
Southern	164	1,051	146
Thames	416	2,755	145
United Utilities	394	1,669	227
Welsh	160	690	224
Wessex	161	705	225
Yorkshire	190	1,164	156

Table 13: Winter runoff volumes based on long-term 3-month winter rainfall

MODEL	ESTIMATION METHOD	FUNCTIONAL FORM	SPECIFICATIONS TESTED
PR14+ network	GLS	Translog	Base, +annual urban runoff, +winter urban runoff
PR14+ network (5 year restricted)	GLS	Translog	Base, +combined annual runoff, +combined winter runoff,
PR14+ botex	OLS, GLS	Translog	Base, +annual urban runoff, +winter urban runoff
PR14+ botex (5 year restricted)	OLS, GLS	Translog	Base, +combined annual runoff, +combined winter runoff

Table 14: Drainage regression specifications

Note: 'Base' variables in PR14+ network models are: length, density, length^2, density^2, length*density, time, and regional wage, all of which are logged. 'Base' variables in PR14+ botex models are: density, load, density^2, load^2, load*density, per cent load in bands 1-3, time, and regional wage, all of which are logged.

4.2.2 ECONOMETRIC ASSESSMENT

Analysis

Modelling tested whether there was a statistically significant relationship between runoff drivers and costs. The monthly urban rainfall dataset was summed annually April to March to correspond to company reporting years, yielding annual and winter urban runoff metrics. These measures were interacted with combined sewer lengths, which were available from the 2016 Ofwat datashare for all companies beginning in 2011, to form metrics representing drainage conveyance in combined systems. All of the drainage variables were logged for analysis.

These drainage variables were tested using the specifications in Table 14. Given the narrative focus on network costs, treatment and sludge models were not included. Drainage coefficients were expected to have a positive sign: their consistency and significance were evaluated. Overall model performance was evaluated using the statistical tests described in Appendix B.

Runoff has **significant explanatory power over cost**

Findings

Runoff has positive, significant explanatory power over costs. Coefficients on both annual and winter drainage variables have a positive in sign and are significant with at least ninety per cent confidence. The sign and significance of the load and density coefficients, which feature as the primary scale variables in the PR14 models, proved to be quite stable to the addition of these drainage variables.

Table 15 displays the results of the PR14 test regressions for the botex OLS models. The table displays the results from adding annual and winter drainage variables to the PR14 network OLS models, as well as the effect of adding these variables interacted with combined sewer length to 'PR14+ 2011+', which is the PR14 model specification estimated using data for 2011-16.

	PR14+	ANNUAL DRAINAGE	WINTER DRAINAGE	PR14+ 2011+	COMBINED ANNUAL DRAINAGE	COMBINED WINTER DRAINAGE
Log load	8.52	8.39	8.84	4.06	1.64	1.94
Log density	21.68	19.22	19.61	-18.92	-39.94	-43.22
Log load²	0.15	0.15	0.15	0.20	0.16	0.17
Density²	1.44	1.74	1.84	5.08	6.48	7.03
Log load x log density	-2.62	-2.63	-2.73	-1.89	-1.16	-1.28
Year	0.02	0.02	0.02	-0.01	0.00	-0.01
Log % bands 1-3	0.13	0.14	0.13	0.06	0.13	0.11
Log wage	0.13	0.19	0.12	-1.04	-0.14	-0.32
Log annual runoff		0.17				
Log winter runoff			0.10			
Log combined annual runoff					0.14	
Log combined winter runoff						0.13
Log combined annual runoff²						
Log annual runoff x log load						
Log combined winter runoff²						
Log winter runoff x log load						
Constant	-103.3	-97.7	-100.9	16.0	72.5	78.5
N	100	100	100	50	50	50
N of companies	10	10	10	10	10	10
N of years	10	10	10	5	5	5
R^2	0.97	0.98	0.97	0.98	0.99	0.99

Table 15: PR14 Drainage test regressions – botex OLS.

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	PR14+	ANNUAL DRAINAGE	WINTER DRAINAGE	PR14+ 2011+	COMBINED ANNUAL DRAINAGE	COMBINED WINTER DRAINAGE
R^2	0.97	0.98	0.97	0.98	0.99	0.99
Variance inflation factor - maximum	21347	21408	21412	33923	35924	36379
Variance inflation factor - median	6118	2618	2622	8312	3486	3422
Ramsey RESET	Fail	Fail	Fail	Fail	Pass	Pass

Table 16: Statistical indicators of model performance largely unaffected by drainage tests

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

There is promise in interacting drainage variables with combined sewer length, though the shorter time series of data available for combined length limits its current usability. The interaction variables that were tested are positive in sign and highly significant when added to the PR14 model formulations.

However, a barrier to the variable's usage is the fact that industry-wide combined sewer length data is only published from 2011, so the reported results are based on a truncated time series from 2011-16. Even before the variable is introduced, the shortened time series that includes only 50 observations has a dramatic effect on the stability of the main drivers. Thus the interaction variables are not suitable for inclusion until a longer combined sewer length time series can be obtained.

Other measures of model performance were largely unaffected by the inclusion of drainage variables. Table 16 lists the test results for the botex OLS drainage test regressions.

Recommendations

There is strong evidence to support the use of drainage measures as the main drivers of economies of scale in benchmarking models. Figure 6 summarises the recommendations for this section. All drainage variables perform well, with the most significant coefficients on the log of combined annual runoff and combined winter runoff. There is strong support for including these in the main model; however, they are currently only available for a limited time period. Their use requires collecting data over a longer period of time.

TYPE	RECOMMENDATIONS
Extend data collection	Collect combined sewer length data over a longer period of time.
Include additional variable	Combined annual runoff and combined winter runoff both perform well in the extended models. With additional data they could be included in the main model.

Figure 6: Recommendations based on analysis of drainage

4.3 ECONOMIES OF SCALE

There is strong evidence that economies of scale exist at treatment works. The unit costs of wastewater treatment are dependent on asset characteristics, in particular the size of the works. The distribution of populations is a primary influence in the number and sizing of wastewater treatment works. This section presents the engineering and econometric evidence for economies of scale in wastewater treatment costs, showing that:

- statistics on works size bandings capture the effect of economies of scale at treatment works;
- economies of scale are most important at the level of treatment works assets. The PR14 models use company-level scale instead; and
- economies of scale and urbanisation seem to be closely linked, with evidence of the former strengthening in the presence of the latter.

4.3.1 ENGINEERING ASSESSMENT

Economies of scale
are most important at the level of treatment works assets

Impacts on business activities and costs

The distribution of population and settlements in wastewater company service areas influences the location and size of wastewater assets. This directly affects the practicality and cost-effectiveness of centralising treatment facilities and taking advantage of potential economies of scale.

The costs associated with maintaining equitable outcomes differ between sparse rural and urban areas for reasons relating to accessibility, logistics and economies of scale. The impact of rural and urban classifications on economies of scale is discussed in more detail in Section 4.4.

Analysis

The distribution of wastewater treatment works was analysed using the Ofwat size bands 1-6. The number of works and the load treated in each size band were analysed in order to determine the impact of economies of scale on cost. The resulting company-level variation is shown in Figure 7 (number of assets) and Figure 8 (load treated).

As the distribution of assets is a factor of the spatial profile, Figure 9 presents an analysis of treatment capacity by rural urban classification (RUC) split. This is discussed further in Section 4.4.

Industry cost data provided in Ofwat’s 2016 datashare and previous June Return reports does not allow for a reliable analysis of the treatment cost variation between urban and rural settings.

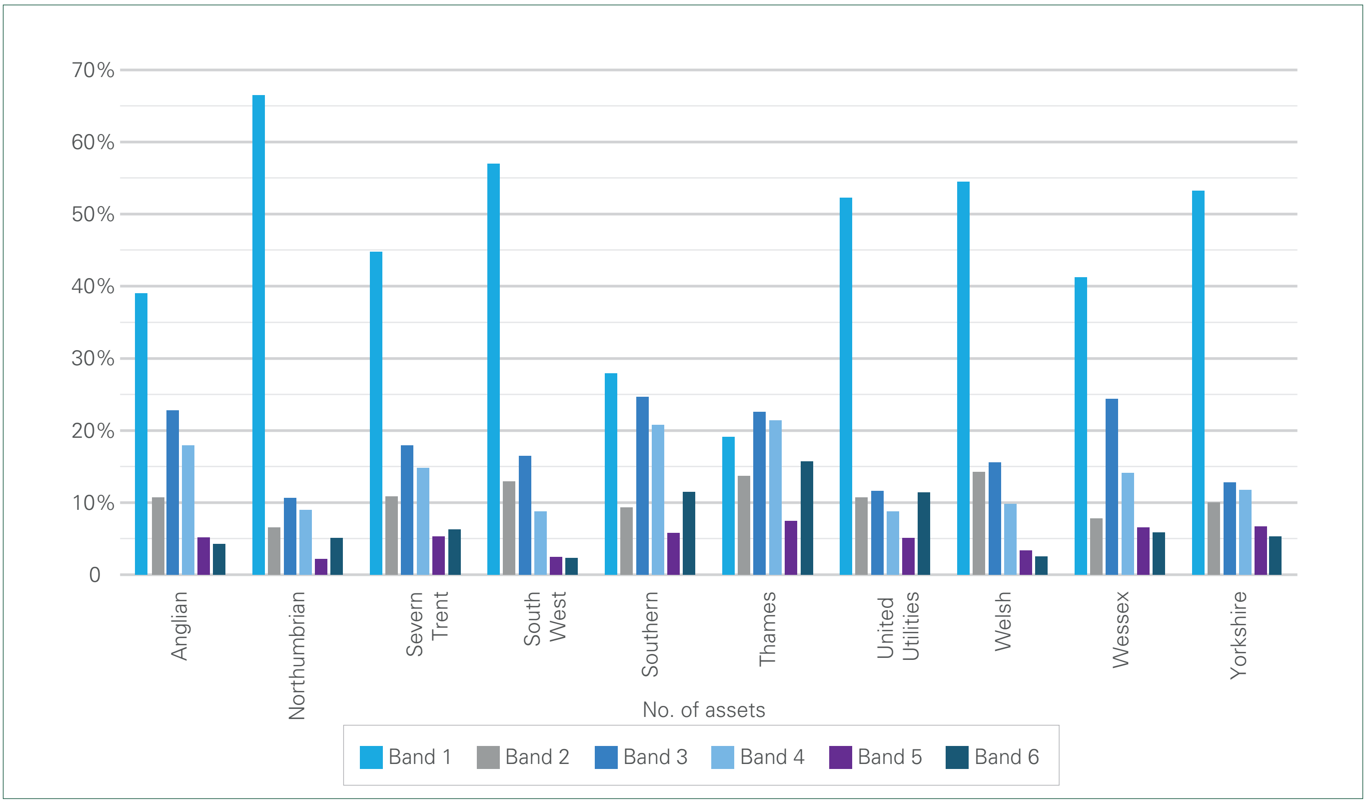


Figure 7: Shares of assets by size band, 2015/16

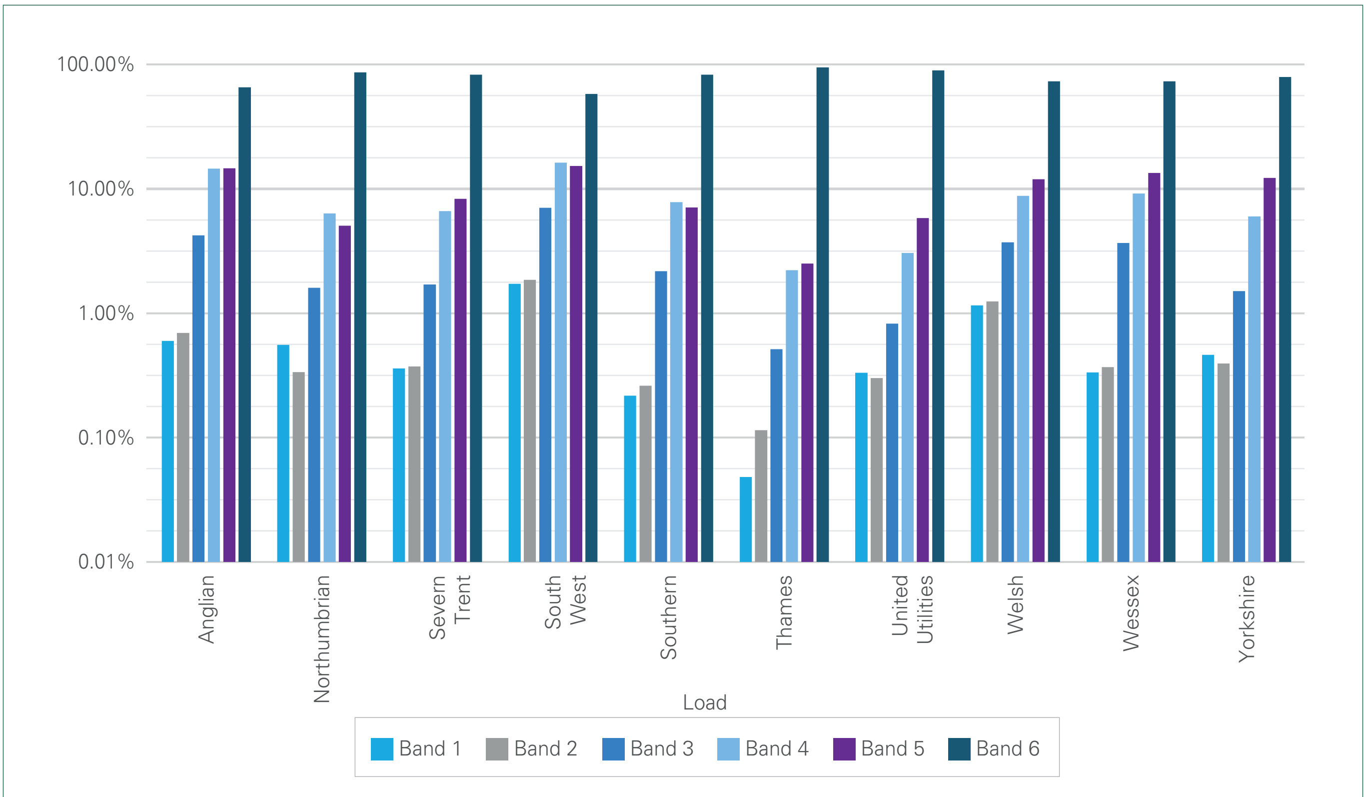


Figure 8: Share of load treated by band, 2015/16

Treatment works in bands 1-3 are associated with the

highest unit costs

Findings

Figure 9 demonstrates that there is variation in the percentage of company assets located within urban and rural areas in terms of treatment capacity. This urban-rural split has been analysed against the results shown in Figure 7 and Figure 8. The comparison shows that companies with a higher-than-average percentage of treatment capacity in rural areas generally show a corresponding higher-than-average proportion of band 1-3 assets, both in terms of number of assets and load treated. In other words, the more rural the company area, the higher the reliance on smaller treatment works. This indicates that the profile of population settlements has an impact on a water company's ability to utilise efficiencies offered by economies of scale.

The greatest proportion of load treated is in Band 6 (58 to 95 per cent), but there is significant variation between companies. Percentage load treated in bands 1-3 is generally low, at 7 per cent or less, yet the number of assets in these bands is highest. Figure 10 shows unit costs for wastewater works ranging from very small (~250 PE) and small (~1,000 PE) to large (50,000+ PE). The data shows that works in the lowest bands (predominantly PE <2, 000 or bands 1-3) are associated with the highest unit costs.

It is widely recognised that the unit costs of treating wastewater in larger plants, are lower than smaller plants due to efficiency savings, in particular energy, which accounts for a significant proportion of operating costs. As Figure 10 shows, works with PE greater than 2,000 (bands 4–6) have much lower unit costs than bands 1-3.

The unit cost analysis shown in Figure 10 further suggests that the unit costs of treating a given load at larger works (25,000 PE or greater) are less than a third those of very small works (250 PE). This is reinforced by a recent study[2] which draws a similar conclusion, and puts this ratio at less than a fifth. Both findings, therefore, support the case for further focus and detail on treatment costs associated with treatment works in size bands 1-3. Finally, significant variation is observed between companies, in terms of both the shares of assets by size band and the share of load treated in each size band.

Asset distribution time series data was compiled for the period 2006/07 to 2015/16 using historical June Return data and the 2016 industry datashare. The time series data was used for the econometric modelling.

[2] Manning, L.J., Graham, D.W. and Hall, J.W., Wastewater Systems Assessment, 2015

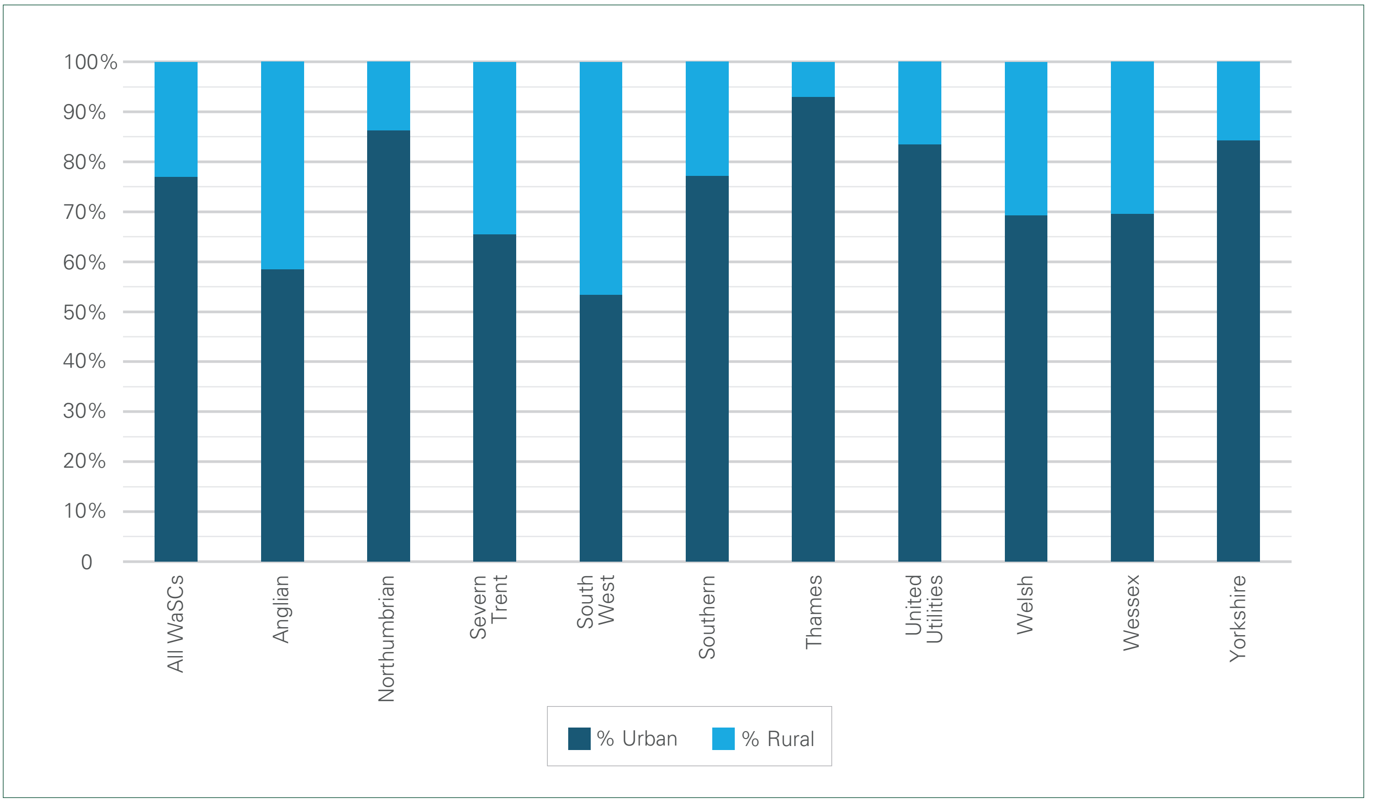


Figure 09: Treatment capacity by rural urban split, 2011

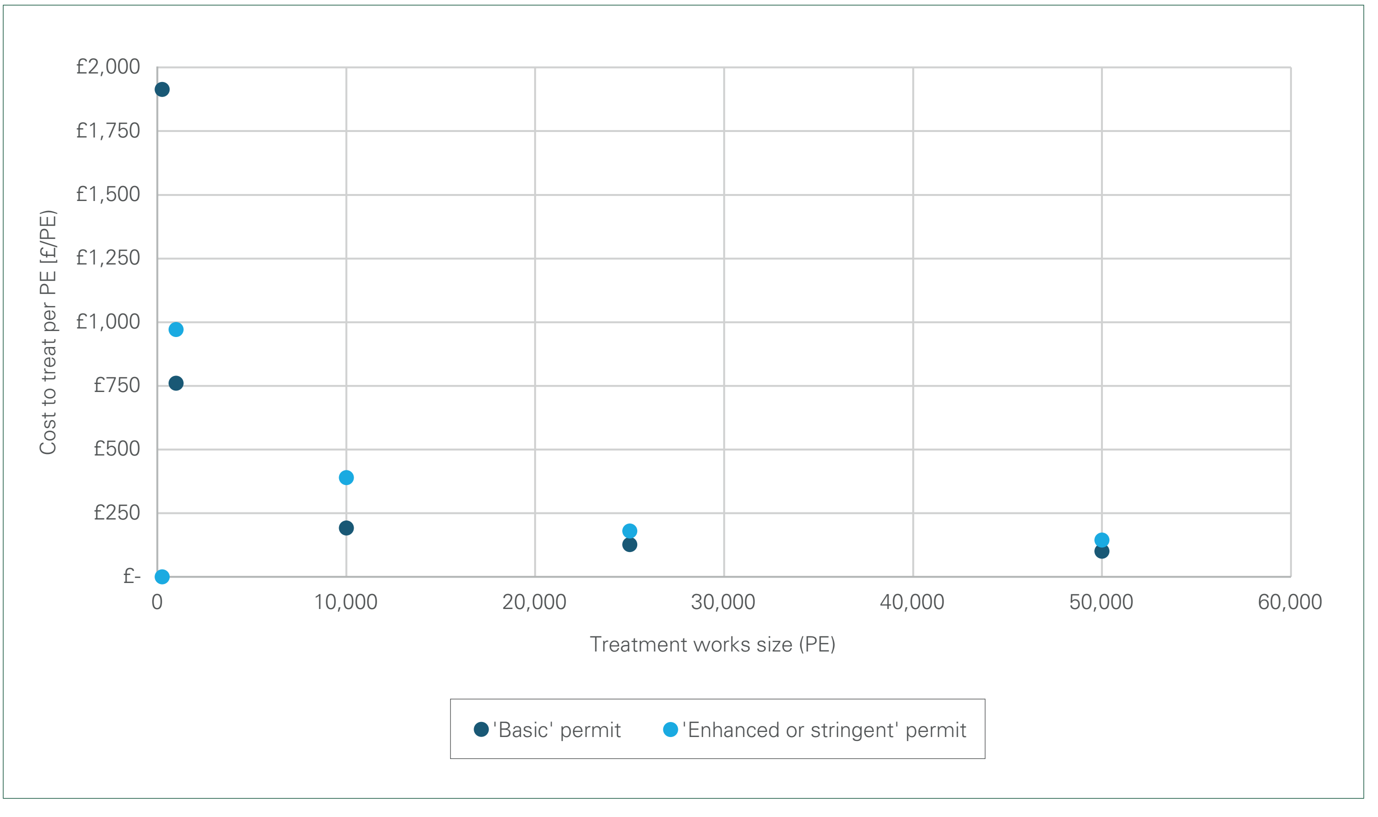


Figure 10: Unit cost to treat in £/PE, for basic and enhanced treatment permits (consents)

4.3.2 ECONOMETRIC ASSESSMENT

Analysis

Though PR14 models attempt to control for economies of scale, for the most part they use company-level variables for which there is weak statistical support. In principle, the relationship between costs and these variables such as load-squared and density could reflect managerial economies that are available to large companies but not to smaller ones. However, the results of the PR14 models do not provide compelling evidence that such relationships hold: as explained in Section 3, coefficient values for load squared (a translog term) and density are not consistent with economies of scale, while density coefficients are generally insignificant.

