artesia

Cambridge Water

Non-household demand forecasts 2020 to 2100

Project reference: 2465

Report number: AR1413

2021-05-13

www.artesia-consulting.co.uk

Report title:	Non-household demand forecasts 2020 to 2100
Project reference:	2465
Report number	AR1413
Date:	2021-05-13
Client:	Cambridge Water
Client lead:	Steve Colella

Project manager:	Rob Lawson
Technical lead	Sarah Rogerson
Contributors:	Francesca Cecinati, Jamie Urwin

Version	Author(s)	Reviewed by:	Description	Date
1	Rob Lawson, Dene Marshallsay, Sarah Rogerson	Rob Lawson	Draft	23/04/2021
1	Rob Lawson, Dene Marshallsay, Sarah Rogerson	Rob Lawson	Final	13/05/2021

Any enquiries relating to this report should be referred to the authors at the following address:

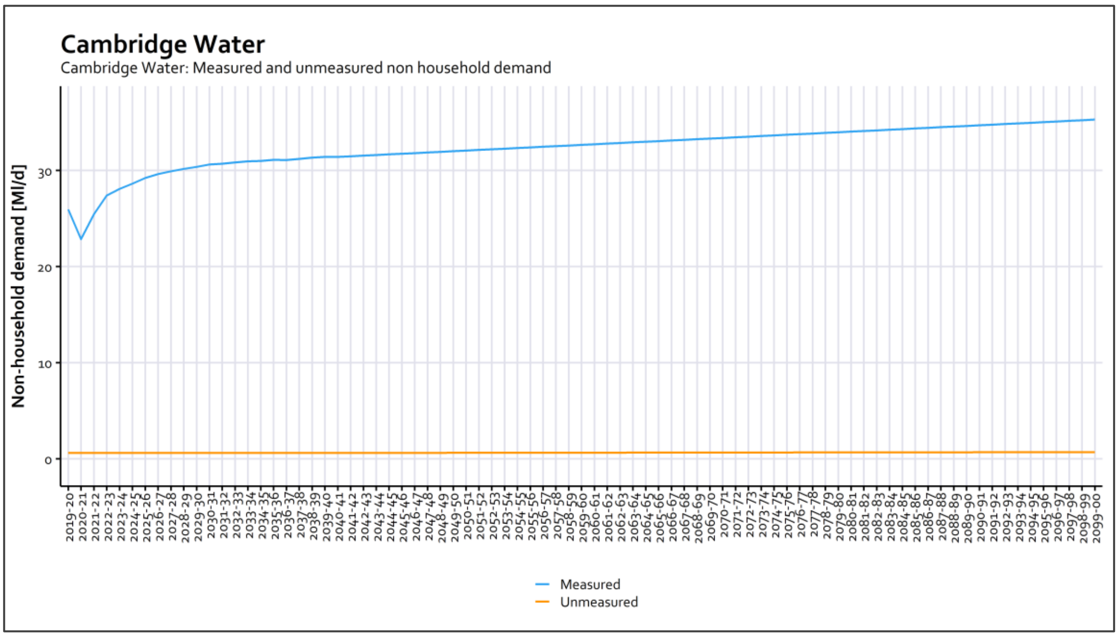
Artesia Consulting Ltd, Unit 2 Badminton Court, Yate, Station Road, Bristol, BS37 5HZ.

Telephone: + 44 (0) 1454 320091 Website: www.artesia-consulting.co.uk

Executive summary

Water companies in England and Wales are required to develop a Water Resource Management Plan (WRMP) under the Water Industry Act 1991 where they set out their plans to ensure that they will have sufficient resources to meet demand under different climate conditions over a minimum of 25 years. Forecasting future demand for water is a key part of the process and consumption by the non-household sector is a major component of demand. This report describes the initial development of the demand forecasts for non-households in the Cambridge Water (CAM) region.

We have produced a set of non-household demand forecasts for Cambridge Water from 2019-2020 out to 2099-2100. These are presented for metered and unmetered properties at company level and disaggregated by industrial sector. The approach used follows existing industry best practice. Robust multiple linear models have been produced for four cohorts of industrial sectors for Cambridge using explanatory factors that include population, gross value-added metrics, employment rates, population density and other factors.



We have presented the forecasts from a base year of 2019-20. The intermediate years 2020-21 through to 2024-25 are presented for information prior to the start of the planning period in 2025-26. These intermediate years are potentially volatile with a number of unknowns around the impact of the COVID-19 pandemic and the impact from Brexit on non-household consumption. Therefore, we recommend that the baseline and scenario forecasts are updated prior to the submission of the final water resource management plans.

There has been an unprecedented change in non-household demand in 2020, due to the policies introduced to combat the COVID-19 pandemic. This is the first year of this forecast and creates added uncertainty going forward as we still do not fully understand what the enduring impacts will be from changes in working practices, such as increased working from home.

The sector also faces a number of future unknowns in demand from non-households, such as population change, Brexit, climate change and how water efficiency will be delivered in the non-household sector. Since the last set of non-household forecasts were completed for

WRMP19, the non-household retail sector has undergone a transformation with the introduction of retail competition. We have observed a change in data quality and consistency since the change in 2017, which has complicated the modelling and has increased the uncertainty around the demand forecasts. Therefore, we have included all these factors in the scenario and uncertainty modelling.

The overall conclusion is that non-household demand in the Cambridge Water region at the start of the planning period (2025), is predicted to be 29.85 Ml/d (as presented in the graph above) within an overall range of 26.58 to 35.99 Ml/d. Much of the early range in possible demand is due to the impact of COVID-19 and uncertainty over the quality of non-household consumption data from MOSL. The baseline demand forecast is predicted to increase by the end of the planning period (2100) to 35.99 Ml/d (an increase of 6.14 Ml/d).

During the course of the work, we have identified a number of improvements that could be implemented for future forecasts. These are included in the recommendations section of the report and cover: improving data quality, investigating different industrial sectors, looking at modelling at a zonal level (taking account of the dominant influence of larger zones where relevant), as opposed to by company, and producing forecasts more frequently to reduce the step change transitions between forecasts every 5 years.

Glossary

Term	Description
A classification of residential neighbourhoods (ACORN)	This is a socio-demographic classification of neighbourhoods published by CACI Ltd. The system is based on the assumption that people who live in similar neighbourhoods are likely to have similar behavioural and consumption habits.
Abstraction	The removal of water from any source, either permanently or temporarily.
Active leakage control (ALC)	Management policies and processes used to locate and repair unreported leaks from the water company supply system and customer supply pipes.
Annual average demand	The total demand in a year, normally measured as the amount of treated water entering the distribution system at the point of production, divided by the number of days in the year.
Annual return	An annual report made to Ofwat by water companies to advise on progress within that Asset Management Period.
Asset management period (AMP)	Five-year period for which water companies are funded by Ofwat according to their Business Plans.
Base year	The first year of the planning period/horizon, forming the basis for the water demand and supply forecasting of subsequent years.
Baseline forecast	A demand forecast of customer consumption without any further water company intervention during the planning period. A baseline customer demand forecast should take account of: customer demand without any further water efficiency or metering intervention, forecast population growth, change in household size, changes in property numbers and the impact of climate change on customers' behaviour. Leakage in the baseline forecast should remain static from the start of the plan to the end of the planning period.
Business plan	Business Plans are produced by the water companies for Ofwat and set out the investment programme for the water industry. These plans are drawn up through consultation with the Environment Agency and other bodies to cover a five-year period. Ofwat accept the Business Plan following detailed scrutiny and review.
Capital expenditure (Capex)	Spending on capital equipment. This includes spending on machinery, equipment and buildings. Capital expenditure is also termed investment.
Central market operating system (CMOS)	This is the computer system that manages all the electronic transactions involved in switching customers and provides usage and settlement data which is used in the billing process.
Consumption monitor	A sample of properties whose consumption is monitored in order to provide information on the consumption and behaviour of households served by the company.

Demand management	The implementation of policies or measures which serve to control or influence the consumption or waste of water (this definition can be applied at any point along the chain of supply).
Department for Environment, Food and Rural Affairs (Defra)	UK Government department with responsibility for water resources in England.
Deployable output (DO)	A measure of the available water resource during a drought year for a given level of service.
Distribution input (DI)	The amount of water entering the distribution system at the point of production.
Dry year annual average (DYAA)	The dry year annual average represents a period of low rainfall and unrestricted demand and is used as the basis of a water company's WRMP.
Dry year critical period (DYCP)	The generic term for the planning scenario which drives investment, i.e., at what point during the dry year (1 in 10 years severity of conditions) is the water supply most at risk of failing to meet planned levels of service.
Environment Agency	UK government agency whose principal aim is to protect and enhance the environment in England and Wales.
Final planning demand forecast	A demand forecast which reflects a company's preferred policy for managing demand and resources through the planning period, after taking account of all options through full economic analysis.
Mega litres per day (Ml/d)	One mega litre = one million litres (1,000 cubic metres) per day.
Meter optants	Properties in which a meter is voluntarily installed at the request of its occupants.
Micro-component analysis (MCA)	Detailed analysis of individual components of a customer's water use.
Non-households (NHH)	Properties receiving potable supplies that are not occupied as domestic premises, for example, factories, offices and commercial premises.
Normal year annual average (NYAA)	The total demand in a year with normal or average weather patterns, divided by the number of days in the year.
Operating expenditure (Opex)	Operating expenditure comprises day-to-day (planned and unplanned) routine expenses, which have no effect on the decline in service potential.
Optant metering	Customer led metering programme.
Peak demand	The highest demand that occurs, measured, either hourly, daily, weekly, monthly or yearly over a specified period of observation.
Per capita consumption (PCC)	The average annual consumption expressed in litres per person per day. Per capita consumption in an area is defined as the sum of measured household consumption and unmeasured household consumption divided by the total household population.
Per household consumption (PHC)	The average annual consumption expressed in litres per household per day. Per household consumption in an area is defined as the sum of measured household consumption and

	unmeasured household consumption divided by the total number of households.
Planning period	An agreed look ahead period for which the WRMP is prepared.
Social tariff	Tariff where the customer charge takes into account factors such as household size, medical needs, income levels or if certain state benefits are claimed.
Statement of response	A document that is produced at the end of the public consultation period for the draft WRMP. The document outlines the comments received from customers and the changes that will be made to the draft WRMP as a result of these comments.
Supply pipe losses	The sum of underground supply pipe losses and above ground supply pipe losses.
Target headroom	Headroom is a margin of safety which serves as a buffer between supply and demand. Target headroom is the threshold of minimum acceptable headroom which would trigger the need for water management options to either increase water available for use or decrease demand.
Underground supply pipe losses	Losses between the point of delivery and the point of consumption.
Void property	A property connected to the distribution network but not charged because it has no occupants.
Water available for use (WAFU)	Deployable output – less any sustainability reductions – plus any bulk supply imports – less any bulk supply exports – less any reductions made for outage allowance.
Water resource zone (WRZ)	The largest possible zone in which all resources including external transfers can be shared, and hence the zone in which all customers experience the same risk of supply failure from a resource shortfall.
Water resources management plan (WRMP)	A water company's plan for supplying water to meet demand over a 25-year period.
Water resource planning guidelines (WRPG)	Guidance produced by the Environment Agency for developing water resource plans.

Contents

Figures

Tables

1 Introduction

1.1 Background

Water companies in England and Wales are required to develop Water Resource Managements Plans (WRMPs) under the Water Industry Act 1991. This is now a statutory requirement. These plans describe how they will ensure that they will have sufficient resources to meet demand under different climate conditions over a minimum of 25 years. WRMPs cover the supply and demand aspects of water resources planning. The plans are updated every 5 years.

Demand is divided into different parts, as outlined in section 6 of the Water Resources Planning Guidance (WRPG):

- Household demand
- Non-household demand
- Leakage
- Minor components (e.g., water taken unbilled, water taken illegally).

Forecasting future demand for water is a key part of the process and demand from the non-household sector is a significant component of demand. Robust assessment of future demand is a pre-requisite for developing credible and resilient plans.

There is now an additional national (for England and Wales) and regional water resources planning context to the company-level WRMPs, which is being implemented for the first time in the planning round for WRMPs to be issued in 2024 (WRMP24). This has been driven by the need to improve resilience and environmental protection, to ensure resources are shared effectively between companies, and to understand and reduce water resource planning risks at the national level.

The Environment Agency are developing the National Water Resources Framework to assess water needs across sectors (not just public water supplies delivered by water companies, but also the water abstracted from the environment by agriculture, industry, etc).

There will also be a comprehensive focus on regional planning in England for the first time. Previously, this had been done on a limited basis, mainly by Water Resources in the South East (due to the fragmented nature of water supply areas in that region) and Water Resources East (due to the large role of non-PWS demand, mainly from agriculture and power) in that region. These two groups have now been joined by three others; therefore, the five regions are now:

1. **Water Resources in the South East (WRSE):**
   Portsmouth Water, SES Water, South East Water, Affinity Water, Thames Water, Southern Water.
2. **Water Resources East (WRE):**
   Anglian Water, Cambridge Water, Essex and Suffolk Water.
3. **Water Resources West (WRW)**:

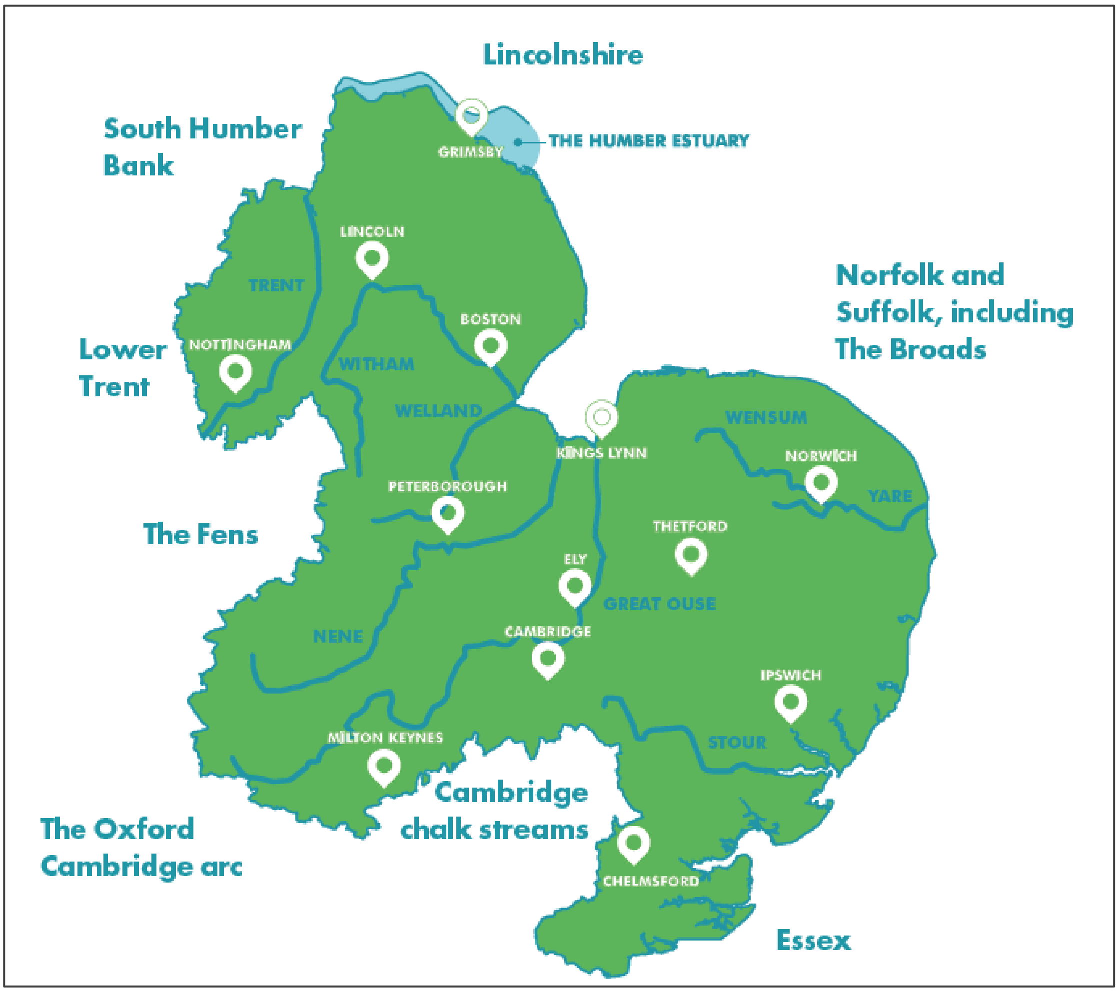
United Utilities, Severn Trent Water, Hafren Dyfrdwy, Cambridge Water, some parts of Dŵr Cymru Welsh Water[1].

4. **West Country Water Resources (WCWR):**
   Wessex Water, Bristol Water, South West Water.
5. **Water Resources North (WRN):**
   Yorkshire Water, Northumbrian Water.

WRW is therefore one of the five regional groups looking to provide strategic oversight and co-ordination of water resources within the context of the new National Water Resources Framework[2]. The aim of the regional groups is to build resilience to drought and other pressures in a cost-effective way, taking account of regional and inter-regional solutions.

The WRW region is presented in Figure 1. This report describes the initial development of the demand forecasts for non-households in the Cambridge Water supply area, which is one of the five principal water companies in WRW.

Figure 1 Water Resources East Region



---

[1] There is no regional plan to cover Wales.

[2] Meeting our future water needs: a national framework for water resources. Environment Agency. 2020.

Note that there are other water companies included in the WRW region who have not been included in this project, so this project cannot provide a regional WRW-level assessment of non-household demand.

The company-level forecasts presented in this report will feed into a regional sustainable regional resilience plan which will in turn inform the Water Resource Management Plans of each member water company within the WRW region. It will set out the schemes, investments and other actions which companies and other stakeholders will need to take to deliver our shared objective. It will also link with the other regional plans across England to form the national picture for water resources management.

1.2 Regulatory requirements

The Environment Agency sets out its expectations and guidance for non-household demand forecasts in the Water Resources Planning Guideline (currently draft).

The latest draft guideline states that water companies should produce an estimate of demand for water in the base year and produce a forecast of their non-household demand over the planning period. The planning period is a minimum of 25 years.