This analysis focuses on the use of asset-level data where the motivation for economies of scale is much clearer. Consistent with the engineering narrative set out in 4.3.1, the analysis measures the availability of economies for any company by considering its treatment assets in size bands 1-3. This grouping is consistent with that previously used in PR14. To analyse the impact of economies of scale, disaggregated information published for PR14 on assets and loads treated by works in different treatment bands is used along with Ofwat datashare information, which includes the same data for 2015-16. By interpolating data for 2013-14 and 2014-15 (see Appendix B), the study constructed time series on the percentage of total load treated in works in size bands 1-3 and the percentage of assets within each size band for each company. It considered logged and unlogged versions of each time series.

Geographical variables were not considered viable. The main downside of using asset-level data is endogeneity, as companies exercise some control on the configuration of their treatment assets. In principle, this could be mitigated through the use of a variable capturing geographical factors that determine the availability of economies of scale, such as population density, topography, ground conditions, and asset inheritances. However, in practice there is no transparent way of distilling all of the exogenous geographical factors into a manageable set of factors, so this was not pursued.

Variables were tested in both Cobb-Douglas and translog versions of all of the PR14+ models (Table 17). It was expected that higher proportions of load or assets in the smaller bands would result in increased costs, resulting in positive regression coefficients for each of the four variables. Overall model performance was evaluated using the statistical tests described in Appendix B.

Findings

Percentage of load treated in bands 1-3 consistently has significant explanatory power. The variable continued to perform well in PR14's botex models but also demonstrated positive and significant coefficients when added to the treatment and sludge models. Per cent load outperformed the per cent assets variables, suggesting that costs depend more on work done at high unit-cost works rather than on the number of such works. Unlogged versions seemed to outperform the original logged variables statistically and other primary PR14 driver coefficients were stable to the addition of either variable. Table 18 displays the results of the PR14 test regressions for the treatment and sludge OLD models.

MODEL	ESTIMATION METHOD	FUNCTIONAL FORM	SPECIFICATIONS TESTED
PR14+ treatment and sludge	OLS, GLS	Translog, Cobb-Douglas	Base, +log per cent load in bands 1-3, +log per cent assets in bands 1-3, +per cent load in bands 1-3, +per cent assets in bands 1-3
PR14+ botex	OLS, GLS	Translog, Cobb-Douglas	Base, +log per cent load in bands 1-3, +log per cent assets in bands 1-3, +per cent load in bands 1-3, +per cent assets in bands 1-3

Table 17: Economies of scale regression specifications

	PR14+ (TRANSLOG)					COBB-DOUGLAS				
	BASE	LOG % LOAD	% LOAD	LOG % ASSETS	% ASSETS	BASE	LOG % LOAD	% LOAD	LOG % ASSETS	% ASSETS
Log load	11.7	10.5	13.7	10.3	11.2	0.76	1.00	0.95	0.87	0.86
Log density	60.3	65.2	74.1	47.1	55.2	-0.11	0.69	0.40	0.02	-0.06
Log load^2	0.16	0.13	0.01	0.16	0.16					
Density^2	-1.90	-3.19	-4.11	-0.87	-1.50					
Log load x log density	-3.47	-2.96	-3.04	-3.12	-3.33					
Year	0.00	0.01	0.01	0.01	0.01	0.00	0.00	-0.01	0.03	0.02
Log wage	0.75	1.02	1.16	0.89	0.82	0.38	0.46	0.16	2.17	1.89
Log % load in bands 1-3		0.14					0.26			
% load in bands 1-3			4.53					5.37		
Log % assets in bands 1-3				0.28					1.49	
% assets in bands 1-3					0.17					1.82
Constant	-207	-211	-251	-170	-193	-5	-11	-9	-12	-12
N	100	100	100	100	100	100	100	100	100	100
N of companies	10	10	10	10	10	10	10	10	10	10
N of years	10	10	10	10	10	10	10	10	10	10
R^2	0.95	0.95	0.96	0.95	0.95	0.90	0.92	0.92	0.92	0.92

Table 18: PR14 economies of scale test regressions – treatment and sludge OLS

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	PR14+ (TRANSLOG)					COBB-DOUGLAS				
	BASE	LOG % LOAD	% LOAD	LOG % ASSETS	% ASSETS	BASE	LOG % LOAD	% LOAD	LOG % ASSETS	% ASSETS
R^2	0.95	0.95	0.96	0.95	0.95	0.90	0.92	0.92	0.92	0.92
Variance Inflation Factor - Maximum	21180	21344	21939	40800	34509	2.15	5.47	3.41	6.80	6.10
Variance Inflation Factor - Median	9274	6133	7171	9919	8295	1.49	2.15	2.20	2.86	2.71
Ramsey RESET	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail

Table 19: Statistical indicators of model performance are largely unaffected by economies of scale tests

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

Measures of model performance were largely unaffected by the inclusion of economies of scale variables. Table 19 lists the test results for the treatment and sludge OLS drainage test regressions.

Recommendations

There is strong evidence to support the use of asset level measures as the main drivers of economies of scale in benchmarking models. In particular, this study supports the use of the unlogged percentage of total load treated in works in size bands 1-3. The variable coefficients tended to be positive in sign, which is consistent with the engineering narrative, and statistically significant while the recommended variables left the main drivers unchanged. Variables based on geography, though in principle more strongly exogenous, are not considered viable.

Modelling will be improved by consistently measured, quality assured data. Problems with the quality and consistency of load data are discussed in more detail in Section 5.2. Issues raised in that section also apply to more granular information on the share of load treated at works in different size bands.

TYPE	RECOMMENDATION
Extend data collection	Ensure that load treated by band, which is already a primary driver crucial to the accuracy of PR14 models, is consistently collected for all years.
Include additional variable	Percentage of load treated in bands 1-3 should be included in both the treatment and sludge as well as the botex models.
Change modelling approach	Variables that are restricted to a range of 0 to 1 should not be logged due to the difficulty of interpretation.

Figure 11: Recommendations based on analysis of economies of scale

4.4 URBANISATION

Density of populations served influences the costs of service delivery through the location of assets in dense or sparse areas. The location of populations and settlements in wastewater company service areas determines where wastewater assets need to be located. Populations living in sparse areas must have the same standard of wastewater services as large, centralised populations in urban areas. This results in capital and operating costs differing substantially between the two demographics. Overall, the weight of evidence points to the total costs of networks and wastewater treatment being higher in urban areas; for sparse areas the picture is more mixed and would benefit from further evidence. This report presents the evidence that:

- networks and assets within densely-populated, urban areas give rise to a different operating environment and associated operating costs to those in rural, sparse areas;
- the costs associated with maintaining equitable outcomes differ between sparse rural and urban areas, for a variety of reasons relating to accessibility, logistics and economies of scale. Despite the cross-over, the arguments for urbanisation and sparsity are distinct from the arguments for economies of scale (Section 4.3);
- differences are evident in the number of treatment assets in both urban and sparse areas, but those in urban areas appear to have the most evident impact on costs;
- engineering-backed variables such as the number of a company's assets in urban areas, the percentage of its capacity in urban areas, or proportion of company area that is urban can be used to express the impact of urbanisation; and
- the cost effects of urbanisation and sparsity can be modelled using one of the engineering variables captured above.

Rural or urban context

is an important factor in the cost of wastewater services

4.4.1 ENGINEERING ASSESSMENT

Impacts on business activities and costs

The profile of population density and rural or urban context is an important factor when considering the cost of delivering wastewater services. The expenditure profile for one company which has a proliferation of small, remote treatment plants serving dispersed communities will differ from another with more densely populated, centralised assets. This is due to: the opportunities for economies of scale (for instance, distance is a key factor in the cost effectiveness of centralisation of treatment) and travel and transportation costs, including indirect costs such as lost time due to travel. For remote sites in rural, sparse areas, this can be due to additional travel time between sites. In urban towns and cities, this can be due to traffic congestion. This has an impact on operation and maintenance resource efficiency, sludge treatment and disposal costs amongst others. On balance, the total costs associated with urban areas are expected to be more significant.

A higher number of connections per length of sewer correlates with higher density areas. Major built-up areas (BUAs, as defined by ONS in the 2011 census) statistically have higher average household sizes than non-BUAs and smaller BUAs, resulting in higher wastewater quantities per household (per connection). BUAs also tend to have larger paved or impermeable areas than non BUAs, resulting in higher runoff volumes per unit area. A higher number of connections per kilometre (that is, higher density) influences costs associated with urbanisation as follows:

- Large diameter sewers in densely populated areas have higher associated costs to maintain and operate, higher capital costs per kilometre, larger plant and equipment.
- Greater storage volumes and more CSOs are required in urban areas to accommodate larger surface runoff volumes (see Section 4.2). These also cost more to operate and maintain.
- Sewer network congestion with other utilities in urban areas affects the ease of maintenance and operation. Sewers are often deeper than in less densely populated areas due to the presence of other utilities.
- Opportunities to optimise sewer runs are limited in BUAs, so they are mainly in roads and highways. This leads to additional installation and reactive maintenance time, logistics and costs to access pipework located in roads and built-up areas requiring permitted access periods.
- Higher frequency of blockage or failure due to higher flows, age of sewers and higher loading. Additional logistics for maintenance associated with congested, shared utilities corridors.

There is a correlation between population density and proximity of catchment to treatment works. Land costs and low availability in high-density major BUAs pushes treatment works further away from the population, imposing additional costs to transfer wastewater.

In addition when tighter discharge permits and higher trade flows occur in densely populated areas, they result in more advanced wastewater and odour treatment processes, and hence impact on cost, to handle higher pollutants and potentially tighter permits.

Sparsity or remoteness influences wastewater network costs in multiple ways. The lower proportion of paved areas and highways in sparse areas results in lower volumes of surface runoff to be accommodated in combined drainage networks. In a similar vein, lower average household sizes[3] in minor and small BUAs and non-BUAs - which are typical of sparse areas - result in lower wastewater quantities to be conveyed and treated. Wastewater networks in sparse areas comprise smaller diameter sewers, plant and equipment, and typically have lower associated operating and replacement costs than in urban areas. Finally, there are typically fewer CSOs and less storage requirements in sparse areas due to the lower surface runoff volumes in sewerage compared to urban areas, although on a per asset basis, there are higher costs that arise due to remoteness and distance from the supply chain. More granular cost and location data is required in order to assess prevalence, for example broken down by areas classed as sparse.

Sparse areas tend to require smaller capacity treatment works due to the smaller foul and runoff flows received, and have smaller, less sophisticated plant or equipment with lower maintenance and replacement costs. However, the downside is that routine maintenance, operational and replacement activities have a higher unit cost component due to remoteness and accessibility. Discharge permit data[4] analysis demonstrates that wastewater treatment plants in rural areas have lower average permit levels for phosphorus (P) and ammonia (NH_3) than urban areas (Figure 4 in Section 2 helps to illustrate this variation).

It is recognised that the picture is mixed, as the lower incoming loads at rural works, and the receiving water course size and ecological condition also exert an influence. Lower quantities of screenings and sludge are produced for disposal; reduced frequency of disposal activities leads to decreased overall cost. However, the additional costs associated with remoteness and accessibility could outweigh the benefit. Industry cost data provided in Ofwat's 2016 datashare and previous June Return reports does not allow for a reliable analysis of the treatment cost variation between dense and sparse conurbations, although existing engineering evidence would suggest that the costs associated with urbanisation are more significant.

[3] "2011 Census: Characteristics of Built-Up Areas" by ONS states that the average household size in Major Built-Up Areas is 2.44, compared with 2.39 in non-Built-Up Areas.

[4] For phosphorus and ammonia only.

Analysis

Analysis of the population and geographical profile within which each wastewater company's assets are located was carried out to identify any material variation in terms of urbanisation and sparsity. This was done by comparison of data by company to establish the ranking of the impact of sparsity or urbanisation on the operational environment profile by identifying assets located within major conurbations and those within 'sparse settings', as defined by DEFRA's 2011 Rural-Urban Classifications (RUC).

According to ONS methodologies, the profile of the geographic and settlement landscape of England and Wales is based on the eight classifications shown in Figure 12.

The analysis focused on two key lines of enquiry: first, the proportion of the company service areas classed as urban, using population densities from the 2011 census, reported at local authority level. This was averaged out across the authority boundary and compared with the percentage of assets within urban areas by treatment capacity.

The second line of enquiry was the respective percentages of assets in sparse areas and urban (major conurbation) areas. The analysis tabulated the proportion of wastewater assets by both number and population equivalent and by small, medium and large asset band (<2,000, 2,000-25,000, >25,000, selected for ease of comparison) to establish the distribution of works as an indication of potential economies of scale.

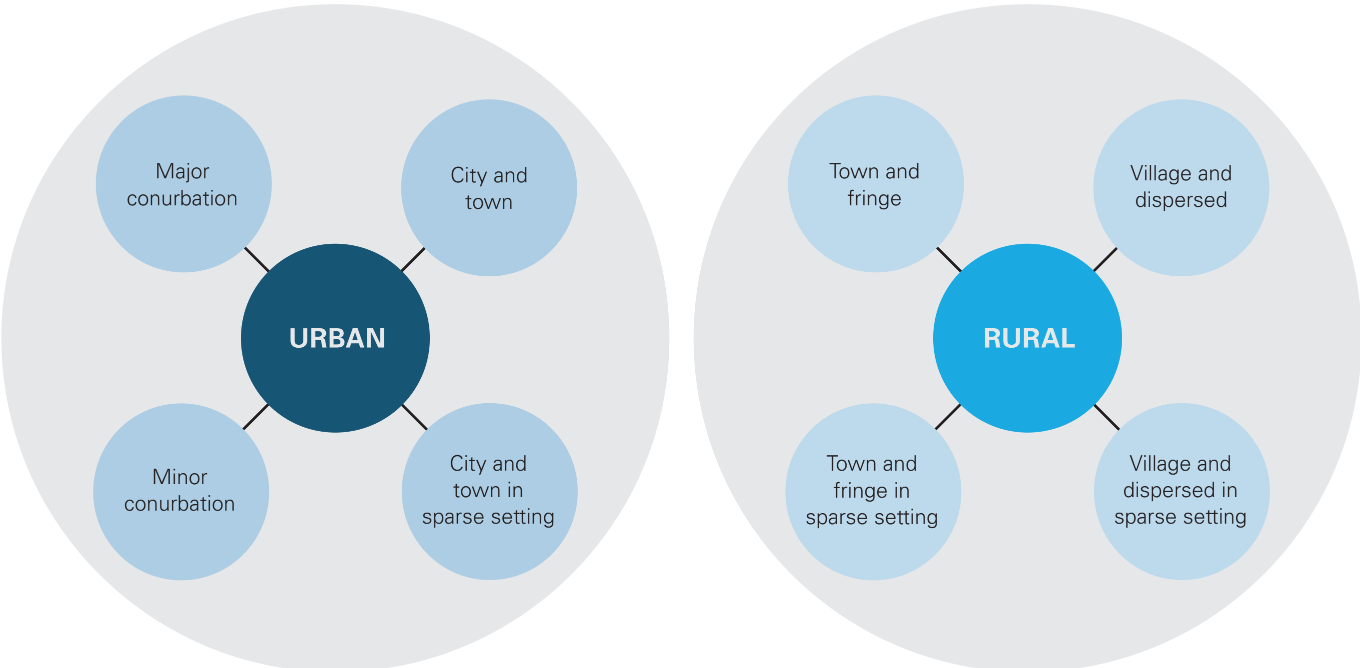


Figure 12: ONS rural and urban classifications

Findings

As has been shown in Section 4.2.1, the percentage of wastewater company service area located in urban areas varies significantly, as does the percentage of works capacity in urban areas, shown in Table 20. The industry average stands at 77 per cent. Those with the highest percentage of works capacity significantly above the average are Thames, Northumbrian, Yorkshire and United Utilities. The lowest is South West at just over 53 per cent.

Corresponding analysis of the number of assets in sparse areas shows that the proportion is significant and high in the north and southwest. The proportion ranges from 14 per cent to 31 per cent. The remaining areas are at 3 per cent or less.

COMPANY	WORKS CAPACITY (m^3)	URBAN WORKS CAPACITY (m^3)	% URBAN CAPACITY
Anglian	8,546,508	4,997,273	58.5%
Northumbrian	3,412,817	2,943,254	86.2%
Severn Trent	11,154,289	7,310,199	65.5%
South West	1,951,396	1,042,192	53.4%
Southern	4,545,355	3,507,462	77.2%
Thames	16,221,729	15,086,306	93.0%
United Utilities	9,454,366	7,888,628	83.4%
Welsh	4,081,807	2,829,474	69.3%
Wessex	3,905,124	2,714,883	69.5%
Yorkshire	5,562,897	4,691,580	84.3%
All WaSCs	68,836,288	53,011,251	77.0%

Table 24: Proportion of treatment works capacity in urban areas

4.4.2 ECONOMETRIC ASSESSMENT

Analysis

Measures of urbanisation were tested in models similar to PR14. PR14 models attempt to capture the effect of urbanisation using a company-wide average density variable. However, the cost effects of urbanisation are better captured using one of the engineering variables discussed in 4.4.1: the percentage of a company's assets in urban areas, the percentage of its capacity in urban areas, or percentage of company area that is urban. These variables were tested in their logged and unlogged forms in PR14's treatment and sludge and botex models, similar to the models used to test the role of drainage and economies of scale. Consistent with the methodology used to evaluate the original models, coefficients were checked for conformance to engineering theory, statistical significance, and impact on overall model performance. Coefficient results are reported in Table 21.

Sparsity variables were also tested, but strong correlations with urbanisation, quality and economies of scale make measurement difficult. This is picked up in Section 6 on new regressions, but more work is required to identify the most appropriate measure of sparsity.

	BOTEX GLS			BOTEX OLS		
	PR14+	% URBAN AREA	% URBAN CAPACITY	PR14+	% URBAN AREA	% URBAN CAPACITY
Log load	3.85	-1.84	-5.59	8.52	7.02	8.51
Log density	1.1	-25.5	-32.7	21.7	17.0	19.1
Log load^2	0.10	0.00	0.15	0.15	0.18	0.13
Density^2	1.83	2.27	3.05	1.44	1.73	1.57
Log load x log density	-1.27	0.50	0.52	-2.62	-2.45	-2.50
Year	0.03	0.02	0.03	0.02	0.02	0.02
Log wage	0.32	0.22	0.35	0.13	0.25	0.04
Log % bands 1-3	0.20	0.68	0.68	0.13	0.09	0.07
% urban area		12.9			-1.29	
% urban capacity			3.72			-0.30
Constant	-29.4	69.4	105.9	-103.3	-84.3	-97.2
N	100	100	100	100	100	100
N of companies	10	10	10	10	10	10
N of years	10	10	10	10	10	10
R^2				0.97	0.97	0.97

Table 21: PR14 urbanisation test regressions – botex models

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	BOTEX OLS		
	PR14+	% URBAN AREA	% URBAN CAPACITY
R^2	0.97	0.97	0.97
Variance Inflation Factor - Maximum	21347	27587	22413
Variance Inflation Factor – Median	6118	7618	4302
Ramsey RESET	Fail	Fail	Fail

Table 22: Statistical indicators of performance are largely unaffected by urbanisation

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

Findings

Urbanisation variables are sometimes statistically significant, though this is sensitive to specification. Results of the GLS and OLS botex models, presented in Table 21, are markedly different from each other. In GLS specifications, load is insignificant and sometimes even negative while percentage treated in small works is significant and positive; in the OLS specification, the reverse is true. Similarly, urbanisation variables are positive and highly significant in the GLS model, but insignificant and negative using OLS. These ambiguous results are consistent with those obtained in the treatment and sludge models and the Cobb-Douglas versions of both model sets.