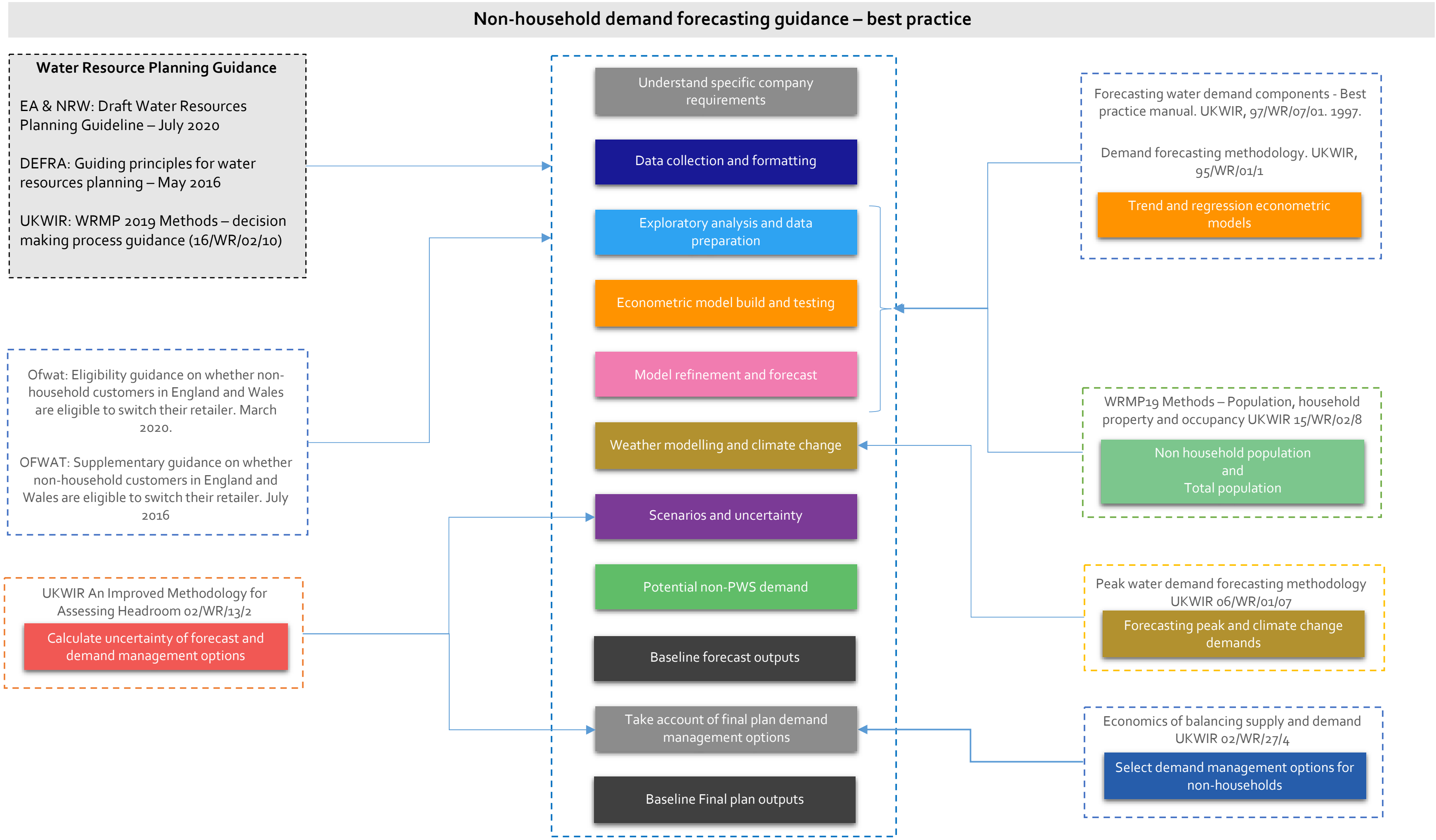
The broad needs of the regulators are:

- A plan that contains an estimated demand forecast for non-households.
- To work with retailers and through regional groups (where applicable) to share information, data and expertise to ensure the forecasts and solutions are robust.
- A description of how figures and assumptions in the forecast have been derived.
- The plan makes use of the Market Operator Services Ltd (MOSL) system that stores retail company data as needed.
- The plan describes the makeup of non-household demand in different sectors either by using the service and non-service split (identifying the main sectors), or by using Standard Industrial Classification (SIC) categories published by the Office for National Statistics.
- We explain the existing water efficiency initiatives planned by both the wholesaler and retailer. The baseline forecast should reflect non-household demand without any further intervention.
- The final plan should include any forecast savings from water efficiency programmes.
- Consideration of non-household water efficiency as an option to manage the supply-demand balance.
- To consider any uncertainty associated with reducing demand and show how you will monitor the water efficiency programme and how the plan can be adapted if required.
- That the plan considers the potential demand for other sources such as: agriculture and those on private water supply in a significant drought.

1.3 Best practice for developing non-household demand forecasts

There are a series of best practice documents in addition to the regulatory requirements, and an overview of these is presented in Figure 2.

Figure 2 Non-household demand forecasting best practice overview



1.4 Your requirements for the non-household demand forecast

Your specification for this initial non-household demand forecast is presented below:

- Non-household potable water consumption (presented as an annual average in megalitres per day) for Cambridge Water for each Environment Agency Standard Industrial Classification (SIC) category. The forecast should be based on an econometric model of non-household consumption, which relates changes in consumption to projected changes in industrial growth taking account of the ongoing efficiency trend in water usage. Variations in trends for different types of customer shall be explored as appropriate. It is anticipated that the forecasting approach will draw on the latest statistical techniques in light of the WRMP guidelines being over 20 years old.

- Uncertainty analysis and confidence intervals around a central forecast – using monte carlo simulation of input variables. Gives an upper and lower bound to forecasts to standard statistical confidence intervals.

- The impact of climate change on the forecast separated out as a volume per annum.

- Up to 10 scenarios to be investigated to explore the impact of different assumptions concerning future economic growth rates, future tariffs, commercial issues and legislative. This would include plausible economic scenarios for the WRE Region, including different growth/decline assumptions, with associated uncertainty, the impact of Brexit, and the impact of current global economic challenges.

The forecast of non-household consumption will take account of the Forecasting Water Demand Components: Best Practice Manual (UKWIR/Environment Agency, 1997) and with the most recent Water Resources Planning Guideline.

- A forecasting modelling tool that shows forecast of water consumption by sector and projections of economic assumptions driving the central forecast and all scenarios.

- The forecasts shall cover the period to 2100 and will be prepared for each Environment Agency Sector.

2 Methodology

This section provides additional details on the methodology we implemented to meet the requirements detailed in section 1.4.

2.1 Data collection and formatting

A consistent data requirement specification was provided to each of the companies involved in this project.

Ref	General data requirements	Data type
1	Data transfer preferences (e.g. email, SharePoint, DropBox, etc.)	Information
2	Key data contact	Information
3	Forecast granularity	Information
4	Number of areas	Number
5	Base year	Year
6	Population (total) forecasts by WRZ (from Base Year)	Population
7	Non-HH property forecasts by WRZ (from Base year) - Split measured and unmeasured	Property
8	Historic annual return: non-HH property numbers split by measured and unmeasured by WRZ	Property
9	Historic annual return: total population numbers by WRZ	Population
10	Pre 2017 annual non-HH consumption data (per property or per segment or industry code)	Consumption
11	2017 to 2020 annual non-HH consumption data (per property or per segment or industry code)	Consumption
12	Data to link non-HH consumption to industry code (SIC, ABP or Land Registry)	Data link
13	Data to link non-HH consumption to WRZ	Data link
14	Weather data for each WRZ: Monthly (or finer) mean temperature and mean rainfall	Weather
15	GVA and employment data by WRZ and industry segment (historic and forecast)	Economic Activity
16	Historic annual return consumption data up to and including base year	Consumption
17	Base year consumption data for each property linked to WRZ and Segment (may be included in Ref. 11)	Consumption
18	Climate change scenario predictions for temperature and rainfall	Climate
19	Scenario trend data	Trend
20	Non-PWS demand predictions	non-PWS
21	WRMP19 non-household consumption forecast outputs	Information

Cambridge Water provided data against these requirements. Any gaps in the data were discussed with relevant persons. Additional data was collected where possible if gaps were identified. In some cases, full data was not available, and in these cases amendments to the process were agreed.

2.2 Exploratory analysis and data preparation

The outputs from the exploratory analysis and data preparation, were a set of consistent data frames. These consisted of:

- Segmented consumption.
- Explanatory variables.
- Annual return data.

2.2.1 *Consumption data*

Data granularity

Consumption data was provided by Cambridge Water at a property level.

Large users

We did not have access to specific large user data, so determined a consumption threshold value above which we could classify users as a large user. We determined that this threshold should be set at 2%, i.e., if a single user consumes greater than 2% of the non-household consumption then we would flag this property as a large user.

Data checks

Data quality checks were performed, looking at the following:

- Proportion of properties that were unclassified or unmatched to a SIC group, split by year.
- Percentage of reported (annual return) volume that is contained within the either classified or unclassified consumption data.

2.2.2 *Population data*

Population forecast data and annual return by year are imported and combined to create a joint population dataset. Populations for overlapping years (2019-20) for both historical and forecast data are compared to check data accuracy.

The populations used for the baseline scenarios are presented in section 2.4

2.2.3 *Industry sector mapping*

SIC classifications are mapped to industry grouping using various mapping files, we developed mapping files for SIC_1980, SIC_1992, SIC_2003, SIC_2007. These were then used to group the properties' consumption into the industrial sectors shown in Table 1.

Table 1 Industry groupings

Industry grouping	SIC_2007 sections	Reference
Agriculture (and other weather dependent industries)	A	1
Non-service industries (excluding Agriculture)	B, C, D, E, F	2
Service industries – population driven	O, P, Q, R, S, T	3
Service industries – economy driven	G, H, I, J, K, L, M, N	4
Unclassified		5

These sectors were chosen due to data limitations and project timescales. They represent a compromise between the previous approach of splitting non-household customers into service and non-service categories, and a more ambitious approach of allowing the exploratory analysis to define the industry groupings based on their usage characteristics.

Table 2 shows the proportion of properties and the proportion of consumption for each company that falls into each of the industry groupings identified in Table 1.

Table 2 Proportion of properties and consumption in each industry group (2019-20)

Company	Industry grouping	Proportion of properties in group	Proportion of consumption in group
Cambridge Water	Agriculture	17.90%	14.30%
	Non-service	6.87%	6.01%
	Service – population	21.90%	24.90%
	Service - economy	43.20%	43.90%
	Unclassified	10.00%	10.90%

2.2.4 *Weather data*

Compiled weather data is loaded with average rainfall and average maximum temperature by year.

2.2.5 *Econometric data*

Econometric data was provided by Oxford economics (OE). This data is formatted into employment and GVA by SIC group and region. Historic data was provided from 1991, and forecast data was provided to 2040. Cambridge is covered by the statistical area of East England.

2.2.6 *Data collation*

A maximal theoretical dataset was created using all combinations of year (from OE, weather, consumption, and population datasets), SIC/industry groups (consumption), with all variables joined to these where available.

This is then aggregated to industry grouping level, with group-specific numerical variables summed (consumption, employment, GVA) and other numerical variables re-joined at aggregated level (weather and population).

Both the SIC and industry grouping aggregation datasets are output for use in subsequent modules.

2.3 Model build, testing and refinement for baseline forecasts

2.3.1 *Non-household forecast modelling*

The non-household forecast modelling is carried out in line with best practice[3].

Choosing the right modelling process is a complex task that needs to take into consideration statistical model performances, but also many other variables that require the modeller expert judgement (availability of variables, reliability of data, overfitting problems, and more). Therefore, the modelling process is based on offering all the statistical tools to the modeller, who then takes a decision based on all considered aspects.

The NHH forecast modelling process is divided in the following steps:

1. Build the MLR model based on past aggregated consumption data, considering Oxford Economic variables and potentially other factors.

2. Calibrate the model for the base year, in this case 2019-20 using the Annual Return (AR) consumption.

3. Apply the MLR model and the calibration to future explanatory variables to estimate future NHH consumption.

The MLR modelling and calibration is done at company level but considers industry groups independently. At each stage adjustments and improvements can be made specifically for each company, depending on the specifics of the data. Therefore, in Appendix A there is a complete modelling report which identifies all the specific modelling details.

---

[3] Forecasting water demand components - Best practice manual. UKWIR, 97/WR/07/01. 1997.

2.3.2 *MLR modelling*

Multi linear regression (MLR) modelling aims at finding a linear relationship between the observed consumption and explanatory variables. In a first moment, all available explanatory variables are considered. Subsequently, the model is refined choosing only the significant variables. The choice is based on:

- model performances excluding the variables one by one.
- interaction between variables
- logical inclusions/exclusions based on the relationship between the expected effect of each variable on consumption, and the estimated coefficients.
- exclusion of outliers
- other modellers' considerations.

Company specific results for each MLR model for each industry sector are included in Appendix A and include the following:

- model term
- estimate
- standard error
- p value.

2.3.3 *Calibration*

The MLR model is based on property-based data, which may not represent the total of the NHH Measured consumption. For this reason, the results of the model need to be calibrated against the Annual Report data for the base year, in this case 2019-20. This also helps accounting for differences between WRZ, not accounted for building the model at company level. Then the total measured consumption is calibrated against AR data at WRZ, deriving a multiplicative factor (factor2). No calibration at industry group level (factor1) could be performed due to the lack of property-level data for the base year.

The calibration factor used for Cambridge for each industry sector is 2.88 (see appendix A).

2.3.4 *Baseline forecasts*

Final NHH baseline forecasts are obtained separately for the measured and the unmeasured component.

For the measured component, NHH is forecast with the following steps:

- apply the MLR model separately for each industry group.

- apply the calibration.

- forecasts are then extended as an extrapolation of the trend from 2040-41 to 2099-00. This is because the econometric forecasts, upon which the forecasts are based, only extend to 2040, and to infer any trends after this point would be over-optimistic.

- minimum consumption is set to 10% of the observed years' average, with exclusion of 2020-21 that is allowed to go to zero considering the COVID crisis.

Given its uncertainty, the unmeasured sector is forecasted as constantly equal to the base year value.

The results are presented in Section 3

2.4 Scenarios and uncertainty

2.4.1 *Introduction*

The concepts of uncertainty and scenarios are often used interchangeably and partially overlap in terms of meaning. Both represent unknowns that may affect water consumption forecasts. For the purpose of the WRMP24 non-household demand forecasts we need to separate the concepts through definitions:

- **Uncertainty** refers primarily to the variability we have in forecasts due to data uncertainty and unexplainable variability uncertainty. Uncertainty is non-zero even in the present figures and grows with time in a gradual way, due to uncertainty propagation. Uncertainty can be described by probability distributions and derived statistics, like mean, standard deviation, or quantiles.
- **Scenarios** refer to the variability in future projections due to foreseeable (at least in terms of happening) events. Scenarios' variability is only applicable to future figures, not to the present, and can grow or decrease in time according to the specific events we are considering. Scenarios are usually represented by a discrete number of alternative forecasts.

As the WRMP24 non-household (NHH) forecasts are derived through a complex process, the sources of uncertainty can be many and very little is known about the quantification of uncertainty. Similarly, the number of factors that can affect NHH water consumption can be large and unexpected events and technologies may alter the way we will consume water; therefore, it is very difficult to consider all plausible scenarios.

In this work, we introduce some approximations to overcome the unknown quantification and the technical limitations involved in modelling both the uncertainty and the scenarios. We first proceed in delineating a large number of foreseeable scenarios, from which we derive plausible central, lower and upper thresholds. Then we proceed in applying uncertainty estimations for quantifiable factors on the three selected thresholds.

Details on the scenarios' definition and the uncertainty quantification are reported in following sections.

2.4.2 *Scenario development*

There are multiple and complex links between non-household demand and a wide range of factors, from international and national macroeconomic trends to local investment strategies and population growth. This complexity could present challenges for forecasting; in terms of what factors to consider and the range of scenarios needed to capture a suitable range of futures. To account for this, we have used three scenarios which are consistent with those developed and used by other companies in the WRE region, central, low and high.

These are described further in the following sections.

Population scenarios

Population scenarios are chosen from the Edge Analytics forecasts. The scenario used are presented in Table 3

Table 3.Summary of population scenarios used

Scenario version	CAM
Baseline pop scenario	Housing-plan-p
High pop scenario	Oxcam-2b-r-h
Central pop scenario	Baseline
Low pop scenario	Ons-18-low-l

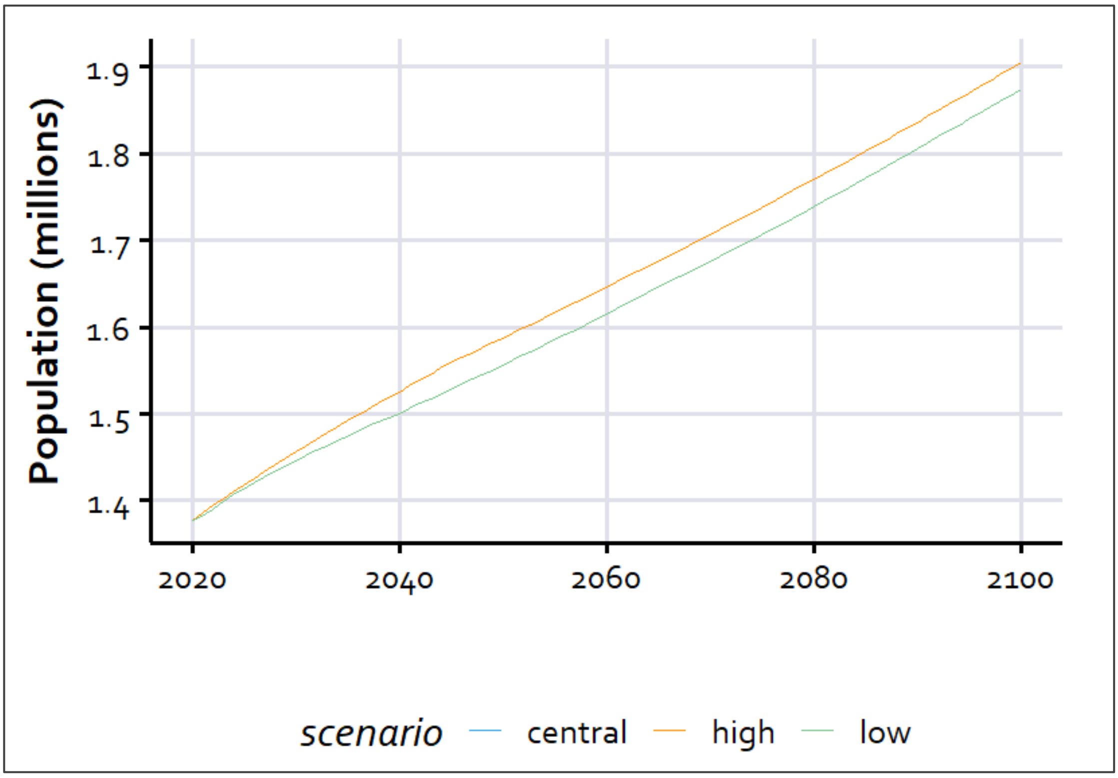
This meant that for Cambridge, the population scenarios map to the scenarios identified in this project are as presented in Table 4.

Table 4.Scenario to population scenario mapping for Cambridge Water

Scenario type	Scenario name
central	Housing-plan-p
low	ONS-18-low-l
high	Oxcam-2b-r-h
median	Oxcam-1b-r-p

The choice of population scenario is illustrated in Figure 3.

Figure 3 Basis for choice of population scenarios Cambridge



Brexit

The United Kingdom has left the European Union and on 24 December 2020, the two parties agreed a trade deal for goods and movement between the UK and EU27. The impact of this deal on the economy and immigration remains unknown at present, for both the short and the long term. However, the short-term forecasts consider both Brexit and Covid-19 impacts on the economy, and these two factors are difficult to separate. So, we decided to apply only the long-term impacts for the Brexit scenarios, as the short-term effects are already represented in the three COVID-19 scenarios.

NHH water consumption is modelled considering GVA, employment and population among other factors, and these factors are the ones impacted by Brexit.

The impact on population is estimated from Lomax, 2019[4], considering the percentage variation between the three reported Brexit scenarios: EU-membership, soft Brexit and hard Brexit. Considering our baseline as the middle scenario, we can consider a change in population of +2.6% by 2040 under the upper Brexit scenario, and a decrease of -2.6% under the lower Brexit scenario.

For employment estimates, we considered the HM Government report *HM Treasury analysis: the long-term economic impact of EU membership and the alternatives*[5], which states that

---

[4] Lomax, N., Wohland, P., Rees, P. & Norman, P. The impacts of international migration on the UK's ethnic populations. J. Ethn. Migr. Stud. 46, 177–199 (2019).