Statistical ambiguity despite a clear engineering narrative for their inclusion reflects the complicated relationship urbanisation has with other variables. The instability described above in part reflects collinearity between urbanisation, economies of scale, and the PR14 translog terms. It may also be related to missing quality drivers, as large treatment works in urban areas often treat wastewater to a high standard in part because of the volume of effluent they must discharge. A further difficulty with the data series is the lack of variation over time, which may lead urbanisation drivers to behave like company fixed effects. That makes it difficult to disentangle their effects on cost from one another statistically. Table 22 shows the results of statistical tests, which are not improved by the specifications tested in this section.

A further difficulty concerns the way in which urbanisation is measured. The drivers proxy for the additional costs associated with operating and maintaining network and treatment assets in urban areas. But the urban area contained within company regions does not measure these assets directly and can be skewed by rural areas in which companies have few assets. The share of treatment capacity within urban areas, though more directly connected with assets, may not reflect network costs where treatment assets that serve cities are located outside urban areas.

Urbanisation drivers should be included in benchmarking models, though more appropriate metrics could be developed. With clear engineering evidence, but ambiguous statistical conclusions, urbanisation should be included in the analysis and treated with care. New regressions discussed in Section 6 pick up on this discussion, and go further in exploring the intricacies of interrelated variables. Figure 13 summarises the recommendations in this section.

TYPE	RECOMMENDATION
Extend data collection	Different measures of urbanisation should be considered, and a time series generated
Include additional variable	Urbanisation should be included in PR19 models due to a clear engineering reason for its presence

Figure 13: Recommendations on urbanisation

SECTION 05

Improved data collection

Cost assessment using econometric models relies on data that accurately represents causal factors consistently across companies.

This section focuses on three factors in which evidence gaps or reliability concerns exist: quality, BOD load measurement, and regional wages. The recommendations motivated in this section would alleviate constraints to both modelling and the wider cost assessment process.

An evidence gap also exists for a fourth factor, sludge (bio-resources), for which a new price control will be developed. Available land bank for sludge disposal is the base dataset for any proposals. Preliminary analysis on land bank and data needs for robust analysis are presented briefly in this section.

5.1 QUALITY

Treatment technology choices are closely linked to the quality of treatment required by permit levels. The permits, of which the most common are ammonia (NH_3), phosphorus (P) and biological oxygen demand (BOD) load, are imposed by the Environment Agency and Natural Resources Wales, and relate to the sensitivity and quality of the receiving water body and the capacity of the treatment works being regulated. This section presents evidence from publicly-available permits data for England and Wales, although the information for Wales is very limited. It finds that:

- permits for NH_3, P, and BOD determine treatment quality required, which affects choice of treatment technology;
- thresholds of approximate or typical population equivalent (PE) exist at which the technology used to achieve permit levels changes;
- engineering assessment shows indicative unit costs increasing by 28 to 47 per cent as permits become tighter;
- available land and plant footprint influence the treatment technology choice and in some cases is the deciding factor; and
- historical (legacy) treatment processes at a site also influence process selection.

28-47%

increase in costs as permits tighten

Analysis of load by secondary or tertiary treatment

Treatment quality can be analysed by load undergoing secondary or tertiary treatment, but such information may not be sufficiently exogenous for regulatory purposes. In order to assess the link between technology and cost, current and historical regulatory data for all companies was collated by treatment works size and treatment type for the period 2007/08 to 2015/16. The dataset showed evidence of significant differences across the industry in the extent of secondary and tertiary treatment carried out. These variables were tested in econometric models, but were not considered to be sufficiently exogenous to be reliable. This issue is discussed in Section 6 on new regressions.

Analysis of Permit Variation

Modelling was conducted to understand variation in treatment technology by works size and permit levels. In order to assess the impacts of exogenous factors, generic treatment works were developed, characterised by representative treatment equipment and processes. Treatment works capacities were grouped by population equivalent (PE) served into 'very small' (250 PE), 'small' (1,000 PE), 'medium small' (10,000 PE), 'medium large' (25,000 PE) and 'large' (50,000+ PE). These five categories of treatment works asset sizes span a range of technologies. Permits were grouped by tightness as basic, enhanced or stringent.

Figure 14 illustrates the number of assets by permit level, while Table 23 notes the treatment types by size and permit. For ammonia permits, a permit level of more than 3 mg/l is regarded as not significant in terms of threshold for treatment technologies and cost. Therefore a 'basic permit' plant has no ammonia removal requirement whereas an 'enhanced permit' plant has an ammonia permit requirement of 3 mg/l or less.

Brief observations on permit type and treatment are noted in Table 24.

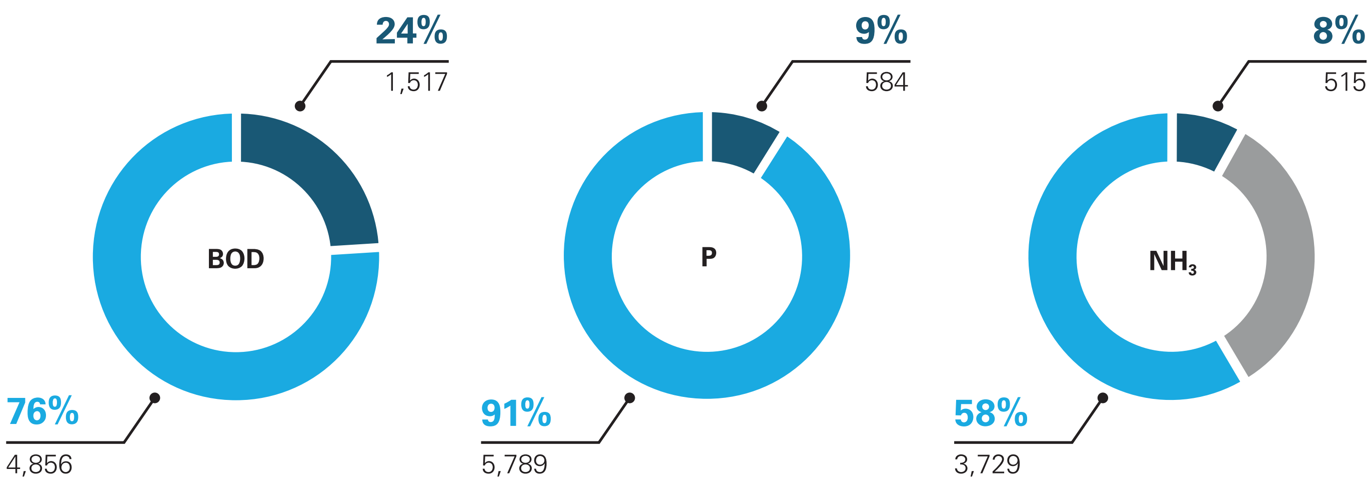


Figure 14: Number of England and Wales wastewater works by permit levels

Key:

- Number of works with a basic permit:
  - > 20 mg/l BOD5;
  - No permit for P;
  - No permit for NH_3-N.
- Number of works with enhanced or stringent permits:
  - ≤ 20 mg/l BOD5;
  - ≤ 1 mg/l P
  - ≤ 3 mg/l NH_3-N
- Other
  - ≥ 3mg/l

TREATMENT WORKS SIZE	BASIC PERMIT 20 MG/L BOD5	ENHANCED PERMIT 20 MG/L BOD5, 3 MG/L NH_3-N	STRINGENT PERMIT 10 MG/L BOD5, 3 MG/L NH_3-N, 1 MG/L P
Very small (250 PE)	(Carbonaceous) trickling filters	Package plant submerged aerated filter (SAF)	Package plant SAF with sand filter and dosing.
OR	(Carbonaceous) trickling filters	Nitrifying trickling filters and tertiary nitrification	As enhanced + chemical dosing and tertiary solids removal
Small (1,000 PE)	(Carbonaceous) trickling filters	Nitrifying trickling filters and tertiary nitrification	As enhanced + chemical dosing and tertiary solids removal
Medium (25,000 PE)	(Carbonaceous) trickling filters	Nitrifying trickling filters and tertiary nitrification	As enhanced + chemical dosing and tertiary solids removal
Large (50,000 PE)	Carbonaceous activated sludge process (ASP)	Nitrifying ASP	As enhanced + chemical dosing and tertiary solids removal

Table 23: Typical treatment technology used to meet discharge permits

PERMIT TYPE	OBSERVATIONS
Load (BOD5) removal	- Trickling filters (TF) commonly used for basic permits, and up to 50,000 PE; activated sludge process (ASP) used for greater than 50,000 PE. - Increase in plant size and/or permit level may require an extension of existing treatment processes. - Treatment technology choice is mostly dependent on site footprint. Where space is unavailable at large works, costly ASP processes are used.
Phosphorous (P) removal	- Removal process is the same as BOD5; can also be managed with TF or ASP. - Choice of technology (TF of ASP) is linked to existing site footprint. - If P permit level is increased ferric dosing might be introduced to increase P removal efficiency. - Treatment technology change threshold is quite low, at 250 PE (package plants).
Ammonia (NH_3) removal	- Treatment technology typically changes thresholds of 1,000 PE and greater than 50,000 PE. - Link with BOD removal, which occurs before NH_3 conversion. - NH_3 removal is driven by BOD removal capacity, available space and site legacy.

Table 24: Observations on permit type and levels

Findings

Technology choices, determined in part by permit levels, have a substantive impact on costs. By tightening basic permits to an enhanced or stringent level, cost increases by varying amounts, depending on the change in technology and the treatment works capacity. Sample industry data suggests a significant cost variation by works capacity.

Results of the analysis are shown in Figure 15 and summarised as follows.

up to
28%
Indicative cost increases observed for small (1,000 PE) works

up to
106%
indicative cost increases observed for medium small (10,000 PE) works a large increase, reflecting changes in technology

up to
44%
indicative cost increases observed for medium large (25,000 PE) works

up to
47%
indicative cost increases were observed for large (50,000 PE) works

The costs above represent replacement capex investments for replacement treatment works, constructed using typical industry costing approaches. The sample data used, however, is from a less efficient wastewater company, and therefore may represent inefficient (worst case) costs.

Figure 16 summarises the recommendations in relation to quality. The works-level costs used for the above analysis were based on a single company, as these are not available through the regulatory data. The analysis could be made more robust through the publication of industry-wide costs by treatment asset (works) in order to improve assessment of the impact of discharge permits across the industry. **The study recommends extending the permit data to works-level and including costs.**

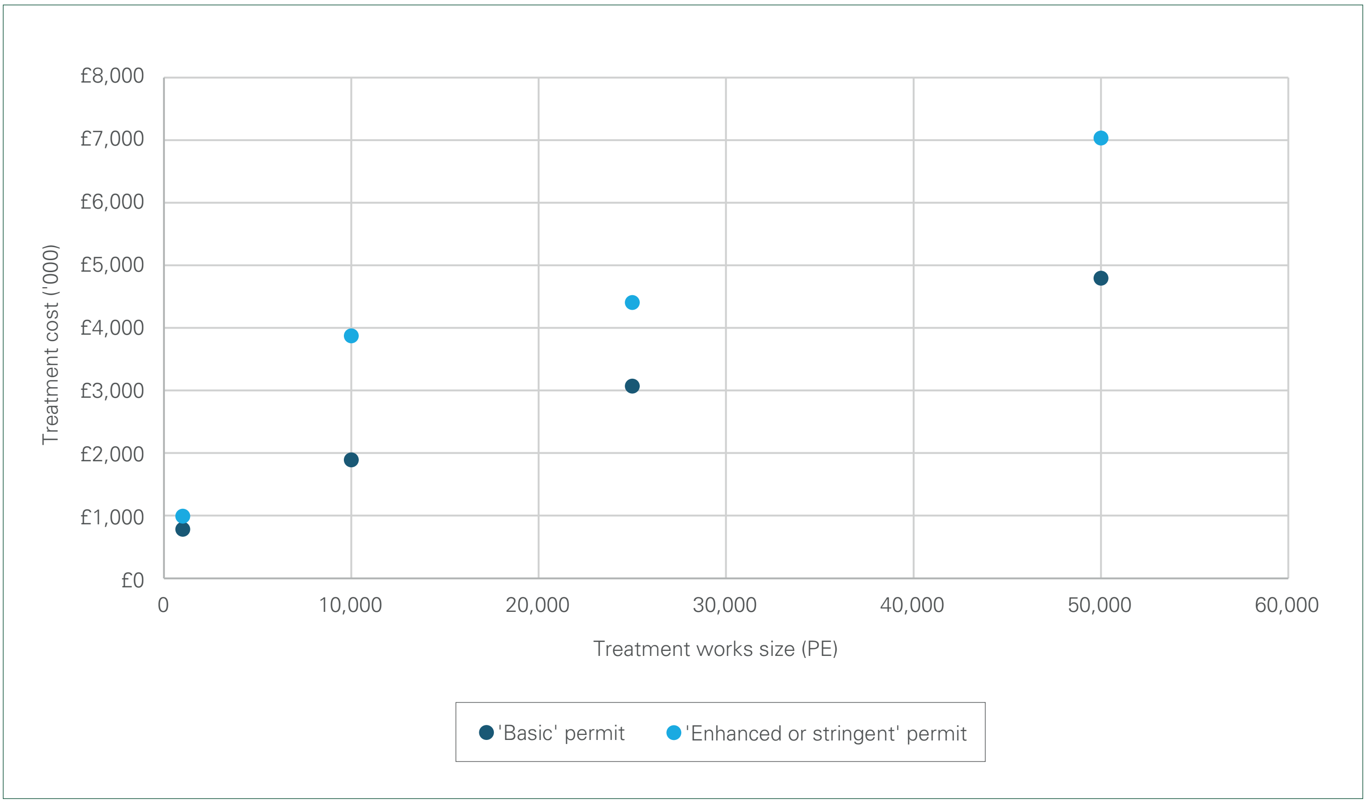


Figure 15: Indicative average treatment cost for various permit levels

TYPE	RECOMMENDATION
Extend data collection	More detailed, reliable works-level data on permits and cost should be published for use in econometric models.
Include additional variable	Permit data should be used to generate an exogenous quality driver.

Figure 16: Recommendations on quality

5.2 BOD LOAD MEASUREMENT

£30-50m

Variation in company costs if a consistent estimation method for PE is adopted

Load is a primary driver in the PR14 models and throughout the modelling exercise, so inaccuracies or inconsistencies in the data can have a pronounced effect on allowed costs. Analysis suggested that adopting a consistent estimation method for population equivalents served could affect company costs by values on the order of 30 to 50 million pounds that vary substantially by company. This analysis tests for errors and inconsistencies, though limited company reporting means results cannot be fully audited.

Measurement error varies by component of load, which in turn vary proportionally by company. Load figures are estimated from four components: resident population, non-resident population, trade effluent, and net imports, as outlined in Table 25. Resident and non-resident population loads, which together are responsible for approximately 88 per cent of total load, are estimated by multiplying resident and non-resident population within the company area by an assumed load per population equivalent (PE). Trade effluent and net imports are estimated separately, and then all four components are aggregated into a single load figure reported to Ofwat. Each component is associated with varying degrees of measurement error, depending on company estimation practices that are not reported. Table 25 highlights the errors associated with different load components when suggested estimation sources are used.

There is a modest underlying load measurement error of approximately one per cent. Errors vary by company depending on the relative composition of load, with higher errors associated with companies that serve areas with more trade effluents or non-residents.

Table 26 illustrates estimated load errors by company when the suggested sources in Table 25 are used: these range from 0.3 to 1 per cent. Note that this error, while it is not large, creates uncertainties around coefficient estimates and therefore cost thresholds of ±£8m for one company.

LOAD COMPONENT	APPROXIMATE SHARE OF TOTAL LOAD	ASSUMED LOAD PER PE	PR14 SOURCE	SUGGESTED SOURCE	ERROR
Resident population	86%	60 g BOD/PE/day	Not reported	UK Census	0.15%
Non-resident population	2%	No universal assumption	Not reported	GB Tourism Survey	2.5%
Trade effluent	12%	N/A	Not reported	Metered data	5% (Typical meter and sampling error)
Net imports	Unknown	N/A	Not reported	No universal source	Unknown

Table 25: Information sources for components of load

	RESIDENT POPULATION (,000)	NON-RESIDENT POPULATION (,000)	TRADE EFFLUENT REPORTED (KG BOD/d)	TOTAL LOAD REPORTED (kg BOD/d)	TOTAL ERROR
Anglian	5,928	254	46,616	416,512	0.76%
Northumbrian	2,669	74	15,054	175,206	0.60%
Severn Trent	8,917	51	75,748	630,972	0.76%
South West*	1,558	143	57,098	106,076	2.01%
Southern	4,502	67	7,426	279,610	0.30%
Thames	14,958	474	32,338	967,611	0.36%
United Utilities	7,362	248	83,878	525,612	0.95%
Welsh	3,083	158	19,538	248,893	0.67%
Wessex	2,711	43	14,571	194,194	0.57%
Yorkshire	5,038	46	57,670	364,724	0.94%
Average					0.79%

Table 26: Estimated load measurement error by company

Note: South West's reported trade effluents are extremely high and likely incorrect. This error causes the impacts on South West to be overstated.

Companies do not use the same assumptions about non-residential loads. Company methodologies note a range of non-resident (NR) assumptions from 20-60g BOD/PE/day. The level at which the non-resident assumption is set also matters a great deal; the final column of Table 27 shows the changes associated with a 20g BOD/PE/day increase or decrease from an assumed 40g BOD/PE/day standard. Though non-residents are a very small portion of total load, cost impacts can be as high as £31m, illustrating the importance of clear standardised guidance on how load should be estimated.

Further inconsistencies appear to reflect different demographic assumptions. It is difficult to audit company estimation practices because not all of the components that make up total load are reported separately, so this study tested for inconsistencies by constructing a measure of load using a single set of sources and assumptions. It operated in the following steps:

- it used the UK Census and GB Tourism Survey to estimate resident and non-resident populations, respectively;
- it assumed that resident population produces 60g BOD per head per day and non-resident population 40g BOD per head per day;
- it added this to reported trade effluents, assuming that net imports are not material.

Comparing these standardised estimates to those reported by company, as reported in Table 26, indicates differences between the reported and standardised estimates of between -3 and +15 per cent, suggesting substantial inconsistencies in how companies estimate load. Data on net imports, if published, could be used to verify this assessment. Cost threshold impacts from standardising how load is estimated, shown in the final two columns of Table 27, approach £100m in some cases.

	TOTAL LOAD REPORTED (kg BOD/d)	NON-RESIDENT POPULATION (,000)	TRADE EFFLUENT REPORTED (kg BOD/d)	TOTAL LOAD REPORTED (kg BOD/d)	TOTAL ERROR
Anglian	416,512	412,440	1%	-£25.0	± £31.0
Northumbrian	175,206	178,167	-2%	£9.8	± £2.0
Severn Trent	630,972	612,779	3%	-£65.9	± £3.7
South West*	106,076	156,304	-47%	£152.5	± £8.5
Southern	279,610	280,223	-	£0.6	± £1.3
Thames	967,611	948,780	2%	-£29.2	± £14.7
United Utilities	525,612	535,524	-2%	£31.2	± £15.6
Welsh	248,893	210,874	15%	-£97.2	± £8.1
Wessex	194,194	178,987	8%	-£51.9	± £2.9
Yorkshire	364,724	361,777	1%	-£8.9	± £2.8

Table 27: Sensitivity of cost thresholds to assumption inconsistencies by company

Note: South West's reported trade effluents are extremely high and likely incorrect.
This error causes the impacts on South West to be overstated.

A final source of inconsistency is the absence of commuters from non-residential calculations. While this is unlikely to be material for most companies, some serve areas in which commuters comprise a substantial proportion of the population actually served. The inclusion of commuters should be considered if an appropriate dataset can be identified.

This study recommends removing the uncertainty surrounding definitions and measurement of load. The driver is relied upon in many of the PR14 models as well as the new models developed in this report. Figure 17 summarises the recommendations from this section, many of which are extremely easy to implement.

TYPE	RECOMMENDATION
Extend data collection	Collect company reports on disaggregated components of load measurement, including net imports
Extend data collection	Standardise demographic sources used to estimate PE
Extend data collection	Establish a clear guideline about the assumptions made for non-resident populations
Extend data collection	Explore the use of commuter data to incorporate into non-resident population estimates

Figure 17: Recommendations on load measurement

5.3 REGIONAL WAGES

Wages directly affect cost of service provision. The costs of labour are a significant component of company costs: employment costs account for 16 per cent of botex across companies, to which further costs of labour associated with outsourced activities can be added. To some extent these costs will reflect labour market conditions that are outside company control and specific to company regions. Cost allowances should therefore account for these differences where they are material.

This section considers whether there is evidence of regional wage variation and how it should be accounted for in cost assessment. There is evidence of statistically significant variation between wages in London and those elsewhere in England and Wales. However, evidence on the magnitude of variation drawn from different sources is mixed and incomplete – and no publicly available source offers geographically granular information on local conditions. Analysis shows that econometric models do not account well for regional wage variation: off-model adjustments to company cost bases will thus account for variation more accurately.

It recommends the use of ex post adjustments to cost thresholds to adjust for regional wages based on evidence provided by companies. As material variation from the average appears to be confined to Thames and Southern Water's regions – but further evidence is needed from companies to establish its magnitude – this report recommends an ex post adjustment.

5.3.1 REGIONAL VARIATION IN WAGES

Analysis

Three approaches to measuring regional wage differences are presented. The first two use ONS wage statistics from surveys based, first, on wages in relevant industrial classifications and, second, on wages in representative occupational classifications, while the third uses proprietary data used by recruiters for salary benchmarking.

Table 28 summarises the metrics considered across all three measurement methods, noting their performance against three tests discussed in this section.

The occupational classification metrics considered have been developed by Ofwat and are more sophisticated than those used at PR14. The PR14 models adopted a measure of regional wages based on two different two digit Standard Occupation Classification (SOC) codes receiving 60 and 40 per cent weightings. Ofwat has subsequently developed more robust metrics using the same wage dataset and weightings based on real company shares of labour costs for different SOC groupings. The one and two digit based occupational codes since proposed by Ofwat are more tailored to wage costs faced by companies than the arbitrary metrics used at PR14, albeit with pros and cons associated with the use of SOC data discussed below, so they were considered ahead of the PR14 metrics for this analysis.