[5] HM Government. *HM Treasury analysis: the long-term economic impact of EU membership and the alternatives*, 2016, Cm 9250, Web ISBN 9781474130905

"unemployment would reach 7% to 8% in 2020, compared with a projected rate of 5% if the UK remained in the EU". Assuming our estimates correspond to the central, we can consider a variability around 3%, so +/- 1.5% for the upper and lower scenarios. Not having further temporal information, we keep this steady in time.

In terms of GVA (proportional to GDP if fixed taxation is assumed), the report proposes wider ranges, going between 1.2% and 2.8%, considering the uncertainty. For consistency we consider 1.5% like for the employment estimates. The summary of Brexit impacts is presented in Table 5.

Table 5 Brexit scenarios and their impact

	Population	GVA	Employment
Upper Brexit scenario	+2.6% by 2040	+1.5% fixed	+1.5% fixed
Central Brexit Scenario	baseline	baseline	baseline
Lower Brexit Scenario	-2.6% by 2040	-1.5% fixed	-1.5% fixed

COVID-19

COVID-19 has had a strong negative impact on the economy and on NHH water consumption, due to lockdown measurements and economic recession, as well as due to remote-working measurements. At the time of writing this report, vaccines are beginning to be rolled out, and the impact of the pandemic is expected to gradually reduce in the second half of 2021. The impact of COVID-19 is modelled in three different ways:

1. GVA and Employment are modified on the short term, according to the expected impact on the economy.
2. Water consumption is reduced across all sectors.
3. Water consumption is shifted between sectors.

COVID-19 impact on GVA and Employment

The impact of COVID-19 on GVA and Employment is estimated from the *Forecasts for the UK economy 2020* by the HM Treasury[6]. The report compares independent forecasts. The baseline was estimated using the Oxford Economic (OE) forecasts for GVA and Employment. From the report the upper and the lower thresholds are estimated for GVA (derived from GDP, Table M1 of the report, with the assumption of proportionality) and for employment (derived from unemployment forecasts, table M5 of the report), using the upper and the lower independent estimate. For GVA, OE is a central forecast, therefore is used as the central scenario, while for employment OE is already the upper forecast, so it is used as the

---

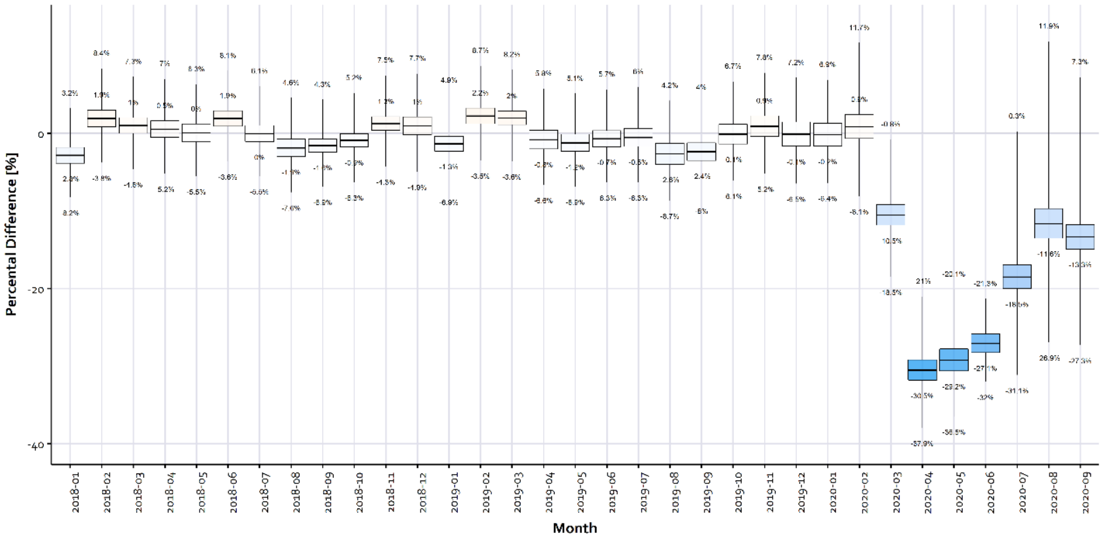
[6] HM Treasury, Forecasts for the UK economy: a comparison of independent forecasts, 2020, No. 397, ISBN 978-1-913635-61-9

upper scenario. The result is a set of percentage changes to apply to the baseline for years 2019-2024. These estimates also include the short-term impact of Brexit.

NHH water consumption reduction due to COVID-19

Beyond the effects on the economy, COVID-19 has an effect on water consumed by businesses and non-household properties due to different operations and remote working. Artesia has conducted an independent study on the impact of COVID-19 on the NHH sector. Figure 4 shows the reduction in water consumption during summer 2020, compared to the previous year, considering weather, holidays, and other influencing factors.

Figure 4 Reduction in NHH water consumption during summer 2020 months compared to previous months.



The three scenarios are considered as follows:

- Upper COVID-19 scenario: no variation on the baseline.
- Central COVID-19 scenario: -12% in 2020-21 and -6% in 2021-22, then baseline.
- Lower COVID-19 scenario: -20% in 2020-21 and -10% in 2021-22, then -3% on the baseline.

Shift between sectors due to COVID-19

The COVID-19 impact on water consumption is due to its impact on the economy and the change of operations due to a mass remote-working approach. However, both these factors, quantified above as a total effect, affect differently the different economic sectors. Therefore, a final step of the modelling is to shift water consumption across sectors.

To do so, we use data from the ONS Business Impact of COVID-19 Survey (BICS) from September 2020[7] (assumed to be the best representation to date to the post-lockdown COVID-19 scenario). The dataset reports both the changes in turnover and the percentage of

---

[7] ONS, BICS Wave 14 edition of this dataset 7 September to 20 September 2020.

workers working remotely, by sector. Combining the two factors we could derive that under the September 2020 conditions, NHH water consumption is likely to have shifted:

- Agriculture +0.4%
- Non-service +9.1%
- Service-economy -4.1%
- Service-population -5.8%
- Unclassified +0.4%

The shift is only considered in the lower COVID-19 scenario, where long term impact of remote-working is considered.

Summary of COVID-19 scenarios

Table 6 lists the summary of the COVID-19 scenarios and their impact.

Table 6 COVID-19 scenarios and their impact

	GVA	Employment	Consumption reduction	Sector shift
Upper COVID-19 scenario	Upper independent forecast	OE forecast	baseline	baseline
Central COVID-19 Scenario	OE forecast	Central independent forecast	-12% in 2020-21 -6% in 2021-22 then baseline	baseline
Lower COVID-19 Scenario	Lower independent forecast	Lower independent forecast	-20% in 2020-21 -10% in 2021-22 then -3%	Agric: +0.4% Non-serv: +9.1% Serv-eco: -4.1% Serv-pop: -5.8% Unclass: +0.4%

Climate change

Modelling residuals

Building the residual models for each company independently is correct theoretically, but due to the low number of points in time it can result in unstable models. Therefore, we used a generalised model we developed for another region, which contained more and more stable data points. To make the residuals comparable, we standardised them, dividing them by the consumption itself:

$$residuals = \frac{(consumption - prediction)}{consumption}$$

Using this method, the resulting model predicts standardised residuals in the future as a function of weather variables (*average rainfall* and *average maximum temperature*). The residuals can then be adapted to each company by multiplying them by the mean consumption of past years.

Modelling historic weather trends

The first step in the analysis is to establish the change in weather patterns that are occurring due to climate change. The weather variables under examination are *average maximum temperature* and *average rainfall*. Figure 5 and Figure 6 show that the trends of these variables over the years can be well represented with linear regressive models.