CLASSIFICATION	METRIC	SOURCE	ENDOGENEITY	RELEVANCE	ROBUSTNESS
Proprietary	Varied	Willis Towers Watson, HAYS	Low	Unclear	Unclear
Occupational	SOC – 1 digit code	ONS	Minimal	Low	High
Occupational	SOC – 2 digit code	ONS	Minimal	Low - medium	High
Industrial	SIC – Water and Waste	ONS	High	High	Medium
Industrial	SIC – Utilities	ONS	Low	High	High
Industrial	SIC – Utilities + construction	ONS	Low	Low	High

Table 28: Regional wage metrics considered

Note: Proprietary datasets, as discussed below, were not considered suitable for regulatory purposes, though the findings broadly support the conclusions reached in the rest of the analysis.

Metrics were compared on the basis of three principles. The goal of an index is to accurately represent the substitution possibilities experienced by the labour market from which wastewater companies hire, but in the regulatory context of cost assessment the use of an index should not introduce adverse incentives for companies or reward inefficiency. The three criteria used to test this are outlined below.

Relevance

Wages should be representative of substitution possibilities. This means that allowable costs for employees should closely approximate the cost of recruiting and retaining labour with suitable skills to work in a WaSC, as evidenced by wages for a representative set of similar roles. One reason indices may fail to be representative could be through the influence of irrelevant sectors: for example, high regional wages differentials observed in the finance and insurance sector should not be reflected in the allowable costs for WaSC employment, since employees seldom substitute between these two sectors and the wastewater sector. Occupational classifications are at higher risk of being skewed by irrelevant sectors than industrial classifications, since they include wages from a broader set of sectors, while proprietary data is designed to measure only wages in comparable sectors so does not suffer from these shortcomings. However, with the information available on company employee categorisations, it is difficult to use proprietary data to measure wages across a full representative set of roles within companies.

Robustness

Sample sizes must be sufficiently large to ensure that differences are reported accurately. Because ONS conducts business wage surveys on a sample of UK employees rather than the entire population, the robustness of any given regional, occupational, and industrial division depends on the number of respondents within that region, occupation, or industry. Some resulting divisions are too small to have enough respondents for a robust measure to be constructed. ONS reports the coefficients of variation (CVs) for all of their results, so these were utilised to ensure that metrics had sufficiently large sample sizes to be usable for cost assessment. Providers of proprietary data do not report the methodological information to assess robustness statistically so the statistical robustness of their results is not clear, although, less formally, the fact that clients pay for access to this data and use it to make pay decisions offers some testament to its robustness.

Exogeneity

Companies should not be able to affect their allowable costs by changing the wages they pay employees. Wage measures must be exogenously determined to ensure that companies are not incentivised to overspend or rewarded for historical inefficiency. This requirement was tested by considering the proportion of employees covered by a metric whose wages are paid by WaSCs: metrics in which WaSC employees make up a high proportion of the total are not suitable for cost assessment. Generally this is not a concern for occupational classifications, which encompass employees from a wide array of sectors, but is an issue for some industrial classifications, which can be dominated by a small number of sectors. For proprietary data, which attempts to measure directly the minimum cost of hiring new staff, wages can be endogenous to the extent that WaSCs have market power in recruiting employees.

Findings

London wages tend to be statistically higher than those of other regions. For most metrics, London exhibited higher wages than other regions, while the other regions were not statistically distinguishable from one another. Table 29, for example, illustrates this for the one digit SOC index over time. This translates into Thames Water's region having statistically higher wages than other companies in the occupational indices, while other companies are not statistically different from one another. Wages in Southern Water's region generally appeared higher than those in remaining companies' regions, though evidence of statistical significance was limited and sensitive to occupational groupings and year.

The metrics assessed show differing magnitudes of regional variation, with differing orders between companies. Table 30 reports regional wage indices. With the exception of the 'Water and Waste' metric, indices show regional wages in Thames Water's area of appointment to be higher than those in other company areas, with Southern Water's region second highest. However, both the magnitude of variation between other companies and the ranking of wages varies quite substantially: the 'Utilities plus construction' shows approximately 50 per cent more variation between the lowest and highest regional wages than the 'Water and Waste' measure; while the ranking of company regions other than Thames and Southern by wage level varies by at least three places.

	2011	2012	2013	2014	2015
East	£14.66	£14.33	£14.09	£13.95	£14.02
East Midlands	£13.97	£13.79	£13.56	£13.41	£13.41
London	£19.36	£18.44	£18.14	£17.75	£17.44
North East	£13.97	£13.64	£13.58	£13.49	£13.71
North West	£14.29	£14.00	£13.83	£13.71	£13.66
South East	£15.53	£15.22	£14.89	£14.74	£14.86
South West	£14.12	£13.77	£13.60	£13.44	£13.47
Wales	£13.58	£13.38	£13.18	£13.15	£13.14
West Midlands	£14.22	£13.89	£13.79	£13.71	£13.59
Yorkshire	£13.93	£13.75	£13.43	£13.32	£13.40

Table 29: Real company wage one digit SOC estimates over time (2013 base)

Note: Occupational classifications were aggregated based on the FTE surveys of WaSCs collected by Ofwat. Paired t-tests using ONS reported coefficients of variation failed to reject the null hypothesis of equal means except for the London region.

COMPANY	SOC1	SOC1 – F	SOC2	WATER AND WASTE	UTILITIES	UTILITIES + CONSTRUCTION	HAYS	RANKING RANGE
Anglian	1.00	1.00	1.00	0.94	0.92	1.00	1.02	5
Northumbrian	0.97	0.96	0.96	0.91	0.82	0.89	0.90	5
Severn Trent	0.97	0.98	0.96	0.95	0.98	0.98	0.99	3
South West	0.97	0.96	0.98	1.09	1.10	1.00	0.99	4
Southern	1.06	1.07	1.07	1.14	1.12	1.14	1.12	1
Thames	1.19	1.16	1.18	1.01	1.16	1.21	1.19	3
United Utilities	0.98	0.98	0.98	0.93	0.98	0.97	0.99	5
Welsh	0.95	0.95	0.94	0.98	0.89	0.86	0.85	5
Wessex	0.97	0.96	0.98	1.09	1.10	1.00	0.99	4
Yorkshire	0.96	0.96	0.95	0.97	0.94	0.95	0.96	3
Range	0.24	0.21	0.24	0.23	0.34	0.35	0.34	

Table 30: 2013 wage estimates by company for occupation and industrial indices

Notes: Index values are expressed relative to the arithmetic mean value across company regions. Regional wage data from ONS was translated to company areas using Ofwat's methodology using proportional assignment by population served. Occupational classifications (SOC1 and SOC2) were aggregated based on the FTE surveys of WaSCs collected by Ofwat. SOC1 – F is 'SOC1* without Finance' presented and defined in Table 31. 'Water and waste', 'Utilities' and 'Utilities + construction' are defined in the notes to Table 32. 'HAYS' is defined in Appendix A. The 'ranking range' column reports the difference between each company's maximum and minimum rank according to the various indices.

Relevance

SOC code based metrics are skewed by the inclusion of irrelevant sectors. ONS data tabulated by occupation and industry allowed for company wages to be constructed both with and without particular sectors.

Table 31 illustrates the impact of removing the finance and insurance sector, for example, which has a large impact upon the estimated differences in wages between companies. The SOC1* wage range in which finance and insurance is removed decreases by 20 per cent from £3.62 to £2.90, for example. This is illustrative of the degree to which occupational codes are influenced by other sectors unrelated to WaSC activities, such as Real Estate, Accommodation and Food Service, or Education. SOC2 wages are slightly less susceptible to this influence, with the range decreasing by 16 per cent, but it remains a concern.

COMPANY	SOC1	SOC1*	SOC1*, WITHOUT FINANCE	FINANCE INFLUENCE ON SOC1*	SOC2	SOC2*	SOC2*, WITHOUT FINANCE	FINANCE INFLUENCE ON SOC2*
Anglian	£14.00	£13.93	£13.84	0.63%	£13.62	£13.54	£13.50	0.30%
Northumbrian	£13.57	£13.34	£13.29	0.34%	£13.07	£13.15	£13.16	-0.06%
Severn Trent	£13.69	£13.59	£13.48	0.80%	£13.13	£13.15	£13.11	0.37%
South West	£13.60	£13.48	£13.29	1.40%	£13.30	£13.24	£13.13	0.84%
Southern	£14.87	£14.82	£14.71	0.72%	£14.55	£14.49	£14.43	0.42%
Thames	£16.73	£16.80	£16.05	4.67%	£16.11	£16.07	£15.56	3.25%
United Utilities	£13.82	£13.63	£13.52	0.78%	£13.32	£13.41	£13.37	0.28%
Welsh	£13.31	£13.17	£13.15	0.19%	£12.79	£12.81	£12.81	-0.03%
Wessex	£13.60	£13.48	£13.29	1.40%	£13.30	£13.24	£13.13	0.84%
Yorkshire	£13.43	£13.34	£13.23	0.81%	£12.91	£12.95	£12.90	0.40%

Table 31: Finance sector influence on occupational classification metrics for 2013

Note: SOC1* and SOC2* were constructed from ONS regional data divided by both occupation and industry. Multiple divisions reduces the reliability of job number estimates, so SOC1* and SOC2* were constructed to be a fairer comparison when the finance sector jobs were removed. This exercise was for testing the influence of irrelevant sectors only; the increased reliability of SOC1 and SOC2 estimates compared to SOC1* and SOC2* means that they would be used if the occupational classification option were pursued.

Relevance

None of the metrics reflects wage variation within ONS regions. This reduces the relevance of the metrics for two reasons. First, regional averaging in ONS regions can obscure variation between companies that operate in different parts of ONS regions where different labour market conditions prevail. For example, the CMA's PR14 determination for Bristol Water observes that Bristol Water faces higher local wage costs than South West Water, though both company areas lie within a single ONS region. Second, the use of regional averages does not account for companies' ability to achieve savings by relocating employment within their regions. This consideration is perhaps most relevant for Anglian and Thames Water, whose areas of appointment intersect London and the South East, as well as other regions where wages are lower.

Robustness

Metrics using ONS data had sufficiently large samples to be used in cost assessment.

Table 32 tabulates coefficients of variation by region for the industrial classification metrics. As a rule of thumb coefficient scores of more than 10 per cent are interpreted as showing a lack of robustness, as such scores suggest that the ranges of variation set out in Table 32 are contained within a 95 per cent confidence interval around the mean. The results suggest that the Water and Waste index exhibits a modest risk of small sample errors, but the Utilities and Utilities plus construction metrics are suitably accurate. Indices based on SOC groupings are weighted averages of a number of occupational classifications, so there is little risk of insufficient sample sizes.

ONS REGION	WATER AND WASTE	UTILITIES	UTILITIES + CONSTRUCTION
South East	5.7%	5.2%	4.0%
London	9.2%	8.2%	4.5%
North East	7.3%	7.1%	6.1%
North West	6.8%	5.2%	3.5%
Yorkshire and The Humber	6.9%	6.0%	4.1%
East Midlands	8.3%	5.0%	4.0%
West Midlands	11.0%	6.8%	4.8%
East	4.0%	5.2%	3.0%
South West	5.9%	6.6%	5.0%
Scotland	7.6%	5.0%	4.0%
Wales	6.9%	6.3%	4.6%

Table 32: 2013 coefficients of variation by region for industrial classifications

Note: SOC1* and SOC2* were constructed from ONS regional data divided by both occupation and industry. Multiple divisions reduces the reliability of job number estimates, so SOC1* and SOC2* were constructed to be a fairer comparison when the finance sector jobs were removed. This exercise was for testing the influence of irrelevant sectors only; the increased reliability of SOC1 and SOC2 estimates compared to SOC1* and SOC2* means that they would be used if the occupational classification option were pursued.

INDUSTRIAL METRIC	TOTAL JOBS	WOC OR WASC JOBS	WOC OR WASC INFLUENCE ON METRIC
SIC – Water and Waste	131,000	54,000	41%
SIC – Utilities	299,000	54,000	18%
SIC – Utilities + construction	1,123,000	54,000	5%

Table 33: WaSC influence on industrial classification metrics for 2013

Note: This assumes that WaSCs account for all of the employment within the water collection, treatment and supply (36) and sewerage (37) classifications. It also excludes the remediation activities and other waste management services (39) sector, for which ONS suppressed job numbers due to insufficient sample size. Both of these assumptions mean that the per cent influence figures are conservatively high.

Endogeneity

Endogeneity concerns around the Water and Waste metric can be addressed by using broader sectoral measures. As Table 33 outlines, WoC or WaSC jobs made up an estimated 41 per cent of the national water and waste sector in 2013, a proportion that is quite consistent over time and across regions. This implies that WaSCs could explain 41 per cent of the additional costs if they chose to raise wages. The utilities metric would reduce that concern by more than half, with companies only recovering 18% of additional spend. Adding construction reduces this still further, though the dominant size of the construction sector means that the resulting wage estimates are less representative of the whole gamut of roles within a utility.

5.3.2 ECONOMETRIC ASSESSMENT

Analysis

Regional wages were tested in regressions to determine their suitability for inclusion in benchmarking models. One way of accounting for regional wage variation is to include wage variables in benchmarking models. However, the evidence presented in the previous section showed that London was the only region statistically distinct from others, which means that regional wages are unlikely to be well suited for regression analysis due to a lack of variability driving estimation. Under such circumstances, results are more likely to be sensitive to specification, exhibit collinearity with a time trend, or be driven by omitted variables.

Regional wage coefficients should exhibit a positive relationship with costs if they are well suited to inclusion in benchmarking models. Coefficient values less than zero or greater than one are implausible, because increasing wages by one per cent should increase costs, but by at most one per cent. Coefficient values of around 0.6 or 0.7 would be more consistent with expectations: these served as the rough guide adopted for this analysis.

Findings

Regional wage regression coefficients lack statistical significance and in many cases are implausible values. Table 34 summarises the regression results using Ofwat's PR14 wage metric and the Utilities index motivated above. Both the PR14 regional wage metric and the Utilities index perform poorly in the majority of specifications, with little evidence of statistically significant variation and some coefficient values with implausible magnitudes.

	T&S GLS		T&S OLS		Botex GLS		Botex OLS	
	PR14+	Utilities	PR14+	Utilities	PR14+	Utilities	PR14+	Utilities
Log Load	6.89	6.42	11.72	10.57	3.85	5.19	8.52	9.11
Log Density	33.93	17.89	60.35	45.11	1.06	1.72	21.68	23.62
Log Load^2	0.10	0.19	0.16	0.16	0.10	0.15	0.15	0.13
Log Density^2	-1.00	1.45	-1.90	-0.57	1.83	2.65	1.44	1.28
Log Load x Log Density	-2.00	-2.38	-3.47	-3.16	-1.27	-1.88	-2.62	-2.68
Year	0.01	-0.03	0.00	-0.02	0.03	0.01	0.02	0.01
Log % Bands 1-3					0.20	0.15	0.13	0.10
Log Wage	1.12		0.75		0.32		0.13	
Log Utilities Wage		-0.29		0.31		-0.14		0.12
Constant	0.1	0.1	-206.8	-165.6	0.1	0.1	-103.3	-110.7
N	100	80	100	80	100	80	100	80
R^2								

Table 34: Regression results for regional wage models

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

5.3.3 CONCLUSIONS AND RECOMMENDATIONS

The evidence shows that wages for relevant employees are higher in London than elsewhere, but no metric clearly demonstrates the magnitude of variation between company regions. Taken together, the evidence in this section supports the following conclusions:

- Prevailing wages for relevant employees are clearly higher than average in London, and perhaps also in the South East, though the statistical evidence for the latter is less clear cut. Evidence on differentials between other regions is weaker: variation observed in the data is not statistically significant and depends on the metric used.
- To measure the magnitude of variation between regions, the SIC code based 'Utilities' performs better than other metrics considered. SOC-code based metrics developed by Ofwat, though more rigorous than those used at PR14, are skewed by wages in irrelevant sectors.
- A critical weakness of all of the metrics considered is the lack of more geographically granular information on labour market conditions. This makes it difficult to draw valid inferences over the extent to which companies can manage labour costs through choices over where to locate activities within their regions, for example by avoiding London or other large cities. The fact that variation in regional wages as measured by the Water and Waste is lower than that in any other index underscores this concern, though the index itself is arguably too endogenous for use in cost assessment.
- Even if exogenous wage variation between company regions could be adequately measured, econometric models are not well suited to capturing factors that only affect one or two companies. Tests show that regional wage variables perform poorly in models similar to those used at PR14.

The above supports a general recommendation for an 'off-model' adjustment to account for regional wage variation. Regional wage variation cannot be accounted for in econometric models, while the risk associated with off-model adjustments – namely that regional wages are correlated with other drivers in the models – appears to be modest for wastewater.

Off model adjustment can take ex post or ex ante forms. Off-model adjustments can either be 'ex ante', where company costs are normalised for the effects of wage variation prior to model estimation, or 'ex post', in which case cost thresholds are adjusted in response to evidence of material variation in regional wages provided by companies. Both approaches have been followed by regulators in cost assessment: Ofgem used an ex ante adjustment to account for regional wage variation at RIIO-GD1, while ex post adjustments, also known as 'special factors' or 'cost exclusions', are widespread.

The nature of the evidence needed lends itself more readily to an ex post adjustment. An ex post adjustment, in which the burden of proof for cost variation falls on companies, can more effectively account for regional wage variation. This reflects both the lack of clear evidence on how variations in wages between ONS regions affect company costs, and the nature of the evidence needed to address this, which is both detailed and company-specific in nature. Companies that apply for adjustments should present evidence of: the relevant labour cost base including contractor costs; in- and outside-region substitution opportunities and relocation costs; and local variation in wage costs. Under an ex ante adjustment, all companies would need to present such information, both historically and projected for next price control – an onerous undertaking given the lack of clear evidence of regional variation.

Figure 18 summarises the recommendations from this section.

TYPE	RECOMMENDATION
Change modelling approach	Do not include regional wage variables in benchmarking models
Change modelling approach	Adopt an ex-post adjustment, expected to focus on Thames and Southern Water.
Collect more data	Companies that apply for adjustments should be encouraged to provide evidence of labour cost base, substitution opportunities, and wage differentials.

Figure 18: Recommendations based on analysis of regional wages

5.4 SLUDGE (BIO-RESOURCES)

A new price control is being developed for sludge (bio-resources) in PR19. Ofwat and the wastewater companies continue to discuss the relevant data needs for this price control. In the interim, the project evaluated the key external factors that differentiate the cost of sludge management, and identified that the most fundamental driver of cost is the availability of land suitable for residual sludge disposal. The study identified that there are three principle routes for residual sludge disposal:

- recycling through spreading on agricultural land;
- disposal via incineration; and
- disposal to landfill.

There are other influencing factors for sludge that are within company management control, including treatment of sludge prior to final residual disposal. Each company has developed its own sludge management strategies over a number of years and AMPs. However from the evaluation, it is still the case that the majority of residual sludge in England and Wales is sent to agricultural land. Disposal to agricultural land is dependent on the net availability of suitable land ('sludge land bank') available local to the sludge production; sludge land bank is an exogenous factor not within the control of companies.

The study subsequently carried out preliminary analysis on available sludge land bank as a possible comparative dataset. Based on high level data, it was observed that companies with access to adequate land bank tend to primarily utilise land disposal as a sludge disposal route, and typically have lower company boundary sludge export rates. However, this may not be the most cost-effective way of managing sludge. Companies in areas with stressed land bank supplement disposal to land bank with alternative disposal routes. The alternative disposal routes are generally more expensive, due to the additional sludge treatment required and transportation to treatment and ultimate disposal sites.

The study found that currently publicly-available datasets do not allow a suitable detailed analysis of net land bank availability. Future work may need to rely on proprietary datasets, for example the ALOWANCE tool developed by the environmental and agriculture research consultancy RSK-ADAS Ltd which covers England and Wales. Future work must also consider development of strategies that recognise sludge as a bio-resource, rather than simply a product for disposal.

SECTION 06

New models

Taken together, the recommendations set out in this report offer substantial improvements to the PR14 models.

Recommendations from the sections throughout this report can be productively combined and implemented using new models in which:

- there is clear engineering evidence supporting the drivers;
- the stability and sensitivity to specification of primary drivers is reduced;
- the primary drivers have significant coefficients and signs are consistent with expectations based on engineering evidence;
- concerns over model specification remain, but the risk of overfitting or omitted variable bias is reduced.

Analysis

New models combine a range of recommended changes to test their collective performance and interactions. OLS regressions utilising the Cobb-Douglas form and time fixed effects were tested for each of the network, treatment and sludge, and botex models, as outlined in Table 35. Various specifications test alternative variables, but load and length remain as primary drivers, supplemented by measures of economies of scale, drainage, and urbanisation. Coefficient results, presented in the tables that follow, were checked for conformity to engineering accounts of costs, statistical significance, and impact on overall model performance.

Network models

A 'basic' network model using length and drainage has credible coefficients and dramatically improved stability, though it lacks an urbanisation driver. Coefficients from regression models are displayed in Table 36. Signs on both length and annual runoff are positive, consistent with engineering accounts, and are robust to the addition of variables in other regression tests. Statistical performance on tests of multicollinearity and misspecification, presented in Table 37, are also markedly improved.

The measure of drainage may be improved when a longer time series of combined sewer lengths is available. As discussed in Section 4.2 on drainage, extended data series on combined sewer lengths may improve the drainage variable. Once available, it should be tested in similar regression models and its performance compared to annual runoff.

The percentage of urban area appears to be a viable proxy for network length in urban areas. Adding percentage of urban area does not change the coefficients on length and drainage, indicative of stability. The percentage of urban area, though it is not significant, does have the expected sign. There is a clear engineering narrative for how networks in urban areas affect cost, for which percentage of urban area is a proxy. However, as noted above this proxy is sensitive to rural areas in which companies may have few network assets. Additional work is required to create and test variables that more closely measure this relationship.

MODEL	ESTIMATION METHOD	RESPONSE VARIABLE	EXPLANATORY FACTORS (LOGGED)	EXPLANATORY FACTORS (UNLOGGED)
New network	OLS	Network costs	Length, annual runoff in urban areas	% urban area, year fixed effects
New treatment and sludge	OLS	Treatment and sludge costs	Load	% load in bands 1-3, % urban area, % urban assets, %sparse assets, % tertiary, year fixed effects
New botex	OLS	Botex costs (T&S+Network)	Load	% load in bands 1-3, % urban area, % urban assets, %sparse assets, % tertiary, year fixed effects

Table 35: New model regression specifications

	NETWORK	
	BASIC	% URBAN
Log length	0.55	0.54
Log annual runoff	0.25	0.21
% urban area		0.56
2007/08	0.04	0.04
2008/09	0.09	0.09
2009/10	0.10	0.10
2010/11	0.15	0.15
2011/12	0.18	0.17
2012/13	0.12	0.13
2013/14	0.21	0.21
2014/15	0.21	0.21
2015/16	0.20	0.21
Constant	-2.95	-2.70
N	100	100
N of companies	10	10
N of years	10	10
R^2	0.83	0.83

Table 36: New network model regression coefficients

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	NETWORK	
	BASIC	% URBAN
R^2	0.83	0.83
Variance inflation factor – maximum	13.11	14.02
Variance inflation factor - median	12.46	11.91
Ramsey RESET	Pass	Pass

Table 37: Test results for new network regressions

Note: See Appendix B for more information on the tests.

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

	TREATMENT AND SLUDGE			
	BASIC	% URBAN	% SPARSE	% TERTIARY
Log load	0.92	0.94	0.88	0.903
% bands 1-3	4.25	5.95	5.78	5.45
% urban assets		2.07	4.46	3.65
% sparse assets			-0.73	-0.59
% tertiary				0.13
2007/08	0.05	0.05	0.05	0.045
2008/09	0.10	0.10	0.10	0.095
2009/10	0.16	0.16	0.16	0.151
2010/11	0.11	0.11	0.11	0.101
2011/12	0.04	0.04	0.04	0.028
2012/13	0.02	0.01	0.01	0.003
2013/14	0.01	0.01	0.01	-0.003
2014/15	0.00	0.00	-0.01	-0.016
2015/16	-0.01	-0.01	-0.01	-0.024
Constant	-6.91	-7.22	-6.54	-6.832
N	100	100	100	100
N of companies	10	10	10	10
N of years	10	10	10	10
R^2	0.93	0.94	0.94	0.94

Table 38: Treatment and sludge new model coefficients

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1% (dark blue: Log load all columns; 2008/09, 2009/10, 2010/11 all columns; Constant all columns)
- Significant at 5% (mid blue: % bands 1-3 BASIC; % urban assets % SPARSE; 2007/08 all columns)
- Significant at 10% (light blue: % bands 1-3 % URBAN, % SPARSE, % TERTIARY; % urban assets % TERTIARY)

Treatment and Sludge models

Basic treatment and sludge models that include load and economies of scale in a Cobb-Douglas form, but no other engineering drivers, also improve performance. Load and the proportion of load treated in bands 1-3 shown in Table 38 are consistently significant and are positive, both of which are consistent with engineering expectations. As summarised in Table 38 multicollinearity issues are much improved, but RESET tests continue to indicate an incomplete functional form, which underscores the importance of testing new engineering drivers.

Fuller specifications including urbanisation, sparsity, and treatment quality drivers better reflect engineering drivers but have some statistical drawbacks. The percentage of assets in urban areas is positive and insignificant when first added in the per cent urban regression in Table 38. This positive coefficient value is consistent with assets in urban areas incurring higher fixed costs than those outside urban areas. However, the coefficient is also sensitive to the inclusion of other variables, including sparsity and tertiary treatment, reflecting difficulties in disentangling urbanisation from other drivers and company fixed effects, as noted in Section 4.3.

Future work should consider alternative classifications of urbanisation and sparsity, collect a dataset that reflects variation in these drivers over time, attempt to find an exogenous measure of quality, and ensure that theoretical engineering drivers are accurately measured and represented in the data.

	NETWORK			
	BASIC	% URBAN	% SPARSE	% TERTIARY
R^2	0.93	0.94	0.94	0.94
Variance inflation factor – maximum	2.35	3.12	7.4	10.17
Variance inflation factor - median	2.34	2.37	4.78	4.06
Ramsey RESET	Fail	Fail	Fail	Fail

Table 39: Test results for treatment and sludge new regressions

Note: See Appendix B for more information on the tests.

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

Botex models

Botex model results are similar to those of treatment and sludge models, with attenuated coefficients. Unsurprisingly, the addition of network costs tends to dilute the explanatory power of variables designed to explain treatment and sludge costs. Models that also included network drivers were tested, but there was insufficient explanatory power to include all of the drivers into the botex model.

Table 40 and Table 41 show the model coefficients and statistical tests, respectively.

	BOTEX			
	BASIC	% URBAN	% SPARSE	% TERTIARY
Log load	0.89	0.89	0.87	0.88
% bands 1-3	3.71	4.85	4.76	4.45
% urban assets		1.38	2.65	1.89
% sparse assets			-0.39	-0.25
% tertiary				0.13
2007/08	0.05	0.05	0.05	0.05
2008/09	0.11	0.11	0.11	0.10
2009/10	0.15	0.15	0.15	0.14
2010/11	0.11	0.11	0.11	0.10
2011/12	0.16	0.16	0.16	0.15
2012/13	0.18	0.18	0.18	0.17
2013/14	0.21	0.21	0.21	0.20
2014/15	0.20	0.20	0.20	0.19
2015/16	0.20	0.20	0.20	0.19
Constant	-6.00	-6.20	-5.84	-6.11
N	100	100	100	100
N of companies	10	10	10	10
N of years	10	10	10	10
R^2	0.93	0.93	0.93	0.93

Table 40: Botex new model coefficients

Note: Further interpretation guidelines can be found in Appendix B

Key:
- Significant at 1%
- Significant at 5%
- Significant at 10%

	BOTEX			
	BASIC	% URBAN	% SPARSE	% TERTIARY
R^2	0.93	0.93	0.93	0.93
Variance inflation factor – maximum	2.35	3.12	7.4	10.17
Variance inflation factor - median	2.34	2.37	4.78	4.06
Ramsey RESET	Fail	Fail	Fail	Fail

Table 41: Test results for botex new models

Note: See Appendix B for more information on the tests.

Key:
- Fails at p<0.001
- Fails at p<0.01
- Fails at p<0.05

Recommendations

Additional work can develop specifications that reflect engineering narratives and perform well against statistical tests. These models can extend the new models presented above, including alternative urbanisation and treatment quality drivers.

A diverse suite of models should be used at PR19, with stability across models tested through analysis of company efficiency rankings. Before being recommended for regulatory use, a diverse suite of recommended models should be tested for consistency in the efficiency scores and company rankings that they produce. Model diversity reduces regulatory risk, and stability in company rankings across such a diversity would add evidence of stability to the overall modelling approach.

Figure 19 summarises the recommendations based on the evidence presented in this section.

TYPE	RECOMMENDATION
Additional modelling investigation	Collect and test an exogenous measure of quality in new regressions
Additional modelling investigation	Collect and test alternative measures of sparsity and urbanisation that vary over time in synthesis regressions

Figure 19: Recommendations for further model development

Appendix A: Engineering background

This appendix details the engineering background to the investigations which identify and explore the underlying causes of cost in the delivery of wholesale wastewater services.

The appendix is structured as follows:

Appendix A1

Appendix A1 summarises the list of fourteen factors investigated by this phase of the study.

Appendix A2

Appendix A2 presents the supporting evidence used in analysing the three factors taken through to full econometric modelling.

Appendix A3

Appendix A3 summarises the evidence base for those factors which preliminary investigations showed to be less significant or difficult to evidence and so were dropped early in the study.

Appendix A4

Appendix A4 presents the datasets used in the study.

A1 LIST OF FACTORS INVESTIGATED

Table 42 lists the fourteen factors investigated by the project, together with a short summary of the final status.

NO	FACTOR	DRIVER	MATURITY OF NARRATIVE	DATA QUALITY AT NATIONAL LEVEL	FINAL STATUS
1	Annual amount, intensity or frequency of rainfall (runoff)	Quantity	High	Fair to good	Detailed investigation and econometric modelling
2	Influence of topography on peak flows in networks	Quantity	High	Fair to good	Preliminary investigation; see Appendix A3
3	Load received or removed	Quality	High	Variable – poor to good	Detailed investigation and econometric simulations
4	Industrial loads received or removed	Quality	High	Poor to fair	Preliminary investigation; see Appendix A3
5	Environmental designations and planning conditions	Other	High	Fair	Preliminary investigation; see Appendix A3
6	Regional wages	Econometrics	High	Fair	Detailed investigation; outside model adjustment
7	Sparsity and urbanisation	Density	Medium	Fair	Detailed investigation and econometric modelling
8	Sludge land bank	Sludge	Medium	Fair	Preliminary investigation only
9	Flood risk	Quantity	Medium	Fair	Preliminary investigation; see Appendix A3
10	Hidden or culverted rivers contributing to sewer flows	Quantity	Low to medium	Poor to fair	Preliminary investigation; see Appendix A3
11	Asset accessibility	Density	Low	Poor	Preliminary investigation; see Appendix A3
12	Customer characteristics	Quality	Low	Poor	Preliminary investigation; see Appendix A3
13	Groundwater infiltration or minewaters contributing to sewer flows	Quantity	Low to medium	Poor	Preliminary investigation; see Appendix A3
14	Asset age	Quantity	Medium	Fair	Preliminary investigation; see Appendix A3

Table 42: Factors investigated by the project (detailed investigations are highlighted in blue)

A2 SUPPORTING EVIDENCE

Appendix A2 presents the supporting evidence collated and developed as part of the engineering and environmental data analysis. The appendix should be read in conjunction with the findings in the main report.

A2.1 ANNUAL AVERAGE AND WINTER RAINFALL ANALYSIS

Annual/Winter Volumetric Run-off

For the annual average analysis, rainfall data was collated on a 1km grid, and was combined with data on urban areas and winter rainfall acceptance potential (WRAP) to develop a metric for annual average urban runoff. The resulting map (Figure 3 in Section 2) shows significant variation in annual average run-off in urban areas and across England and Wales.

The secondary data produced through the analysis was tabulated (Section 4.2.1) to enable a ranking of total estimated run-off volume by company. **The final measure of total annual urban run-off volume (m^3) was then determined for use in the econometric analysis.**

For winter rainfall, a similar approach to the annual average methodology was followed. In this case, average winter rainfall was mapped using the five HadUKP-defined rainfall zones across England and Wales (Figure 20) for the three wettest winter months: November, December and January. **The annual catchment winter run-off volume was calculated for use in the econometric analysis.**

A secondary metric 'urban Annual Effective Run-off volume in m^3/km^2 was created by totalling the overall volume in each urban area within company boundaries, and dividing by the total urban area served, for both annual and winter rainfall. However, this was considered to be an inferior metric at explaining drainage costs to total annual urban rainfall.

Datasets used

CEH/HR Wallingford Standard Annual Average Rainfall 1941-1970

UK Met Office/ Hadley Centre HadUKP UK regional precipitation series

UK Met Office/ Hadley Centre 3-month winter rainfall

UK soil permeability data

HR Wallingford Winter Rainfall Acceptance Potential (WRAP)

CEH 5-yr 25-day rainfall; CEH 2-hr maximum rainfall; CEH 24-hr maximum rainfall

ONS urban areas 2001

ONS urban areas 2011

Figure 20 HadUKP map showing regions for which England & Wales Precipitation (EWP) time series data is available

Source: Reproduced from http://www.metoffice.gov.uk/hadobs/hadukp

Annual/Winter Volumetric Run-off Time Series

Two time series datasets were constructed for use in the econometric models:

- Annual urban run-off volume by company for 2001-2015;
- Winter urban run-off volume by company for 2001-2015.

The long-term mean annual rainfall was converted to urban run-off using the 2001 ONS urban area dataset and the WRAP adjustment. A similar dataset was compiled for 2011, using the long-term annual rainfall and 2011 ONS urban area dataset plus WRAP adjustment. By interpolation, a time series dataset was created covering the period between 2001 and 2011. This was further extended to 2015, using assumptions based on the ONS population growth to estimate urban area growth from 2011 to 2015.

Two additional datasets were created using the HadUKP regional long-term annual and winter rainfall dataset, for five climatic regions across the UK, and pro-rated to provide a corresponding long-term rainfall dataset by WaSC area. This was then combined with the ONS urban area dataset and WRAP to generate the following two datasets:

- Annual urban run-off volume constructed from long-term rainfall data, by company for 2006/07 – 2015/16;
- Winter urban run-off volume constructed from long-term rainfall data, by company for 2006/07 – 2015/16.

A2.2 BOD LOAD MEASUREMENT

The BOD load treated in each treatment works size band is a key metric reported to Ofwat by WaSCs. As a primary driver used in PR14 and throughout the econometric modelling exercise, the accuracy of company-reported load will have a substantive influence on modelled cost.

Current approach to load measurement and its shortcomings

Ofwat has defined an approach to load measurement which the companies use when reporting to the regulator. Ofwat's guidance to the companies is that the BOD load should consist of load from the resident and non-resident populations. In addition, Ofwat's guidance specifically **excludes** commuters and day visitors from non-resident population figures. However, the definition does not **explicitly** rule out loads associated with trade effluents or imports.

The study found that there is a lack of consistency within the industry in regard to the interpretation of Ofwat's definitions and guidance on load measurement. The study analysed the load and PE data reported by companies in the Ofwat 2016 data share and their commentaries (where relevant) and identified differences in the calculation of loads.

Methodologies and assumptions used to calculate the different load components should be more transparent to enable direct comparison between companies.

Further findings from analysis of the 2016 industry datashare are as follows:

- average trade effluent concentrations vary widely, from 1,240 to 3,500 mg/l across the industry (excludes an outlier in the dataset);
- average sewage concentration across the industry also varies, but less so, with the range being 410 to 660 mg/l (excludes an outlier in the dataset); and
- no alphanumeric Confidence Grades are provided by the companies for BOD load figures in the Ofwat datashare. It would be useful to include this with the data.

	OFWAT	COMPANY A	COMPANY B	COMPANY C
Resident population	✓	✓	✓	✓
Non-resident population	✓	✓(Implied)	✓	✓
Trade effluent	? (Not included in definition, but included on 'Sewage Treatment' tab)	✓	? (Silent)	✓
Imports	?(Silent)	-	-	✓
Other	-	-	Load discharged to other companies' works included	-
Source	Rows 49 to 54 of Ofwat's definitions spreadsheet: "*Ofwat Wastewater Variables for Cost Assessment.xls*," August 2016	Company A June return 2016 Wastewater Explanatory Factors (WEF) explanatory note	Company B commentary on December 2016 update of industry datashare	Company C commentary on December 2016 update of industry datashare

Table 43: Load estimation components - selected companies and Ofwat guidelines

A2.3 DRAINAGE COST MODELLING

Capital cost information for drainage (surface water sewers, combined sewers) were obtained from a recent standard industry database and used to represent and analyse capital costs for UK waste water networks.

A drainage cost model was constructed and used to simulate increasing flows, by changing set parameters in turn. This was repeated for three different size catchments. Overall base costs were shown to vary by 13 to 17 per cent when dry weather flow (DWF) volumes pumped were varied by up to 25 per cent. When storage volume alone was analysed, a similar (20 per cent) increase in storage requirement on existing networks resulted in 23 to 26 per cent increase in base capital cost, depending on the typical size of the existing network in each case (shown below).

Capex increase	Base storage	20% storage increase	30% storage increase
Small catchment 1,000 PE	100%	123%	149%
Medium catchment 25,000 PE	100%	125%	151%
Large catchment 50,000 PE	100%	126%	154%

Figure 21: Relationship between variation in network storage available and base capex cost

A3 EVIDENCE ON AREAS FOUND TO BE LESS SIGNIFICANT

Appendix A3 presents factors identified in the study which were not taken beyond preliminary investigations for several reasons, including: limited impact on regulated cost; low priority compared to other factors with a similar impact; lack of nationally comparable data and difficulty in evidencing without targeted industry data collection.

The list of factors is presented in Table 44; further detail on each is provided in the subsequent sections of the appendix for completeness.

EXOGENOUS FACTOR	VARIABLE INVESTIGATED	RATIONALE FOR NOT TAKING BEYOND PRELIMINARY INVESTIGATION
Rainfall intensity	Storm intensity	Urban areas situated in areas of high annual rainfall, display rainfall intensity increases in longer storm events typically associated with winter frontal rainfall, whist shorter storm events display no discernable intensity difference between urban zones across the country. Winter rainfall assessed instead.
Topography	Impact on run-off	'Flashiness' aspect of run-off typically associated with steep topography is difficult to measure, assess and evidence nationally. Further work may provide opportunities to assess.
Industrial loads	Industrial loads received	No material effect on costs not already measured by overall load.
Environmental designations and planning conditions	Percentage of assets subject to National Park Authorities, etc.	Significance of effect on costs variable and difficult to evidence consistently across the industry without detailed data from all companies.
Flood risk	Fluvial and sea flood risk	Difficult to evidence in a consistent manner across the industry without further work, as datasets do not show the impact of flood defences, for example.
Hidden river flows into sewers	Network infiltration	Varying levels of importance across the industry. Contributions range from local to more significant. Impact of storm run-off considered more significant. Lack of publicly-available dataset. Opportunity to collate national dataset.
Groundwater infiltration	Network infiltration	Varying levels of importance across the industry. Contributions range from local to more significant. Impact of storm run-off considered more significant. Lack of publicly-available dataset. Opportunity to collate national dataset.
Customer socio-economic characteristics	Customer deprivation indices	No clear, consistent evidence of link with wholesale wastewater activities.
Asset age	Asset age	Difficult to establish a clear link with cost variance due to complexities relating to choices made by companies regarding asset upgrade programmes.
Asset accessibility	Travel times	Very dependent on local circumstances including where manpower is located. Remoteness of sites and populations better addressed under sparsity and urbanisation.

Table 44: Factors for which preliminary investigations only were undertaken

A3.1 RAINFALL INTENSITY

The intensity of rainfall received in company supply areas affects peak run-off volumes and design of networks. The study used UK Met Office rainfall statistics (mm/hr) and HR Wallingford design storm rainfall statistics (also in mm/hr) to assess rainfall intensity variation.

The results showed that extreme rainfall in UK urban regions is often associated with summer convectional thunderstorms and that there is less of a relationship between topography and rainfall intensity. This would suggest it is equally possible to receive high intensity events across the country when climatic conditions allow. The highest hourly intensities recorded in the UK typically occur in the summer. When comparing high rainfall and low rainfall areas, the data show that any increases of note in rainfall intensity would tend to occur in long duration rainfall events i.e. winter rainfall (Table 45). As a result, winter rainfall accumulations were assessed instead of intensity.

Datasets used

UK Met Office rainfall statistics (mm/hr)

HR Wallingford design storm rainfall statistics (mm/hr)

RETURN PERIOD				1 IN 1 YEAR			
STORM DURATION	**60**	**120**	**240**	**480**	**720**	**960**	**1440**
MANCHESTER (STANDARD ANNUAL AVERAGE RAINFALL = 881 MM)							
Ave. Intensity (mm/hr)	11.38	7.19	4.51	2.80	2.11	1.72	1.29
Depth (mm)	11.38	14.38	18.03	22.40	25.28	27.57	31.15
CAMBRIDGE (STANDARD ANNUAL AVERAGE RAINFALL = 551 MM)							
Ave. Intensity (mm/hr)	12.80	7.76	4.65	2.76	2.02	1.62	1.19
Depth (mm)	12.80	15.52	18.61	22.08	24.29	25.98	28.56
COMPARING MANCHESTER AND CAMBRIDGE							
Depth difference (mm/hr)	-1.42	-1.14	-0.58	0.32	0.99	1.58	2.59

Table 45: Rainfall intensities - North West region compared to South East region

Datasets used

Centre for Ecology and Hydrology, CEH/HR Wallingford Standard Annual Average Rainfall 1941-1970

UK soil permeability data

HR Wallingford Winter Rainfall Acceptance Potential (WRAP)

ONS urban areas 2001, 2011

A3.2 TOPOGRAPHY

As expected, the annual average catchment run-off pattern follows the east-west pattern of annual rainfall. The exception is in North East urban areas, such as Newcastle and Sunderland where the influence of topography results in catchment run-off volumes comparable to those in the North West and The Midlands. In general, runoff versus topography analysis shows that urban areas which receive annual catchment run-off of 500 mm (or greater) are located close to areas of high elevation, that is the North West, Wales, Yorkshire and Humber and the North East.

Preliminary GIS mapping also showed that variation in topography across England and Wales has the potential to impact volumes and speed of flows into urban wastewater systems, which would require an operational response. However, it is not straightforward to assess this impact without further information. A suitable measure to describe the 'flashiness' aspect of run-off associated with steep topography could possibly be identified, but more thought needs to go into this.

Datasets used

Environmental Permit Regulations – Industrial (limited, publicly available dataset)

A3.3 INDUSTRIAL POLLUTION LOAD RECEIVED

GIS mapping of industrial licences was carried out using a limited, publicly-available dataset from the Environment Agency. The data identifies all rural and urban industries in England that require a permit to operate. The data highlighted potential regional variations in trade discharges or effluent discharge volumes, mainly linked to the variation in numbers and types of industries.

This is better represented through trade effluent data reported by the companies to Ofwat and currently accounted for in the load received calculation. Furthermore, from the preliminary data snapshot, it was evident that the permitted industry data did not identify whether the industry discharged to sewer or not, therefore further investigations of this dataset were not pursued.

A3.4 ENVIRONMENTAL DESIGNATIONS

There is evidence of variation in the number of company assets located within designated areas. There is also evidence of variation in the relative proportion of company service area designated. The relative treatment capacity by company located in a designated area and the number of company effluent permits located in designated areas was assessed. Both of these measures showed significant differences between companies, as shown in the table below.