Figure 5: A plot showing the trend of peak daily temperatures for Cambridge

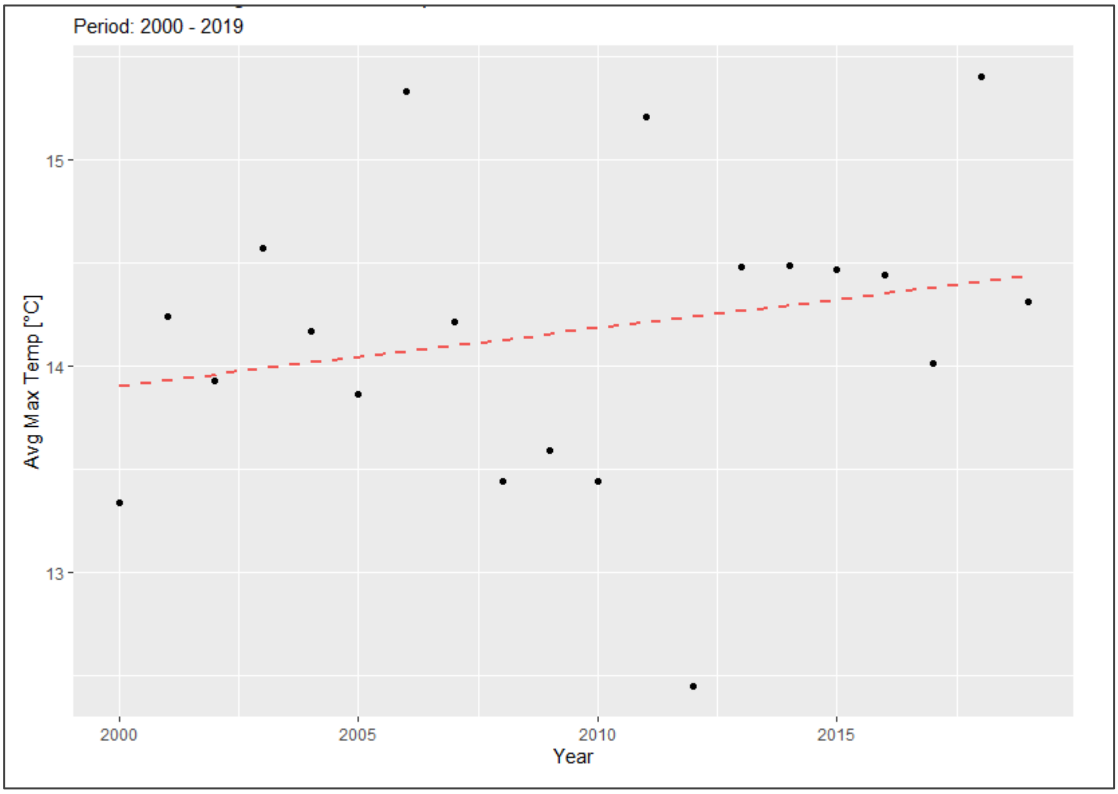
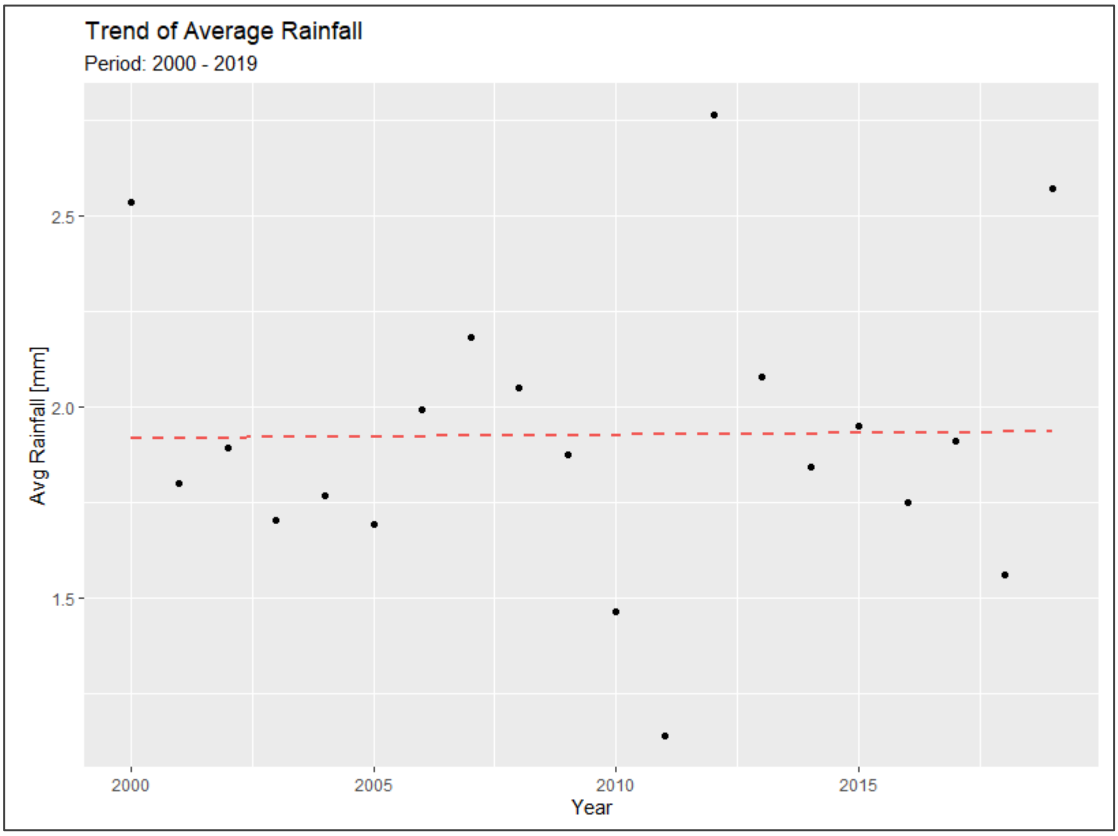


Figure 6: A plot showing the trend of average daily rainfall for Cambridge



Forecasting Weather and Climate Change Residuals

The weather models developed in Figure 5 and Figure 6 are used to forecast *average maximum temperature* and *average rainfall* through the forecast period.

We used additive climate change models in conjunction with the weather forecasts. These models provide 12 scenarios of potential temperature and rainfall patterns.

The forecasts of the weather variables are each summed with the 12 relevant climate change scenarios to produce 12 forecasts for *average maximum temperature* and *average rainfall*. The 12 scenarios for each are then fed into the residual model to obtain residual forecasts.

However, all 12 scenarios are not required for this analysis, only a *low*, *central*, and *high* scenario. To extract three scenarios from the 12, the 10th, 50th, and 90th quantile of the scenarios are taken for each financial year.

The climate change scenarios only go up to the year 2080, whereas we need the forecasts up to the year 2100. The forecasts must therefore be extended to meet client needs. To perform this extension, a linear regressive model is fit to each of the *low*, *central*, and *high* scenarios and used to predict the final 20 years to the desired end year, 2100.

Water efficiency

The evolution of technology and regulations is expected to contribute reducing NHH water consumption, by improving water efficiency.

The three water efficiency scenarios below were selected in consultation with the WRSE and WRW steering groups, and have been used in this instance for Cambridge Water:

- Upper water efficiency scenario: water consumption is reduced by 2% by 2050-51.
- Central water efficiency scenario: water consumption is reduced by 7.5% by 2050-51.
- Lower water efficiency scenario: water consumption is reduced by 16% by 2050-51.

2.4.3 *Modelling uncertainty*

Every single element of the complex WRMP24 NHH forecasts is affected by a certain degree of uncertainty, but the quantification is difficult. Therefore, we decided to focus on the elements the have the biggest impact on the forecasts:

- the explanatory variables used in the model.
- the model
- climate change
- MOSL reporting.

The quantification of uncertainty for each component is described in the following sections.

Explanatory variable uncertainty

Each explanatory variable is affected by a different degree of uncertainty. It is not easy to separate the uncertainties and to evaluate the effects of each on the resulting water consumption. However, thanks to the linear nature of the model, if we consider the explanatory variables to have the same uncertainty, e.g., ±10%, we can derive that the same uncertainty will affect water consumption. The following explanatory variables are considered for uncertainty:

- GVA.
- Employment.
- Population.

Other minor explanatory variables are expected to have a lower uncertainty and to affect the water consumption estimations to a smaller degree.

Observing the population scenarios from Edge Analytics, we can observe that their uncertainty is very small in the present and grows steadily in the future, reaching a value of ±6% to ±12% depending on what scenarios we consider.

In terms of GVA and employment as we can observe in the *Forecasts for the UK economy 2020* by the HM Treasury, the larger uncertainty is actually in the short term and varies between ±30% to ±50% for GVA to ±1.5% to ±3% for Employment.

Considering the uncertainties estimated above, the general uncertainty for the explanatory variables is estimated as:

- ±8% of the water consumption in 2019-20.
- Growing to ±12% of the water consumption in 2025-26.
- Growing to ±18% of the water consumption in 2099-00.

Model uncertainty

Model uncertainty is estimated separately for the considered industry groups and companies, as different models are used. A model's R^2 value represents the variability in the data that the model is able to explain. We estimate the model uncertainty as $1 - R^2$, i.e., the variability in the data that the model is not able to explain. This is a simplification, as effects like overfitting can increase the R^2 value beyond what the real capabilities of the model are, but overall, it is a good proxy for the model uncertainty.

Climate change uncertainty

Climate change uncertainty has been estimated from the UKCP18 Climate Change Over Land infographic, that estimates the following:

- Rainfall is expected to show a variability up to ±25-30% in summer and ±12-19% in winter by 2060-79. It can be approximated as a ±20% on a yearly basis by 2060-79.
- Temperature is expected to show a total variability between 2.5-3.5 °C in winter and 3.3-4.7 °C in the summer, so about 4 °C on a yearly basis by 2060-79. Assuming an average yearly temperature around 15°C, that is about 15°C±2°C, i.e., ±13%. by 2060-79.

Combining the two estimates, we can consider a climate variability of about 16% by 2070, so we assume 18% by 2099.

MOSL

The liberalisation of the water market for the commercial sector has had an impact on the water consumption reporting, operated by MOSL, the market operator for the water retail market in England. During this time MOSL has failed to deliver some of its targets for improving data quality (notably in the "Long term unread meter category" and the "level of properties flagged as vacant" areas)[8]. The MOSL annual market performance report identifies that 1 in 6 premises is now flagged as vacant, and meters unread for more than a year have increased from 7% at 2017 to 15% at March 2019, with one-third of these not being read since market opening.

The effects are observable as the difference between the reporting before 2016 and after 2016, which the modelling could take into account as a flag, set to zero before 2016 and set to 1 after. However, there are commitments from MOSL to improve in this area and signs in 2019 that progress is being made. We are unsure how these improvements will impact reporting in the future, depending on how the water retail market evolves. Therefore, the uncertainty on the MOSL reporting flag is estimated as growing to 30% by 2030, and then remain 30% after.

2.4.4 Application of uncertainty

Once the uncertainty of the single components is defined as in the previous sections, they are then combined in a quadratic way:

---

[8] Annual Market Performance Report 2019/20. MOSL.

$$u = \sqrt{u_{EV}^2 + u_{model}^2 + u_{climate}^2}$$

The resulting uncertainty, estimated for each Company, industry group and year, is applied on the three derived scenario thresholds.

3 Results

3.1 Summary of results provided

Baseline forecast outputs are provided in the file "CAM_baseline_20201209". This file includes the following breakdown of baseline non-household consumption forecasts from 2019-2020 through to 2099-2100:

- 01: Forecasts of measured non household demand for each industry sector.
- 02 Forecasts of unmeasured non-households at company level
- 03 Forecasts of measured non-households at company level
- 04 Company total forecasts (measured plus unmeasured)

Scenario forecast outputs are provided in the file "CAM_scenario_20201209". Within this folder, are subfolders using the following naming convention: Artesia-CAM_NonHousehold-Demand-Forecasts_XX_Scenario_preliminary_result_20201209.

Where the XX is replaced in each file name by the relevant scenario number (see Table 3). These files include the following breakdown of scenario non-household demand forecasts from 2019-2020 through to 2099-2100:

The results for Cambridge are presented in the following sections. Firstly, we present the company-level results for measured and unmeasured non-household consumption, for baseline and scenario. Finally, we present the MLR model metrics for each model to identify the drivers for the forecast.