Assets in designated areas are commonly subject to additional costs to meet environmental and planning requirements. However, national data on this are not publicly available. PR14 evidence obtained from one company with a significant proportion of assets in designated areas suggested that the penalty was potentially about 2 per cent of company capital on-costs, comparable with operation and maintenance on-costs. It was unclear, however, whether this accounted for planning costs. For the purposes of context, the most significant on-costs were design and management (~70 per cent) and capital overheads (~20 per cent).

WATER AND SEWERAGE COMPANY	EFFLUENT PERMITS IN DESIGNATED ALL AREAS (NO.)	PE FLOW* WITHIN DESIGNATED AREAS ('000 M^3/D)
Anglian	79	253
Northumbrian	43	41
Severn Trent	103	225
Southern	170	462
South West	210	331
Thames	116	520
United Utilities	146	1,217
Welsh	22	370
Wessex	95	175
Yorkshire	118	178

Table 46: Treatment capacity and effluent permits located in designated areas

Note: Flow based on assumption of 130 l/h/d.

Datasets used

Environmental and ecological designations including AONB, National Parks, SPA, SAC and SSSIs

Heritage designations including Scheduled Ancient Monument, Parks and Gardens, World Heritage Sites

Datasets used

Environment Agency National Flood Zone 2 dataset v201608

Environment Agency National Flood Zone 3 dataset v201608

EU Urban Wastewater Treatment Directive Database

A3.5 FLUVIAL AND SEA FLOOD RISK

When they occur, floods can interrupt the operation of wastewater assets, including treatment works sites. This results in operational costs to recover the sites and continue to provide services, as well as capital costs to repair and/or replace damaged assets. The study used the datasets shown to assess flood risk to wastewater assets.

The Environment Agency flood risk data represents the following probabilities:

- Zone 2: Medium probability (between a 1 in 100 and 1 in 1000 annual probability of river flooding, and between a 1 in 200 and 1 in 1,000 annual probability of sea flooding).
- Zone 3: High probability (1 in 100 or greater annual probability of flooding or a 1 in 200 or greater annual probability of flooding from the sea).

Company wastewater assets were mapped on to the Zone 2 and 3 flood maps, by number of assets and cumulative load treated by assets located within the flood plain. There is significant variation both in asset numbers and cumulative treatment load. However, the Environment Agency maps only identify natural flood plain, and do not account for the existing standard of protection from man-made flood defences. It is not possible to obtain an accurate picture of assets and treatment capacity at risk from river and sea flooding in Zones 2 and 3 without further significant work to identify all flood defences and the extent of coverage. This would also necessitate accounting for on-site flood resilience measures, where they exist.

Industry-wide, it is not uncommon for flood risk recovery costs to be covered by insurance payments, but further evidence would be needed to verify how widespread this approach is. Given the effort required to compile detailed, consistent evidence across the industry, flood risk was not explored further in this study.


2
6
.90
.80
.70
.60
.50
.40
30

Datasets Used

Published Academic PhD dataset for Sheffield City

A3.6 HIDDEN RIVERS

In addition to groundwater infiltration (A3.7), unidentified historical assets such as culverted watercourses can be "captured" into combined sewer systems, affecting sewer base flow and sewage concentration. This increases the risk of surcharging and spills, and increases pumping and treatment costs[5]. A published PhD study dataset of 'hidden' rivers compiled for the city of Sheffield indicates that just over 20 per cent of river length in the town is culverted, but 50 per cent of the original stream length is lost (unknown). Although not all of these would be connected to sewers, this highlights the potential scale of the problem. The proportion of total river length culverted varies by region, but national data enabling this spatial analysis is no longer freely available from the Environment Agency.

Although the proportion of culverted watercourses is significantly greater in urban areas, it is not just an urban issue, and probably affects other parts of catchments. A suitable publicly-available dataset is currently lacking, and would need to be compiled from sources such as the Ordnance Survey and National Library of Scotland archives (http://maps.nls.uk/index.html), using published methods[6]. Removing captured watercourses from sewers needs a different approach to tackling infiltration-inflow. It can be achieved through stream daylighting, and so deliver wider environmental and social benefits. There are good examples of this in North America (Pittsburgh) and Europe (Zurich).

[5] Broadhead, A.T., R. Horn and D.N. Lerner (2013) 'Captured streams and springs in combined sewers: a review of the evidence, consequences and opportunities', Water Research, 47, 4752-4766.

[6] Broadhead, A.T., R. Horn and D.N. Lerner (2015) 'Finding lost streams and springs captured in combined sewers: a multiple lines of evidence approach', Water and Environment Journal, 29, 288-297.

A3.7 GROUNDWATER INFILTRATION

As with hidden rivers and culverts, groundwater infiltration into sewers is a source of sewer base flow, affecting sewer capacities and influencing sewage treatment through dilution and introduction of minerals. A groundwater emergence map for England identifies areas of shallow groundwater at a broad scale. Areas of groundwater emergence are evident in all company supply areas - to varying degrees - apart from the Northumbrian Water area. This is linked to regional geology. The north eastern, eastern and southern areas of England are most affected.

Wastewater assets in the bedrock will typically interact with groundwater aquifers, although this is unlikely to be a common occurrence due to typical sewer depths. Assets in superficial deposits - the majority of wastewater assets – are more likely to be significantly impacted by shallow groundwater infiltration and groundwater quality. Some evidence on the rebound of groundwater associated with changes in industrial abstraction or mining was collated through a literature review and identified regional variation in the impact of anthropogenic forced changes to groundwater level. However, the dataset was insufficient for detailed analysis. Groundwater infiltration issues are considered to vary in scale, but most are expected to be highly localised. For this reason and due to the lack of a comprehensive publicly-available dataset, this issue was not investigated further.

Datasets Used

Environment Agency groundwater emergence map (England only)

Environment Agency groundwater rebound data (limited)

A3.8 CUSTOMER CHARACTERISTICS

Variation in customer characteristics and behaviour - affect delivery and management of services within the supply areas. Particular characteristics of a service area (for example, deprived or affluent areas) may result in a higher cost to serve for a variety of reasons, such as blockages due to sewer use habits or maintenance access issues. From sample data reviewed, the engineering evidence of the impact of blockages on company costs is consistent, but what is less clear is the causality. Due to the absence of a clear, consistent narrative, this factor was not pursed further.

Datasets used

Ofwat Wastewater Datashare, 2016 – Network

A3.9 ASSET AGE

Asset age and condition influence the operational efficiency of sewer networks and wastewater treatment works and drive asset replacement programmes. The 2016 industry datashare includes a breakdown of sewer age profile as follows: sewers constructed before 1880; sewers constructed between 1881 and 1900 (and in each subsequent twenty-year period until 2000) and sewers constructed after 2001. The data (Table 47) show that about 12 per cent of sewers in use today were constructed in 1900 or earlier. Sewer construction rates were highest between 1921 and 2000, with the proportions constructed in each 20-year period ranging from about 17 per cent to about 22 per cent. Overall, this period accounts for 75 per cent of sewers by construction age. Only 6 per cent of sewers have been constructed since 2001.

Comparing individual companies reveals that the sum of sewers over 50 years old varies from 38 per cent (Wessex Water) to 65 per cent (Thames Water), highlighting the exogenous nature as this is related to historical development of each area. The industry data are presented only for 2015/16 reporting year. Despite the evidence of variation in asset age profile between companies, establishing a cost variance is a complex undertaking which would need to understand and control for choices made by companies regarding their asset upgrade programmes. For this reason, and due to the limited regulatory data available, this factor was not studied further.

	ANH	NES	NWT	SRN	SVT	SWT	TMS	WSH	WSX	YKY	INDUSTRY TOTAL
Sum of Sewer age profile (constructed pre-1880)	1.26%	0.24%	8.86%	6.07%	8.19%	1.13%	10.01%	6.45%	0.49%	8.04%	**6.33%**
Sum of Sewer age profile (constructed 1881-1900)	13.85%	0.38%	0.11%	1.92%	1.44%	3.36%	10.00%	11.13%	7.63%	0.02%	**5.61%**
Sum of Sewer age profile (constructed 1901-1920)	5.54%	5.50%	8.44%	7.14%	4.71%	2.48%	11.67%	4.61%	0.58%	6.99%	**6.77%**
Sum of Sewer age profile (constructed 1921-1940)	13.51%	15.94%	15.29%	15.73%	21.07%	4.12%	23.01%	7.30%	16.16%	19.48%	**17.23%**
Sum of Sewer age profile (constructed 1941-1960)	16.20%	25.84%	19.51%	14.63%	20.26%	9.81%	10.43%	14.73%	13.65%	22.52%	**16.84%**
Sum of Sewer age profile (constructed 1961-1980)	22.13%	23.76%	23.37%	27.07%	24.34%	37.40%	14.38%	23.78%	29.33%	18.52%	**22.08%**
Sum of Sewer age profile (constructed 1981-2000)	25.44%	23.83%	23.57%	19.17%	13.20%	35.03%	11.08%	20.76%	18.37%	21.87%	**19.07%**
Sum of Sewer age profile (constructed post 2001)	2.07%	4.51%	0.85%	8.26%	6.79%	6.67%	9.42%	11.24%	13.79%	2.56%	**6.08%**
Sum of over 50 yrs old	50.36%	47.91%	52.21%	45.50%	55.66%	20.90%	65.12%	44.22%	38.52%	57.05%	**52.77%**
Sum of up to 50yrs old	49.64%	52.09%	47.79%	54.50%	44.34%	79.10%	34.88%	55.78%	61.48%	42.95%	**47.23%**

Table 47: Asset age profile based on Ofwat 2016 datashare

A3.10 ASSET ACCESSIBILITY

The distribution of key populations in the company supply areas and associated distribution of assets in the supply area influences operation and management of the assets. Access to remote assets is one area of impact which results in greater operational cost/person associated with smaller, more rural assets; staffing requirements are another. High-level analysis was undertaken to explore the variation in journey times across two water company areas, based on the national roads database and the Highways Agency's journey time 'Accession' database which provides information on road speed.

However, the construction of evidence was problematic due to a weak narrative and difficulty establishing realistic travel times in multiple scenarios, as staffing needs are met from decentralised locations (rather than from a single, central location) and may vary significantly on a day-to-day basis. The impact of remote populations and disproportionate cost of remote assets was considered to be more usefully addressed under the sparsity factor.

A4 BASE DATA

Appendix A4 summarises the key publicly-available (freely obtainable by download or straightforward request) datasets which were used to develop the engineering and environmental evidence. In addition to public data, key drainage data was purchased from the Met Office (annual rainfall and rainfall statistics) and Centre for Ecology and Hydrology (soil permeability data).

The wastewater industry datashare 2016 was made available to the project, as well as some carefully identified data from the project sponsor United Utilities, which was used for testing and verification of particular hypotheses or identification of gaps.

A4.1 ENVIRONMENTAL AND ENGINEERING DATASETS

NO	DATASET	SOURCE
1	EU Urban Wastewater Treatment Directive Database (>2000 PE)	European Environment Agency Waterbase (last modified March 2016)
2	EA Consented Discharges to Controlled Waters with Conditions (2015)	Environment Agency via data.gov.uk
3	Master Wastewater Datashare, October 2016	Ofwat; updated December 2016
4	Standard Annual Average Rainfall (SAAR) 1941-1970,	CEH
5	Met Office SAAR maps 1981 – 2010, rainfall days	UK Met Office
6	HR Wallingford Winter Rainfall Acceptance Potential (WRAP)	CEH
7	Met Office/CEH 5yr 25-day rain, 2- and 24-hr max rainfall	UK Met Office
8	HR Wallingford/ UKCP rainfall intensity mapping/ statistics	UK Met Office
9	Environmental Permit Regulations – Industrial (limited, publicly-available dataset)	Environment Agency via data.gov.uk
10	Heritage designations dataset	Heritage England via data.gov.uk
11	Environmental designation dataset	Natural England via data.gov.uk
12	Environment Agency groundwater emergence map (England only) Environment Agency groundwater rebound data (limited)	Environment Agency
13	National Flood Zone 2 dataset v201608 National Flood Zone 3 dataset v201608	Environment Agency via data.gov.uk
14	Published Academic PhD dataset for Sheffield City	Arup
15	Agricultural Land Classifications	Defra via data.gov.uk
16	Rural and Urban Classifications, RUC	Defra via data.gov.uk
17	June returns (Historical 2006-2011)	Ofwat via Ofwat.gov.uk archives
18	National Roads Database – isochrones travel data Accession database (road speed)	Highways Agency
19	IMD scores by LSOA	Office of National Statistics, ONS
20	Urban Regions 2001 & 2011 GIS layers	Office of National Statistics, ONS
21	Local Authority IMD scores	Office of National Statistics, ONS
22	UK housing types	Office of National Statistics, ONS
23	Consented Discharges to Controlled Waters with Conditions (2017)	Natural Resources Wales

Table 48: Environmental, demographic and engineering datasets

A4.2 UNITED UTILITIES SAMPLE DATASETS FOR TESTING AND VERIFICATION

- Phosphorus permits (consents);
- sample CSO permits; and
- sewer flooding data

A4.3 THIRD-PARTY REGIONAL WAGE DATASETS

Data for 2015 from two third-party datasets was obtained from pay benchmarking companies working across a variety of sectors in the UK. These are:

- Willis Towers Watson (TW) Manufacturing, Distribution and Services (MDS) Survey 2016, for which a sample of the survey data was obtained, which covered benchmarking information for engineering, including the water industry. Typically, the last 3 years of data are available to purchase.
- HAYS UK Salary and Recruiting Trends 2016, which represents the most recent salary survey data produced by the company. The datasets are freely available, and were produced 2015 and 2014 as well.

Commentary on the datasets

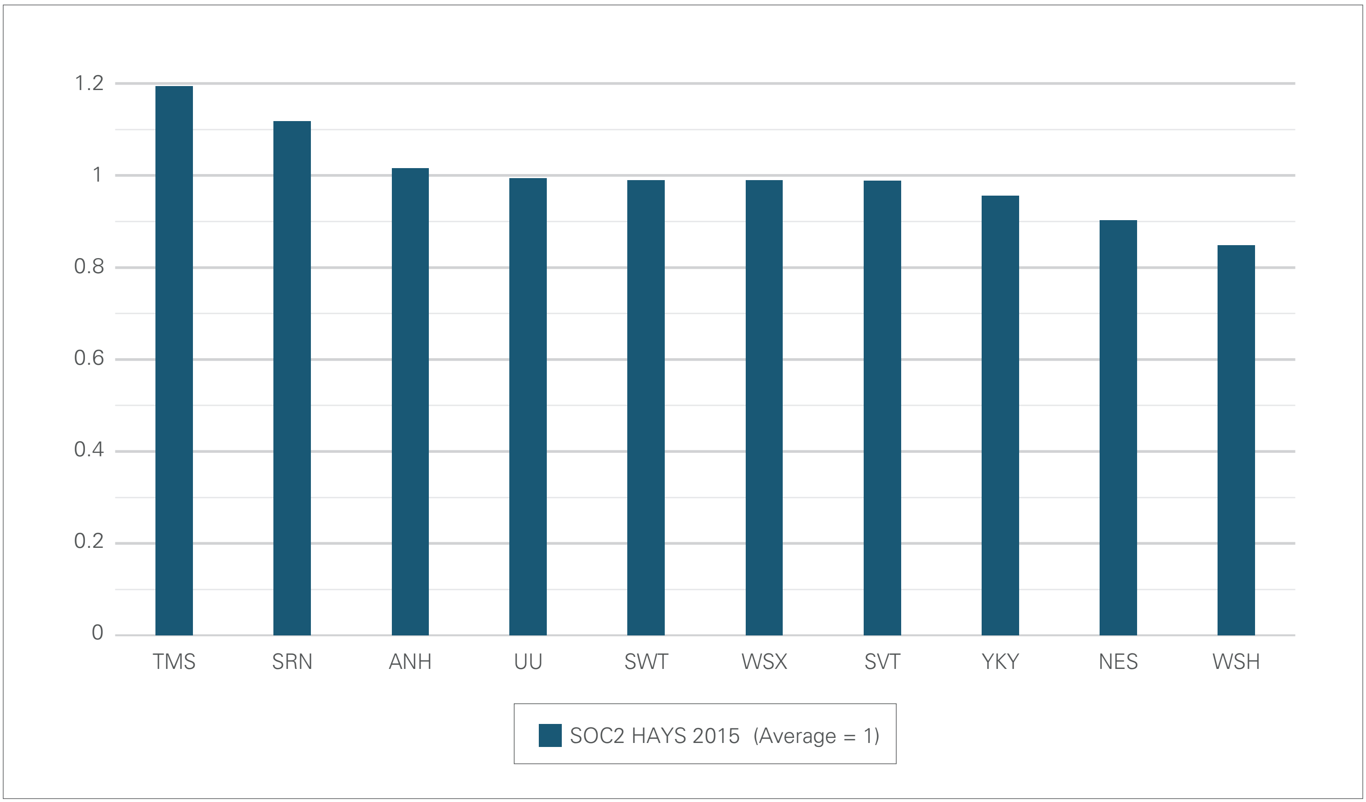
The private datasets are commonly used to benchmark salaries across a variety of sectors, and are built up from salary and recruitment surveys covering a broad range of sectors and jobs. These include construction, engineering, manufacturing, utilities, retail, services, aerospace, defence, procurement and supply and office support. HAYS data is compiled from client jobs filled every year (actual salary, rather than advertised salary), while TW data is compiled from salary surveys with participating companies, including utilities and those in the water and wastewater sector. Private data sets are, therefore, likely to be a more accurate reflection of the market in any given year. However, the richness of the datasets is dependent on survey respondents or on job demand and can be variable from year to year.

Geographical (regional) categorisation of the datasets is based on ONS regions, although not all jobs and roles in the surveys cover all regions. For example, some engineering and manufacturing jobs were missing in some regions in the HAYS dataset. The methodologies used to collate the datasets are generally not stated or provided, and statistical analysis to confirm suitability of the datasets not possible. Job descriptions, however, are better defined and more appropriate to the sector than the ONS job descriptions.

Analysis

The HAYS data sets was used with Ofwat's spreadsheet which was used to develop company-specific labour cost indices for wastewater only. ONS data was replaced with HAYS data for SOC1 and SOC2, using the occupational weights provided by the WaSCs. The wastewater company indices constructed in this way were compared to the respective average Ofwat company indices for 2015, and showed that Thames Water and Southern Water were above average, while the rest of the companies were at or just below average (Figure 22).

This demonstrates the potential for third party datasets to be used to construct labour indices, although the dataset limitations make them unsuitable for regulation. The limitations are discussed in detail in Section 5.3.



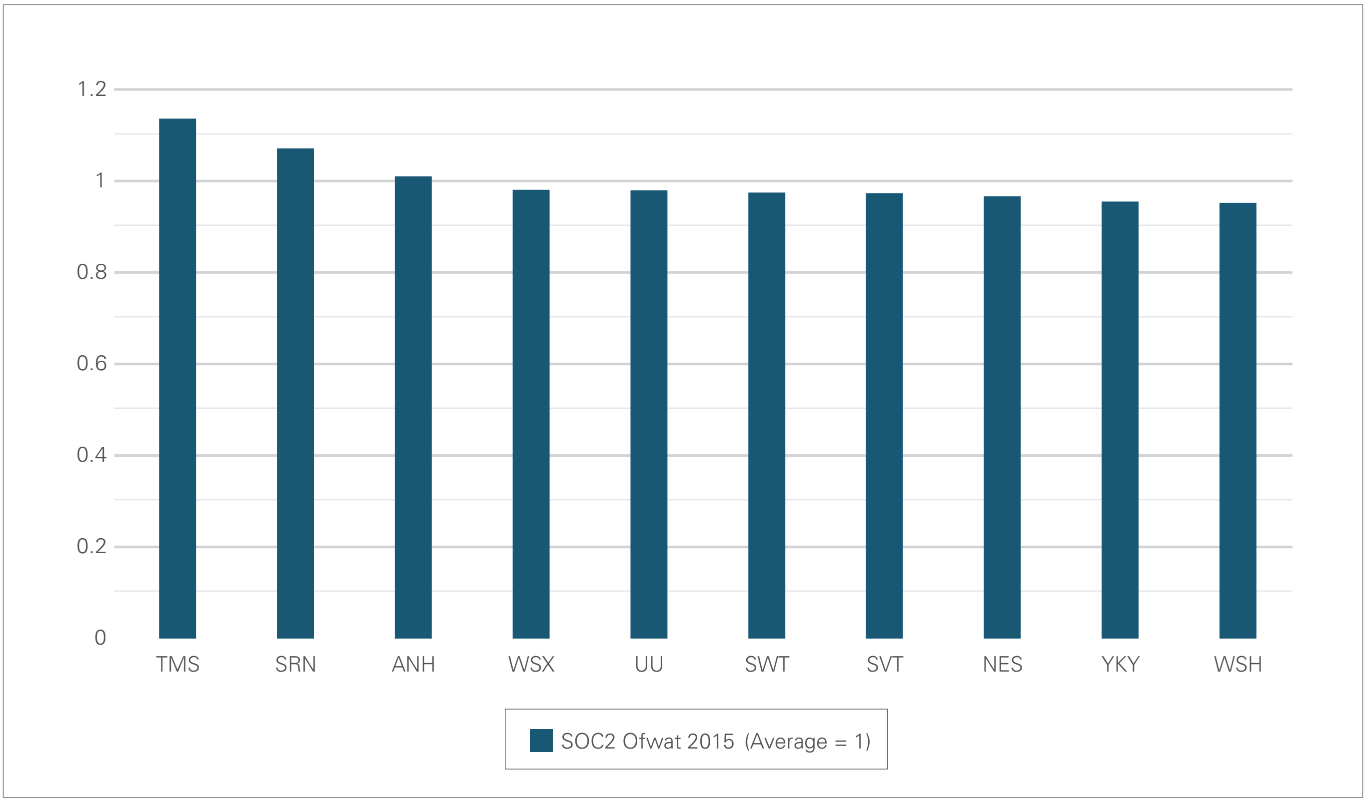


Figure 22: Company labour index (wastewater only): HAYS 2015 vs Ofwat 2015

Appendix B: Econometric background

This appendix details the econometric modelling exercise which identifies and explores issues related to the wastewater modelling approach to be taken by Ofwat in the 2019 Price Review.

Modelling work is structured as follows:

Replicating PR14

In order to explore the desirability of alternative approaches to the status quo (that is, the PR14 models), this exercise first replicates the PR14 methodology. This ensures that model estimation produces the same results as those estimated by Ofwat and CEPA during PR14.

Updating PR14 models with recent data

PR19 will use the most recent data available to determine the allowable cost thresholds for the industry. Thus even if the same models are used, the estimates may change with an updated data series. The updated estimates were used as a base case, to explore whether and how the modelling approach might change. The statistical tests used in PR14 were replicated as closely as possible before supplementing with additional tests.

Exploring alternatives to PR14 models using alternative specifications and an expanded dataset

Alternative models were tested on a dataset that was expanded both with recent observations of PR14 data lines and with alternative data lines collected by Arup and Vivid. The use of these alternative data lines was motivated by engineering and economic hypotheses about weaknesses in the PR14 models.

B1 PR14 REPLICATION

PR14 tested ten models, summarised in the table below, five of which were selected by CEPA and Ofwat as sufficiently robust to be reflected in the final determination of costs. Replicating the five selected models was a priority, but SW2 was also replicated because of the availability of the data lines the model requires. The data lines unique to the other models that were rejected from PR14 were not available, and thus no attempt was made to replicate or update their results.

There were two barriers to replicating the PR14 models: slight data changes and different software approaches to implementing Feasible Generalised Least Squares (FGLS). The published data available from Ofwat appears to have truncated some of the decimal places used in the original CEPA coefficient estimates, so the coefficient estimates from this estimation exercise differ slightly from the published originals. The magnitude of this difference is quite small, a tenth of a per cent or less in most cases. Thus the models that use OLS replicate the published originals quite closely, but the second barrier results in substantially different estimates for GLS models.

MODEL	APPROACH	REPLICATION ATTEMPTED?	REPLICATION SUCCESSFUL?
SW1	Sewage network GLS RE	Yes	No – FGLS implementation
SW2	Sewage network COLS	Yes	Yes
SW3	Sewage treatment and sludge (full) GLS RE	No – rejected from PR14, requires extra data lines	
SW4	Sewage treatment and sludge Cobb-Douglas GLS RE	No – rejected from PR14, requires extra data lines	
SW5	Sewage treatment and sludge (refined) GLS RE	Yes	No – FGLS implementation
SW6	Sewage treatment and sludge (refined) COLS	Yes	Yes
SW7	Wholesale sewage (full) GLS RE	No – rejected from PR14, requires extra data lines	
SW8	Wholesale sewage (full) COLS	No – rejected from PR14, requires extra data lines	
SW9	Wholesale sewage (refined) GLS RE	Yes	No – FGLS implementation
SW10	Wholesale sewage (refined) COLS	Yes	Yes

Table 49: PR14 models

Note: Sources are Vivid Economics and Annex B CEPA report.

Many of the PR14 models use GLS panel methods to estimate model coefficients. As noted in Section 3.3, because all of the observations are drawn from 10 companies over time, there is reason to believe that the error terms are correlated with one another (serial correlation) or have non-constant variance (heteroscedasticity). This implies that a correctly specified OLS model will be consistent, but not efficient. GLS corrects for these non-spherical errors by transforming the variance-covariance matrix into one that is uncorrelated (homoscedastic and serially uncorrelated). GLS can only be performed if this transform is known, but if the structure of the variance-covariance matrix is not known, a consistent estimate of the covariance of the errors can be generated and used instead. There are a number of ways to generate estimates that generate extremely similar results for correctly specified models with large sample sizes.

The Stata and Limdep implementations of the GLS Random Effects (RE) regressions generate coefficient estimates that differ substantially. This has been independently confirmed by both Vivid Economics and Ofwat's academic advisor. Section 3.3 presents evidence from using three different methods: Stata FGLS, Stata MLE, and Limdep, which implements a Fuller-Battese transformation. All three simply represent different approaches for estimating the structure of the variance-covariance matrix of the main regression of interest.

The three estimation methods are all consistent, but have different small sample properties. The standard FGLS approach in Stata is probably the most common, and would likely be considered the 'default' approach. The MLE is the most conservative method in the possible presence of model misspecification because it is non-parametric. Fuller-Battese is expected to be the most efficient provided that the error correlations have a nested structure, which is plausible but more restrictive than either of the other approaches.

Variance decompositions suggest that most of the variability comes from differences between companies rather than differences over time. The table below tabulates the percentage of variance that GLS estimates assign to differences between companies. These figures vary substantially depending on the total variability in models, which depends on specification, but the high percentages suggest that there is little variability over time. This lends additional evidence to GLS being poorly suited to this modelling exercise.

	NETWORK	T&S	BOTEX
PR14	73%	27%	59%
PR14+	68%	61%	66%
New models	84%	71%	86%

Table 50: Variability attributable to differences between companies in GLS models

Note: Reported figures are estimates of rho, which is the between company variance divided by the sum of within and between variances

B2 PR14 UPDATE

In order to extend base cost thresholds to account for updated information through to the 2015/16 financial year, the following two steps are required.

1. Generate model coefficient estimates in Stata, replicating PR14 model methodology as closely as possible.

2. Use the model coefficients to generate new base cost thresholds by using the 2015-2020 forecasts generated by Jacobs in 2013 for PR14. Note that this real 2015 data was used to generate coefficients (step 1) and those were then tested using 2015 forecasts in an effort to illustrate how the models perform with more recent information.

The remainder of this section describes the first step.

A ten year panel was created by combining two published Ofwat datasets. In PR14, nine years of sewerage data was available, the first two of which were deemed anomalous due to extreme flooding events, resulting in a seven year time series between 2006 and 2013. The Ofwat datashare includes information between 2011 and 2016, allowing the panel to be extended to the ten years between 2006 and 2016. The resulting data was used as a seven year series (an ancillary regression, for consistency with PR14's seven year series) and a ten year series, which was subsequently used as the base case for comparisons with alternative estimation methods.

The available data sets used to create a panel have small discrepancies between the overlapping 2011-2013 data or have missing observations. Such gaps reinforce the points made in Section 5 regarding the importance of collecting accurate industry information in future datasets. The remainder of this section outlines the steps taken to prepare the existing datasets in turn.

Sewer length experiences a 2-3 per cent decrease in 2011/12 and 2012/13 values between PR14 and the Ofwat datashare. This drop is likely to be artificial and due to a change in measurement or estimation method, and its magnitude varies by company. The number of connected properties experiences a similar difference, though it is smaller in relative magnitude. Three alternative methods to address these discrepancies were considered:

1. **Not use historic data from PR14.** This option would shorten the time series to only 5 years, raising concerns about the robustness of limiting each company to only five data points, or a total of 50 points when pooled across years and company.
2. **Use historic data from PR14 without modification.** This option would use the PR14 data from 2006-2011 and Ofwat datashare information from 2011-2016. The switch would tend to attenuate coefficients.
3. **Use historic data from PR14 with modification.** This option would modify the 2006-2011 PR14 data proportionately based on the average ratio of the 2011 and 2012 PR14 data to the Ofwat datashare. The ratio would be company specific, because it is likely that measurement changes may vary substantially by company. It seems unlikely that companies are strategically misreporting sewer length information in a way that would bias results due to the high penalties for misreporting.

The third option was the one adopted in the estimation in this report.

Load is recorded in the Ofwat datashare, but only for 2015/16, leaving a two year gap where reported load information that includes total loads is incomplete. There are again three possible options:

1. **2013/2014 and 2014/2015 information could be dropped from analysis.** Leaving two of three recent years out of an analysis intended to test updated information would both reduce an already limited panel length and runs contrary to the point of the exercise.
2. **A linearity assumption could be made for 2013/14 and 2014/15, treating 2012/13 and 2015/16 as endpoints.** This would mean reduced variability for estimation in the dataset and is likely to increase multicollinearity with the time trend used in PR14 style specifications.
3. **Total load could be estimated using reported PE served.** The datashare records information about PE served, and Ofwat calculates PE based on an assumption of 60g BOD/person/day. This means that total load can be estimated for the missing years, though the total loads calculated in this way deviate from reported total loads by up to 5 per cent for 2015-2016. Part of this error may be due to the fact that companies estimate PE differently, treating non-resident and migrant populations in a variety of ways, as discussed in Section 5.2.

The third approach is the one adopted for this report.

Percentage of load treated in works in size bands 1-3 also has a two year gap for 2013-14 and 2014-15. As with load, there is a gap of two years for load distribution between works in size bands 1-6. This was addressed by assuming the proportion of load treated by each band changes linearly between 2013 and 2015. Multiplying the proportion by the estimated total load for the 2013/14 and 2014/15 observations yields an estimate of load treated in each band. The linearity assumption does not appear to be particularly accurate, especially for lower bands (representing smaller treatment works). Other functional forms have not been explored and graphs of past data do not suggest an obviously appropriate form.

Information on regional wages was not collected in the Ofwat datashare, so there is no official information for the most recent three years. CEPA analysis for PR14 used ONS data from regional hourly wage surveys (ONS Table 15.6a) to construct company-specific average wage estimates. The following four steps were followed attempting to recreate CEPA's estimates:

1. SOC21 and SOC53 wages were collected for each ONS region and regional averages were constructed using the CEPA weighting of 60 per cent SOC53 and 40 per cent SOC21
2. Wastewater companies that fell exclusively within a single ONS region were assigned that regional average wage.
3. Wastewater companies that fell across multiple regions were assigned a population-based weight, calculated by Ofwat, to give the weighted average of the wages in the relevant regions.
4. The resulting nominal company wages were then adjusted for inflation. It is not clear from the CEPA methodology how this was done, but this analysis has averaged monthly ONS 2015 RPI from April to March (to align with the reporting years used by Ofwat), rebased the RPI to 2013, and then used it to convert nominal wages into real wages. The real wages calculated differ from the regional wages reported in PR14 by ~1-5 per cent. It is likely that this discrepancy comes from changes in the ONS methodology to collect SOC information since PR14.

The same methodology was used to extend the sample to the three most recent years.

PR14 uses smoothed wastewater costs, but there was a difference between costs reported in PR14 versus those in the datashare. Both PR14 and the Ofwat datashare collect opex and base capex for the total, collection, and combined treatment and sludge, though the Ofwat datashare offers further disaggregation. Base capex was smoothed using a moving average including the year and the four previous years, consistent with practice at PR14. As with the variables discussed above, there is some discrepancy between 2011-2012 and 2012-2013 data between the two data sources, most likely reflecting adjustments made since publication of the earlier dataset. In part because the updated cost values are likely to be more accurate, and in part because it is unclear how one might try to apply a correction factor to unsmoothed costs which naturally vary substantially between years, Ofwat datashare cost information was used to replace PR14 values during the two years of overlap.

Table 51 summarises the adjustments made to data in preparing the panel for analysis.

VARIABLE	COMPARABILITY ISSUE	ADJUSTMENT MADE
Sewer length	2-3% difference for overlapping years	Replace overlap years with more recent data, create company specific adjustments and apply to PR14 data
Number of properties	2-3% difference for overlapping years	Replace overlap years with more recent data, create company specific adjustments and apply to PR14 data
Load	Missing 2013/14 and 2014/15 data	Estimate total load based on PE
Share treated in bands 1-3	Missing 2013/14 and 2014/15 data	Assume linearity in proportion of total load treated in each band
Regional wage	Missing 2013/14 – 2015/16 data	Recreate methodology with newly gathered data
Wastewater costs	Difference for overlapping years	Replace overlap years with more recent data

Table 51: Summary of data adjustments for PR14 update

B3 COMPARING ALTERNATIVE MODELS

The approach adopted for PR14 uses several methods to select models for use in cost predictions in different ways.

Determining variables to be included.

This was based on three criteria: statistical significance, sector significance and whether the result was consistent with expectations. The first is a purely statistical criterion while the latter two rely on economic theory and intuition to decide whether a given variable should be in a regression model. For some variables, there was a prior on both sign and magnitude, such as wages, while for others there was only a prior on sign.

Testing model agreement.

Company efficiency scores are based on the deviations from modelled trends, resulting in a ranking of company performance. A comparison of the resulting rankings across different model specifications was used as a way to test model agreement and stability.

Examining robustness to outliers.

The approach in PR14 considered the effect of dropping variables and outliers on the stability of coefficients of variables in the regressions.

Alternative panel data specifications.

The Hausman test has also been used to select between panel data random effects (RE) and fixed effects (FE) models.

This estimation exercise has adopted the first two comparison methods and has added the following additional statistical indicators.

Correlation matrices

Correlation matrices: correlations were generated for new variables prior to including them into a regression as a first check for multicollinearity.

Variance Inflation Factor

After estimation, a Variance Inflation Factor was created. Variables with a VIF of much more than ten suggest a high degree of collinearity with other variables in the model. Though VIFs of more than ten should not disqualify a model from use, it can be an informative diagnostic about model multicollinearity.

Joint variable F tests

Overall model fit was assessed using joint variable F tests in addition to the adjusted R squared values mentioned in PR14. It should be noted that adjusted R squared is not a robust statistical measure for the GLS class of models, and the values were quite high for all models, limiting their utility as a comparative metric.

Ramsey Regression Equation Specification Error Test (RESET)

The functional form was considered using the Ramsey Regression Equation Specification Error Test (RESET) using robust standard errors, which is a test for broad model misspecification. One possible explanation for poor performance on this test is omitted variables to the extent they are correlated with the higher order variables used in this test. However, omitted variables are extremely difficult to detect statistically due to the fact that they are absent from analysis.

The results of all statistical tests form a part of the evidence base used to evaluate a model and are not relied upon exclusively. Coefficient significance in particular is easily over-emphasised, since the overall goal of the models should be to accurately explain and predict industry costs, which is not necessarily directly correlated with the significance of particular driver coefficients. With an extremely small sample size, this report approaches the explanation of costs by relying in addition on clear engineering evidence to motivate inclusion or exclusion of particular drivers.

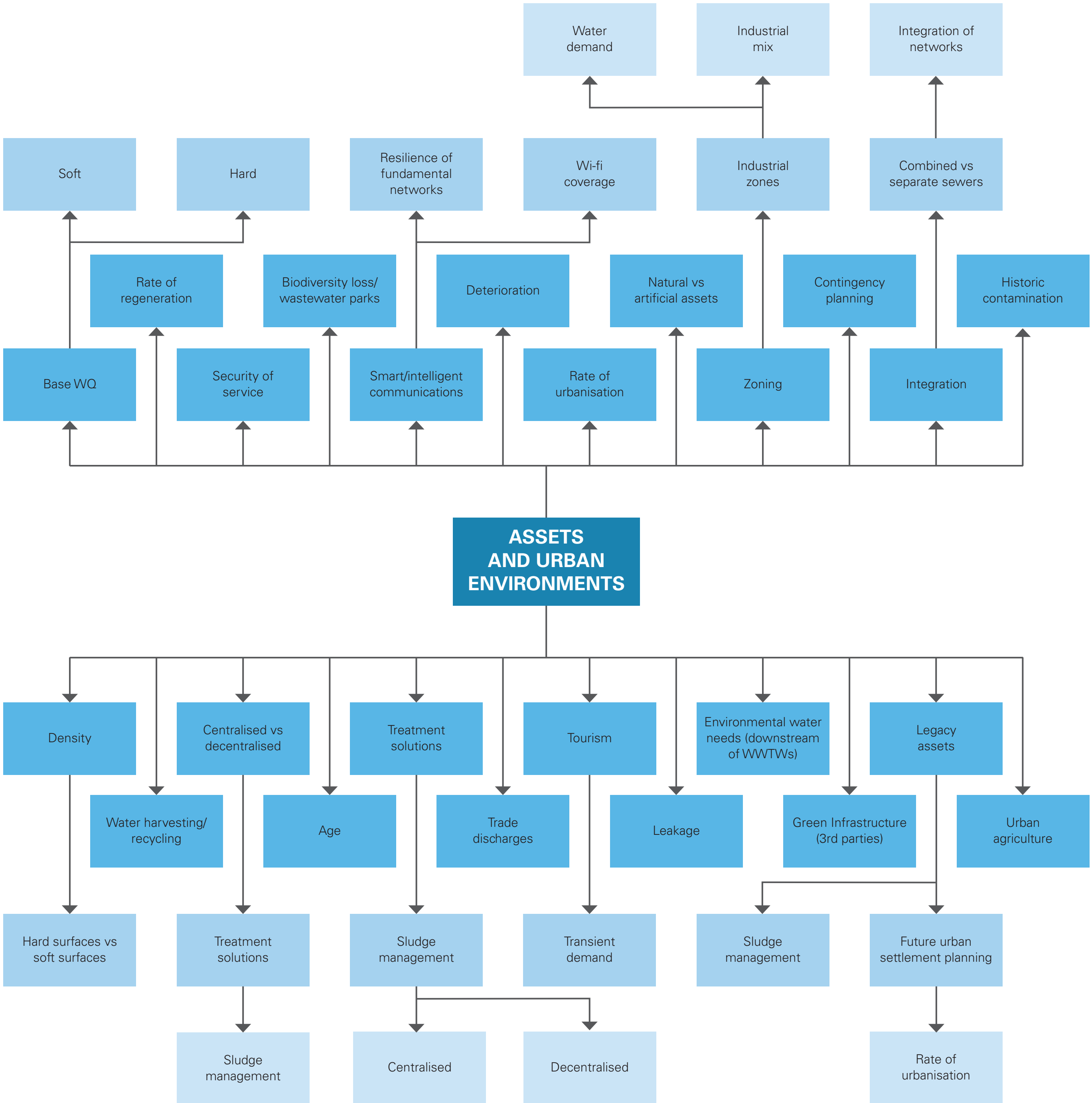
Appendix C: Long list of 200+ possible factors

Appendix C presents the 200+ factors which were explored at the outset as possible drivers of wastewater service cost.

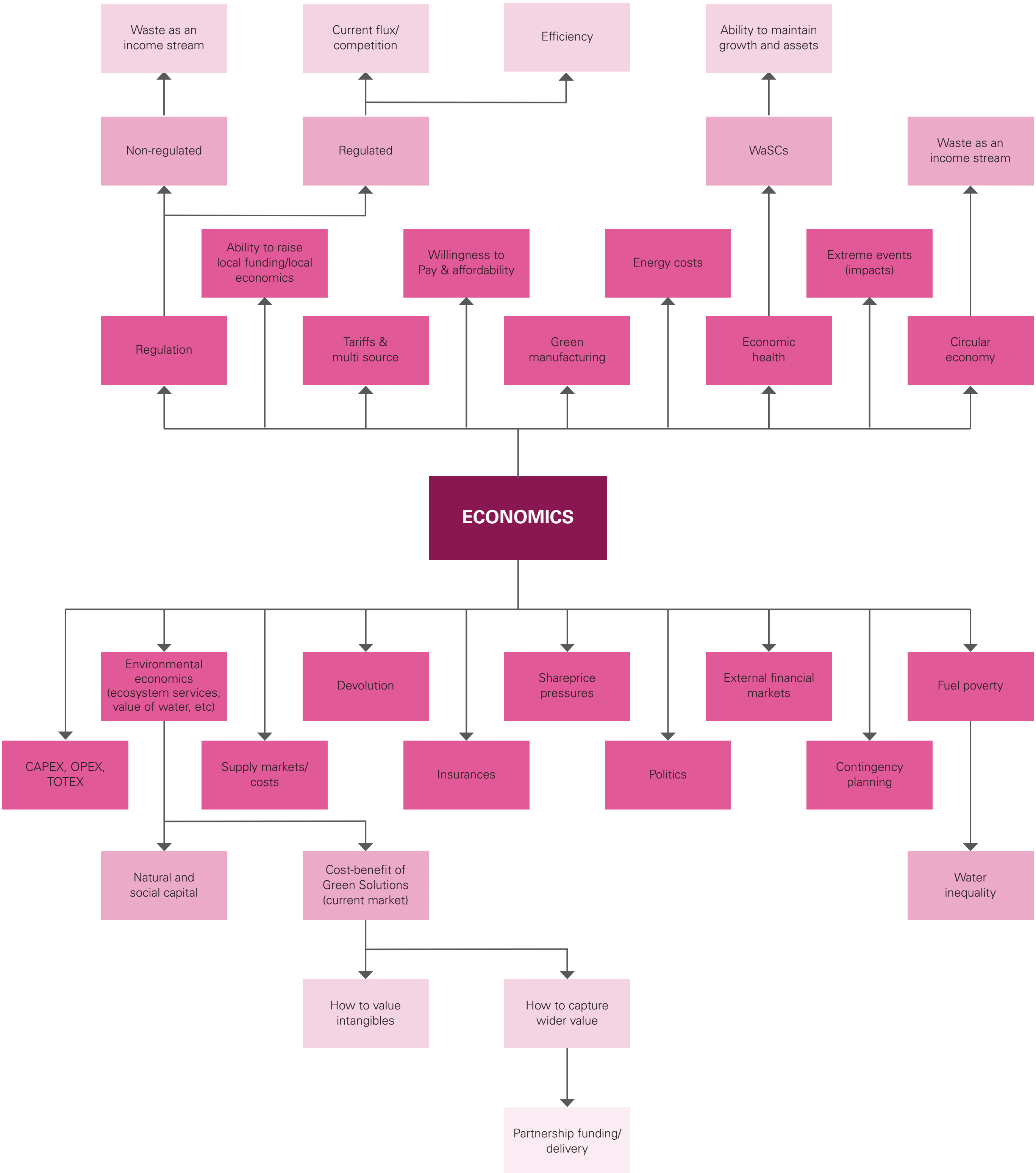
These are presented in the form of mind maps in order to show the possible linkages that were discussed.

The mind maps are presented for information only, and are grouped into seven categories: assets and urban areas; economics; geography; governance; people and society; stakeholders and other third parties and weather and climate.