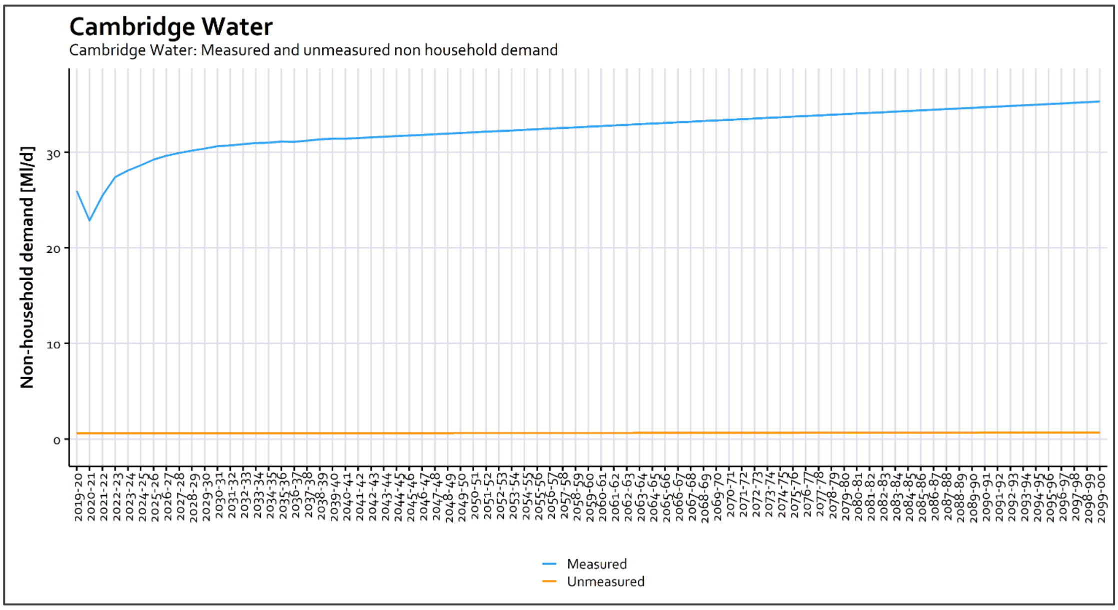
The calibration factor applied across all sectors was 2.88 (See Appendix A)

3.2 Total measured and unmeasured baseline non-household demand

At start of the planning period (2025), the Cambridge Water total non-household demand is predicted to be 29.85Ml/d within an overall range of 26.58 to 35.99 Ml/d. Much of the early uncertainty is due to the impact of COVID-19 and lack of non-household consumption data from MOSL.

By the end of the planning period the non-household demand is predicted to be 35.99 Ml/d (an increase of 5.93 Ml/d) within a range of 26.58 Ml/d to 35.99 Ml/d.

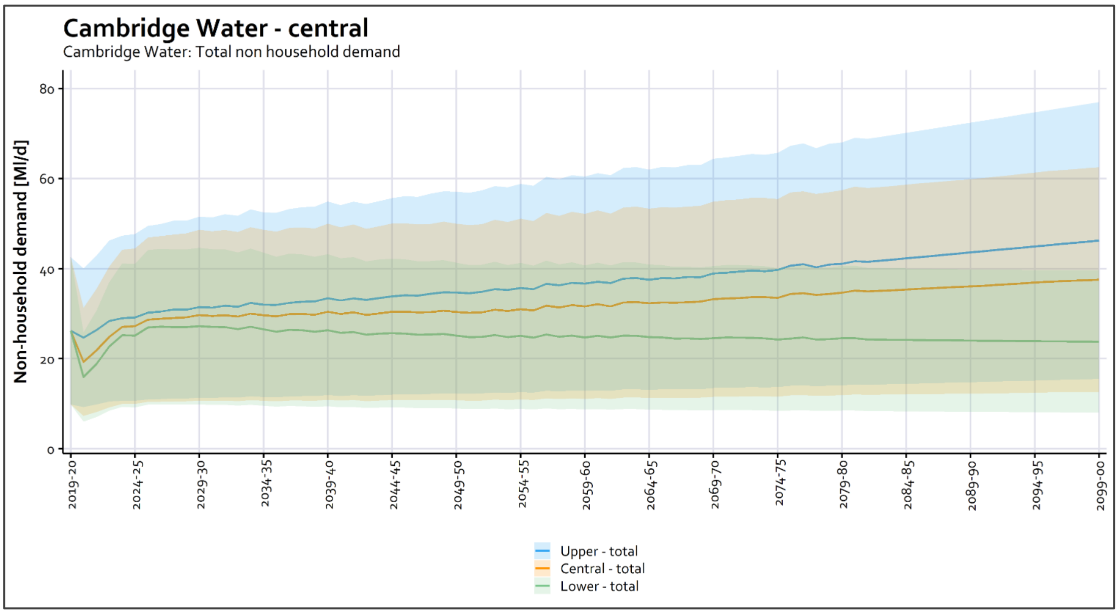
Figure 7 Cambridge measured and unmeasured non-housheold consumption



3.3 Scenarios for total baseline non-household demand

Figure 8 shows the non-household demand forecast for the Cambridge region, for the three different scenarios identified in section 2.4 of this report. It shows a wide range of uncertainty over the planning period with uncertainties around the higher and lower scenarios approaching plus or minus 25 Ml/d. A large proportion of this uncertainty is generated from the start of the forecast, out to 2029-30, driven by large uncertainties associated with COVID-19, Brexit and population uncertainty.

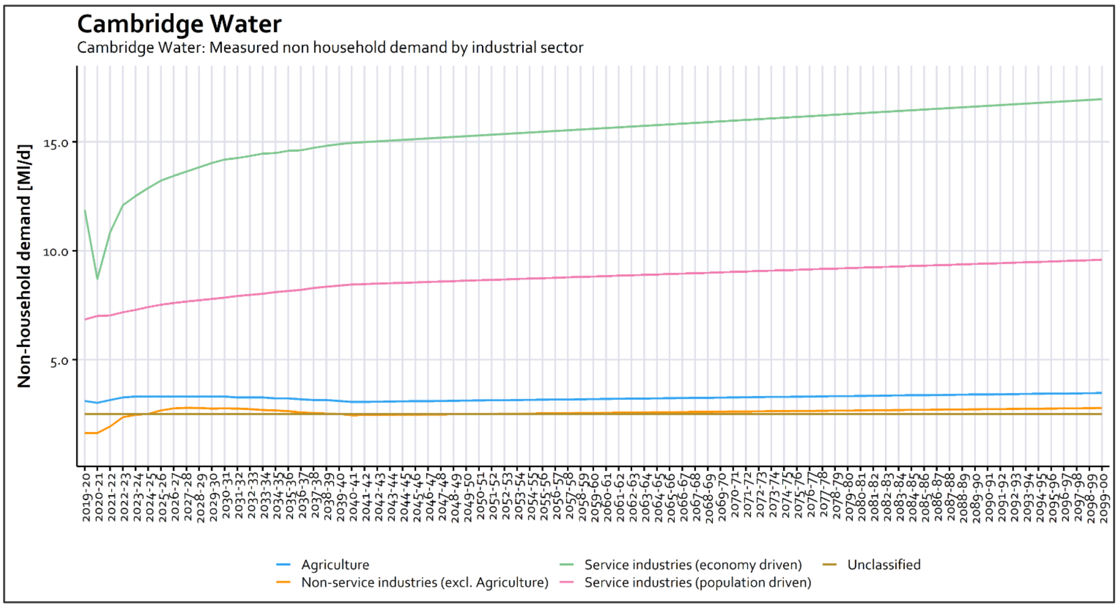
Figure 8: Scenarios for total non-household demand for the Cambridge region



3.4 Total baseline non-household demand broken down by sector

The majority of non-household demand in the Cambridge Water region is from the Service Industries (economy driven) sector. This is seen in the forecasts presented in Figure 9. The next most significant sector in the region is Service Industries (population driven). There is a steady increase in demand in these two sectors. The remaining sectors remain relatively consistent across the forecast period.

Figure 9: Total non-household demand for the Cambridge region, broken down by sector



4 Discussion of findings

This section discusses the overall results and findings that are consistent across WRW, any company specific findings are discussed in Appendix A. The discussion is broken down into the following sections:

- The modelling approach used for this study.
- Uncertainty in the predictions.
- Data issues.
- Potential alternative industry segments.
- Weather impacts.
- Other potential improvements.

4.1 Modelling approach

In the process of developing the non-household consumption models we started with the different drivers for each sector group. However, some models were weak, and therefore we allowed the data for each sector's model to guide us in the selection of the explanatory variables, i.e., all explanatory variables were applied to each sector model and the models refined through a process of variable reduction until the strongest models remained. The resulting significant variables for Cambridge Water are shown in Table 7.

Table 7 Significant explanatory variables in each sector model

Company	Sector model	Population	GVA	Employment	MOSL flag	Other
Cambridge	Agriculture		X		X	
	Non-service	X		X	X	
	Service-economy		X	X	X	
	Service-population			X	X	
	Unclassified	X			X	

What we see is quite a mix of explanatory factors, with population having a strong influence in the non-service and unclassified models. Economy (GVA) is a factor in the non-service and service-economy models. Employment is a factor in the agriculture, service-economy and service-population models. The MOSL flag (discussed in sections and 4.3) is significant in nearly all models. Note, that for the unclassified sector there was too much variation in the data to establish any robust models, therefore these were held flat across the planning period.

All the models developed and used in this report are robust, but some will have wider uncertainty bands.

4.2 Uncertainty in predictions

We have estimated unknowns that may affect water consumption forecasts through applying uncertainties and scenarios around the baseline forecasts. We have estimated uncertainties due to the data uncertainty and unexplained variability and applied these across the forecasts such that they grow in a gradual way over time. It should be noted that missing consumption data since 2016-17 results in a large uncertainty in the forecast. For unknowns in future projections, we have used scenarios to estimate the future variability. We created scenarios that include future variations in population, Brexit impacts, COVID-19 impacts, MOSL data quality, climate change and water efficiency. These are all explained in section 2.4.2. Modelling uncertainty is covered in section 2.4.2 and includes uncertainty for the MOSL data as well as climate change, the modelling variables and the model itself.