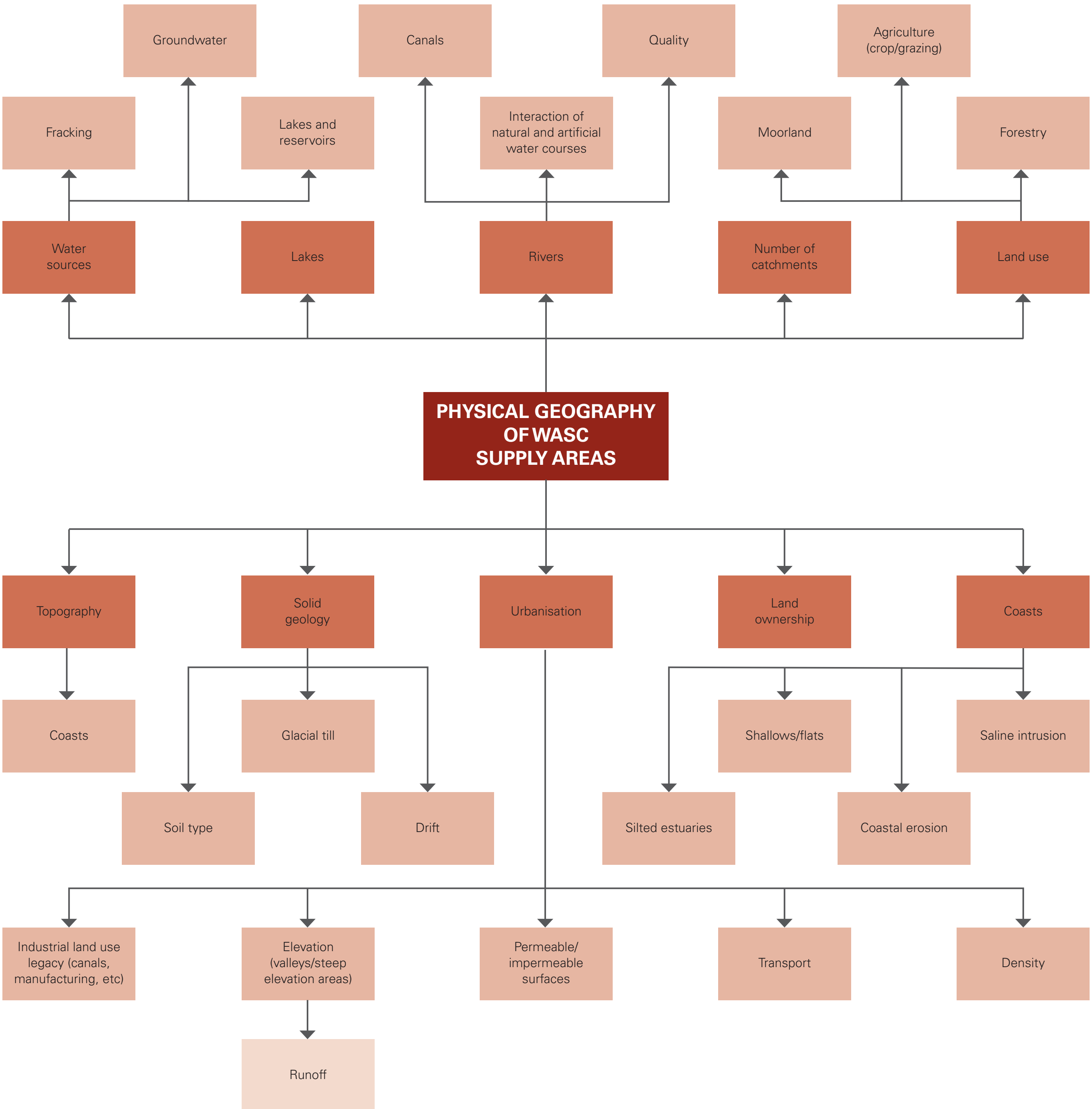
C1 ASSETS AND URBAN ENVIRONMENTS



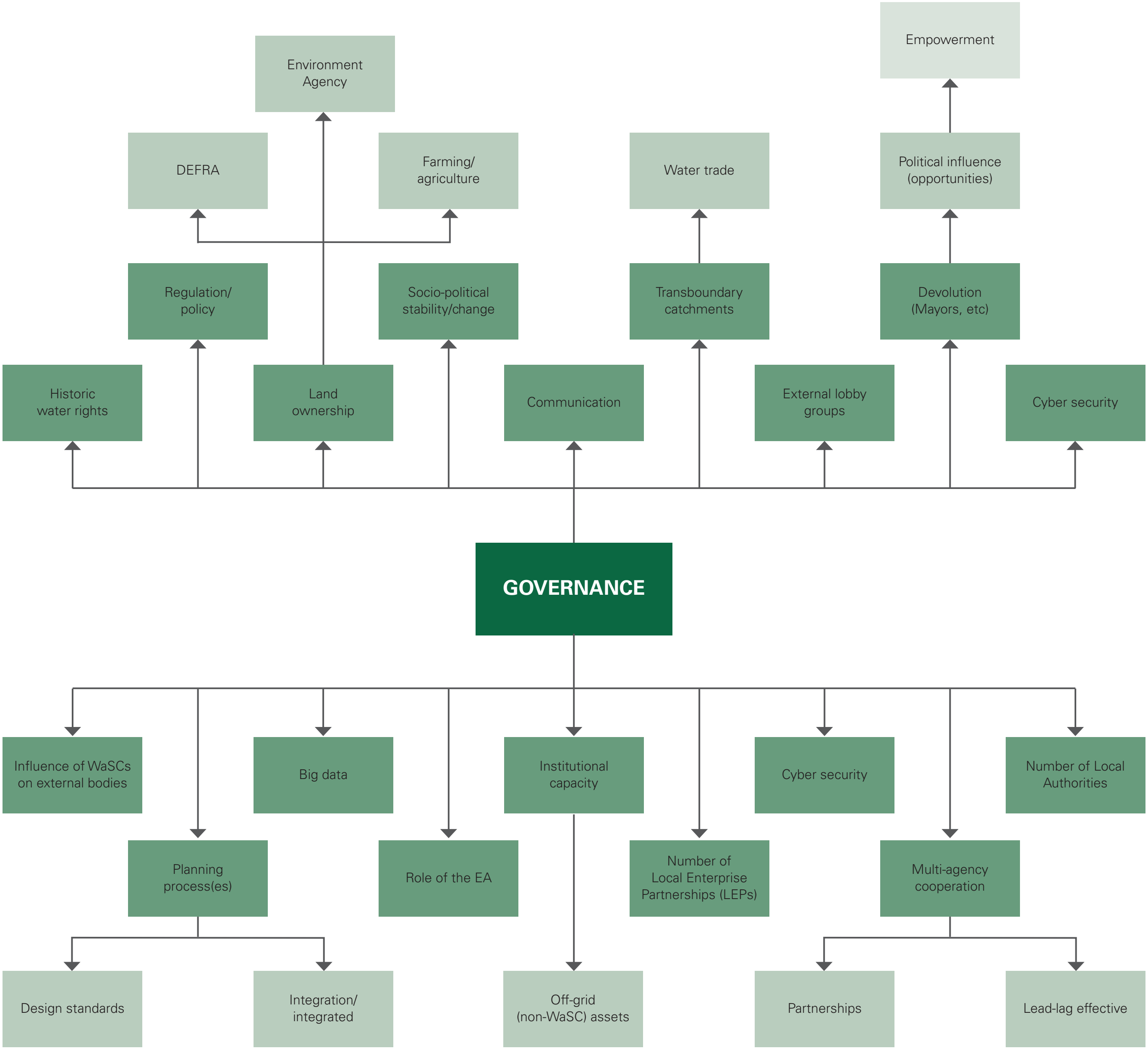
C2 ECONOMICS



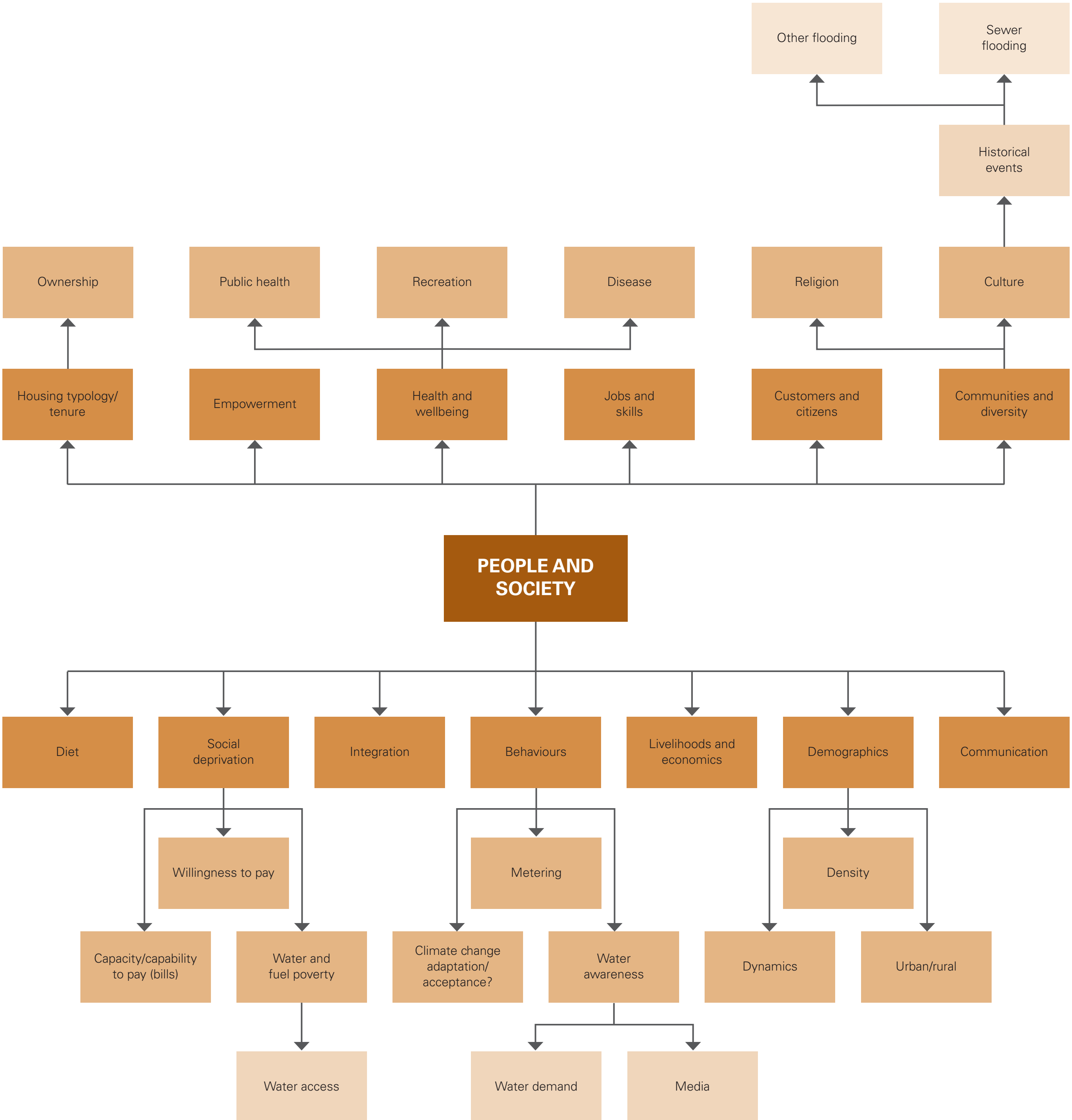
C3 PHYSICAL GEOGRAPHY OF WASC SUPPLY AREAS



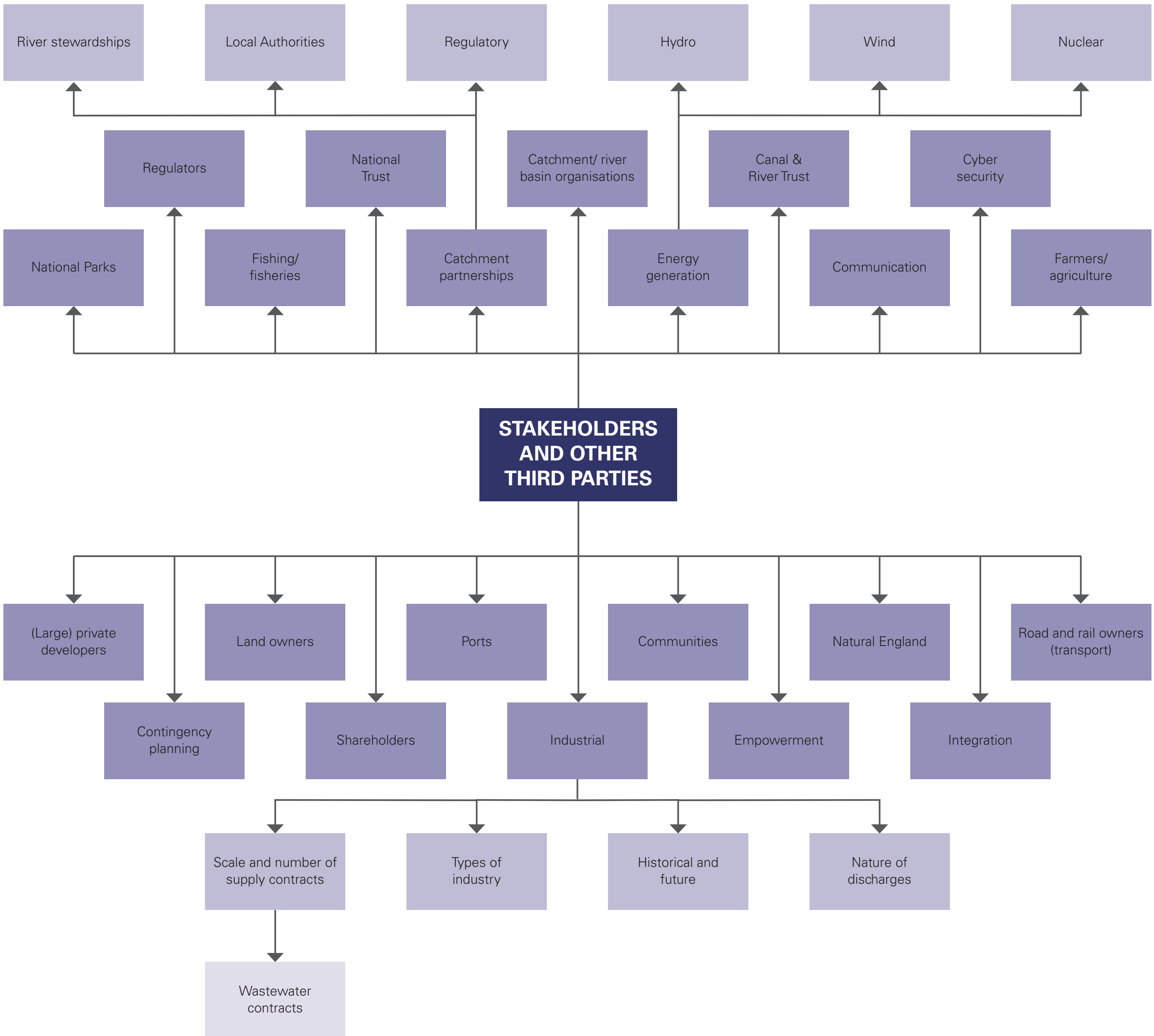
C4 GOVERNANCE



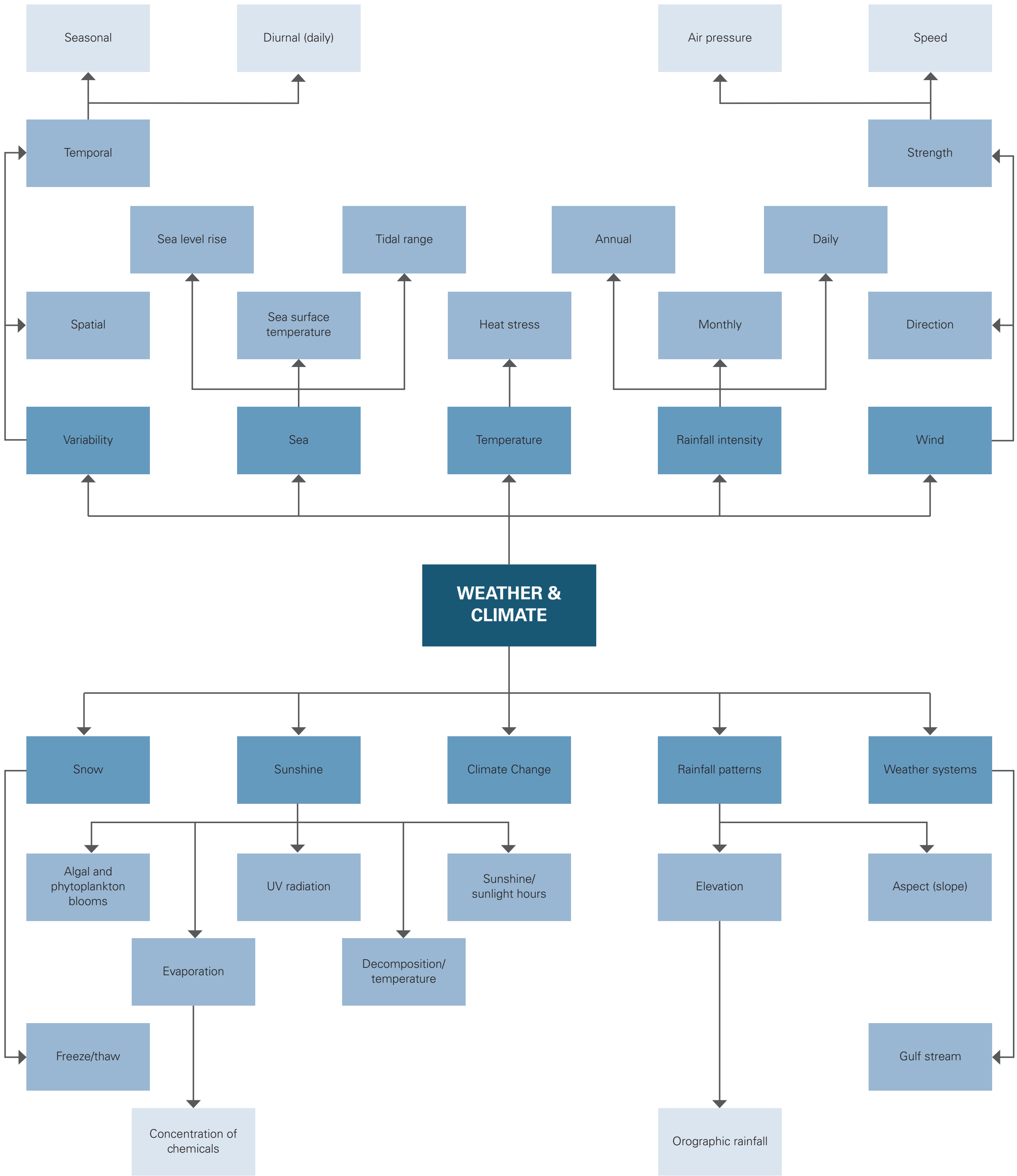
C5 PEOPLE AND SOCIETY



C6 STAKEHOLDERS AND OTHER THIRD PARTIES



C7 WEATHER AND CLIMATE



LIST OF ABBREVIATIONS

ABBREVIATION	MEANING
AMP	Asset Management Period
ASP	Activated Sludge Plant
BOD	Biochemical oxygen demand
CEH	Centre for Ecology and Hydrology
CSO	Combined Sewer Overflow
DWF	Dry Weather Flow
EA	Environment Agency
GLS	Generalised Least Squares
OLS	Ordinary Least Squares
ONS	Office of National Statistics
PE	Population Equivalents
PR14	The price review undertaken in 2014
PR14+	The models or supporting time series dataset used at PR14, extended by three years to 2015-16 using Ofwat 2016 datashare
Ramsey RESET	Ramsey regression equation specification error test
RIIO ED1 and RIIO GD1	Ofgem price reviews for electricity and gas distribution network operators, employing the RIIO (Revenue = Incentives + Innovation + Outputs) approach to setting company revenue.
SAF	Submerged Aerated Filter
VIF	Variance Inflation Factor
WaSC	Water and Sewerage Company
WoC	Water only Company
WRAP	Winter Rainfall Acceptance Potential, which indicates infiltration potential

CREDITS

PROJECT LEADS

Arup Project Director: **Ian Gray**

Arup Project Manager: **Philip Songa**

Vivid Economics Project Director: **Robin Smale**

Vivid Economics Project Manager: **Oliver Walker**

INTERNAL PEER REVIEW ERS

Dr. Mark Fletcher
Global Water Director, Arup

Steven Lloyd
Transactions Director, Arup

Robin Smale
Director, Vivid Economics

EXTERNAL PEER REVIEWERS

Dr. Julia Martin-Ortega
Associate Professor of Ecological Economics, University of Leeds

Dr. Thijs Dekker
Lecturer in Transport Economics, University of Leeds

Dr. Kieran Conlan
Strategic Business Director, Ricardo Energy & Environment

Dr. Paul Leinster
Professor of Environmental Assessment, Cranfield University

Dr. Ralf Martin
Assistant Professor in Economics, Imperial College Business School

Dr. Melvyn Weeks
Assistant Professor in Economics, University of Cambridge

ACKNOWLEDGEMENTS

The project team wishes to thank all of the staff at Arup and Vivid Economics who have contributed to the report, as well as the peer reviewers and other parties who have kindly shared knowledge, supporting information and insight, including the project sponsors, United Utilities.

DISCLAIMER

This report takes into account the particular instructions and requirements of our client. It is not intended for and should not be relied upon by any third party and no responsibility is undertaken to any third party. The information included in this work, while based on sources that the authors consider to be reliable, is not guaranteed as to accuracy and does not purport to be complete. The authors confirm that auditing of calculations and estimates was not part of the peer reviewers' scope.

:vivideconomics ARUP