The result is some significant uncertainty around the future projections of non-household consumption. For example, the results show an uncertainty (including the scenarios) of approximately +66% and -66% by 2100.

The forecasts start from a base year of 2019-20 and some of the early uncertainty will be due to Covid-19 impacts. Within the baseline models Covid-19 causes a decline in the GVA forecast from Oxford Economics in 2020/21, then a return to normal, however we have shown in section 2.4.3 that the effect on Covid-19 on demand is likely to be larger than this and so a larger adjustment for Covid-19 is made in the scenario forecasts.

Therefore, companies may consider selecting a forecast which differs slightly from the baseline, but within the scenario ranges, depending on their own local knowledge and approach to risk.

4.3 MOSL data

The lack of MOSL data had a direct impact on the quality of the forecasts and the associated uncertainty. Even if MOSL data are obtained, it would be useful to understand better whether the changes in the annual return aggregated property level data from MOSL are a long-term change or a short-term result of data issues during retail separation.

MOSL's report for 2019-20 identifies that there are specific problems with long term un-read meters and high numbers of vacant properties (some with high consumption values)[Error! Bookmark not defined.]. Therefore, we have included three future scenarios, one where the data improves, one where it deteriorates and one where it stays as it is currently.

For future forecasts, we need to consider further whether can we make better adjustments for the effects from retail separation and consider if this is the new normal (e.g., due to the redefinition of NHHs to HHs) or is it due to erroneous data, that will eventually be resolved. It might be useful to flag to MOSL, Ofwat and the Environment Agency the significance of these data errors.

4.4 Property level consumption data

Suitable models have been developed for South Staffs Water using our established approach using historic property level data aggregated to industry groupings of; non-service, agriculture, service-population and service-economy, as well as unclassified. The models are a combination of population, employment and GVA depending which factors were able to establish valid relationships. Large users are included within the data as the removal led to very unstable groups. Having property level consumption data improves the identification of large users and voids, it improves consistency of data and allows for better quality checks on the data; all of which will improve the model results. It would be better to have a consistent smaller set of properties that are representative of an area, than try and reduce the overall size of the unclassified group.

4.5 Weather

The outputs from this modelling are raw un-normalised forecasts of non-household demand, calibrated to the reported values for AR 2019/20. This is in line with the scope of the project. The Water Resource Planning guidance does not specifically require companies to apply weather factors to non-household demand.

Companies wishing to derive and apply NY (normal year) or DY (dry year) factors to the non-household demand forecasts derived in this project should consider the relative size of the weather driven non-household demand (i.e., from the agricultural sector) in their region and individual WRZs, compared to other non-household and total demand. They should also take into account the quality of the data for deriving these factors.

4.6 Other improvements

There is clearly an impact on the forecasts from the quality of the data. There should be further work to help water companies improve the quality the data they use for forecasts. It might be more cost effective to do this as a regional group, rather than individually.

The current best practice (developed in 1997) suggests the econometric approach, and this has been applied quite consistently by individual companies over the past few WRMPs. However, some of these relationships are quite weak and there might be alternative forecasting techniques that might be better given the quality constraints on some of the data. Cambridge Water should consider whether it is worth working with others in the region, or nationally to carry out some wider industry research to evaluate alternate methods for modelling and forecasting.

TO HERE

Another option worth considering is a greater level of aggregation within the WRE region. We have seen that there are limitations in the data, and it might be possible to look at all the WRZs across the region and group together WRZs based on how their non-household consumption behaves (as opposed to a company geographical boundary). This may allow more data to be pooled, and this may result in stronger models being developed.

5 Conclusions

We have produced a set of non-household demand forecasts for Cambridge Water from 2019-2020 out to 2099-2100. These are presented for metered and unmetered properties at company level and disaggregated by industrial sector.

The approach used follows existing industry best practice. Robust multiple linear models have been produced for 4 cohorts of industrial sectors for Cambridge Water, using explanatory factors that include population, gross value-added metrics, employment rates, population density and other factors.

Since the last set of non-household forecasts were completed for WRMP19, the non-household retail sector has undergone a transformation with the introduction of retail competition. A significant impact from this is that metered non-household consumption data is now the responsibility new retailers, managed by the new Market Operator Services Ltd (MOSL). We have observed a change in data quality and consistency since the change in 2017. This has complicated the modelling (which relies on a consistent set of time series data) and has increased the uncertainty around the demand forecasts. This has been taken into account in the models, uncertainty and scenario estimates.

The first year of the forecast (2020) has seen an unprecedented change in non-household demand due to the policies introduced to combat the COVID-19 pandemic. This increases uncertainty going forward as we still do not fully understand what the enduring impacts will be from changes in working practices, such as increased working from home. Therefore, we have included the COVID-19 impact in the scenarios and uncertainty estimates.

The sector also faces a number of future unknowns in demand from non-households, such as population change, Brexit, climate change and how water efficiency will be delivered in the non-household sector. Therefore, these have also been included in the scenario and uncertainty modelling.

The overall conclusion is that non-household demand in the Cambridge Water region at the start of the planning period (2025), is predicted to be 29.85 Ml/d (as presented in the graph above) within an overall range of 26.58 to 35.99 Ml/d. Much of the early range in possible demand is due to the impact of COVID-19 and uncertainty over the quality of non-household consumption data from MOSL. The baseline demand forecast is predicted to increase by the end of the planning period (2100) to 35.99 Ml/d (an increase of 6.14 Ml/d).

We have identified a number of improvements that could be implemented for future forecasts, and these are included in the recommendations.

6 Recommendations

Cambridge Water should use the baseline and scenario forecasts presented in this report to select an initial WRMP baseline forecast for the metered and unmetered non-household demand forecast lines in the Environment Agency's water resource planning tables.

We have presented the forecasts from a base year of 2019-20. The intermediate years 2020-21 through to 2024-25 are presented for information prior to the start of the planning period in 2025-26. These intermediate years are potentially volatile with a number of unknowns around the impact of the COVID-19 pandemic and the impact from Brexit on non-household consumption. Therefore, we recommend that the baseline and scenario forecasts are updated prior to the submission of the final water resource management plans.

During the course of the work to develop the non-household demand forecasts we have identified a number of potential improvements to achieve more accurate forecasts. These are set out below.

Cambridge Water should inform MOSL of the importance of getting consistent good quality data on non-household consumption for forecasts. MOSL's report for 2019-20 identifies that there are specific problems with long term un-read meters and high numbers of vacant properties. These have caused some volatility in the consumption data since the introduction of market reform, which have impact on the robustness of future forecasts.

The ability to allocate non-households to specific industry sectors through tools such as SIC or AddressBase Premium also aids the robustness of forecasts. Cambridge Water should investigate the most efficient way of improving this information for future forecasts. Note, it is important that there is good quality and consistency of data over time for a good coverage of industry types, which in turn means that the variation that we observe is genuine and can be modelled better.

The forecast modelling in this study has been carried out using a functional programming approach that allows forecasts to be run and evaluated more efficiently. This approach allows forecasts to be produced more frequently, potentially sub-annually, as data is updated. A more continuous forecasting approach would remove the step-like transitions between AMP forecasts and could improve the robustness of the forecasts. The functional approach would also allow for different sector groupings to be applied quickly and efficiently.

Appendix A: Cambridge modelling results

Note about Cambridge NHH forecasts:

- Provided time series have been integrated with time series from WRMP19 to obtain 4 years of data.
- Property-level data was not available after 2016-17.
- The Unclassified sector has been forecasted as constant, keeping the last available data value steady.

MLR modelling

MLR modelling aims at finding a linear relationship between the observed consumption and explanatory variables. In a first moment, all available explanatory variables are considered. Subsequently, the model is refined choosing only the significant variables. The choice is based on:

- model performances excluding the variables one by one.
- interaction between variables
- logical inclusions/exclusions based on the relationship between the expected effect of each variable on consumption, and the estimated coefficients.
- exclusion of outliers
- other modellers' considerations

The result in the case of Cambridge Water NHH consumption MLR model is reported in the following tables.

MLR model summary for the industry group "agriculture"

term	estimate	std.error	p.value
(Intercept)	-1.18	2.14	0.636
GVA	0.0015	0.0013	0.3704

MLR model summary for the industry group "nonservice"

term	estimate	std.error	p.value
(Intercept)	-8.55	4.46	0.306
population	0.000016	0.0000096	0.3421
employment	0.012	0.0084	0.3834

MLR model summary for the industry group "serviceeconomy"

term	estimate	std.error	p.value
(Intercept)	-6.62	1.34	0.127
GVA	0.00016	0.000034	0.13
employment	0.01	0.0026	0.1598

MLR model summary for the industry group "servicepopulation"

term	estimate	std.error	p.value
(Intercept)	-2.32	4.47	0.6549
employment	0.0056	0.0056	0.4194

MLR model summary for the industry group "unclassified"

term	estimate	std.error	p.value
(Intercept)	-14.225	6.09	0.1445
population	0.000042	0.000017	0.1321

Calibration

The MLR model is based on property-based data, which may not represent the total of the NHH Measured consumption. For this reason, the results of the model need to be calibrated against the Annual Report data for the base year, in this case 2019-20. This also helps accounting for differences between WRZ, not accounted for building the model at company level. Then the total measured consumption is calibrated against AR data at WRZ, deriving a multiplicative factor (factor2). No calibration at industry group level (factor1) could be performed due to the lack of property-level data for the base year.

The calibration factors for Cambridge are reported below.

wrz	industry_grouping	factor2
Cambridge Water	agriculture	2.883
Cambridge Water	nonservice	2.883
Cambridge Water	serviceeconomy	2.883
Cambridge Water	servicepopulation	2.883
Cambridge Water	unclassified	2.883