artesia

Cambridge Water

WRMP24 Household consumption forecasting – Micro-component model

Project reference: 2464

Report number: AR1399

2020-12-15

www.artesia-consulting.co.uk

Report title:	WRMP24 Household consumption forecasting – Micro-component model
Project reference:	2464
Report number	AR1399
Date:	2020-12-15
Client:	Cambridge Water

Version	Author(s)	Reviewed by:	Description	Date
1.0	Shana Meeus		Report template	15-10-2020
2.0	Rob Lawson, Giles Fox	Rob Lawson	Cambridge Water report	08-01-2021
3.0	Rob Lawson. Giles Fox	Rob Lawson	Final version	09-02-2021

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 have a statutory duty to develop Water Resource Management Plans (WRMPs) under the Water Industry Act 1991. Forecasting the demand for water is a key element of this plan, and household demand is, in turn a significant part of overall demand.

Companies are now working in a more extensive and co-ordinated way within the context of regional plans, which have been implemented across England in the run up to the next round of WRMPs, to be published in 2024 (WRMP24). Regional plans have been implemented to improve resilience and environmental protection, and to better understand how resources may be shared between companies.

This report sets out the initial development of household demand forecasts for Cambridge Water (CAM) for the Water Resources East regional plan. This household demand forecast has been developed within the context of regulatory requirements and technical guidance. In addition, for this round of plans, Artesia has developed an updated and improved modelling framework which sets out the detailed steps required to develop the household demand forecast.

The forecast set out in this report has been developed based on micro-component modelling methods, which model household water use based on estimates of specific water using activities within the home. This is a well-established and extensively used approach to modelling and forecasting household water demand. This method is suitable for water resource zones with a normal level of water resource planning concern.

This report describes the steps involved in producing a micro-component-based household demand forecast. A key step is to split population and property forecasts into metered segments, including unmeasured, existing measured, compulsory measured, optants and new properties. Assumptions are made about these segments in order to ensure consistency within and between the segments for key variables such as household occupancy. Calibration ensures consistency with zonal population, property and occupancy totals. These values are then rebased in an agreed way to match the base year values.

Micro-component modelling uses the most recent available data on micro-component use and occupancy to determine statistically significant relationships between these variables. A linear model has been developed for toilets, showers, baths, washing machines and taps based on this analysis. Trends are then added to the model to reflect likely technology developments, and to explore scenarios associated with these, over the planning period.

Weather modelling is then used to derive normal year, dry year, and (where needed) critical period factors. Scenarios have then been produced to reflect a range of potential variations in population, property and meter forecasts.

The results of the forecast give a 16.99 Ml/day increase in household consumption for normal year demand scenarios including the impact of climate change, over the planning period (2019/20 to 2099/00), this is an 39.55% increase for the company. This is largely driven by a 75.44% increase in the property forecast.

In contrast, total PHC decreased by 20.45% over the forecast period and PCC showing a smaller decrease of 6.9%. The reason for this disparity is due to decreasing occupancy. If occupancy is forecast to decrease, then per household consumption will be more greatly affected than PCC, as the relationship between the two variables is not linear. This reflects the 'economies of scale' inherent in the occupancy model which means that the proportional increases in consumption reduce as more people live in a property.

Acronyms

The following acronyms may be used as part of this report and have the following meanings.

Acronym	Description
AA	Annual average
ACORN	A classification of residential neighbourhoods
ALC	Active leakage control
AMP	Asset management plan
AR	Annual review
BL	Baseline
Capex	Capital expenditure
CMOS	Central market operating system
CP	Critical period
CSL	Customer side leakage
Defra	Department for Environment, Food, and Rural Affairs
DI	Distribution input
DMA	District metered area
DO	Deployable output
DYAA	Dry year annual average
DYCP	Dry year critical period
EA	Environment Agency
EBSD	Economics of balancing supply and demand
FP	Final planning
HH	Household
HHCF	Household consumption forecast
IHM	Individual household monitor
MCA	Micro-component analysis
mHH	Measured household
Ml/d	Mega litres per day
MLR	Multiple linear regression
mPCC	Measured per capita consumption
mPHC	Measured per household consumption
NHH	Non-household
NYAA	Normal year annual average
Ofwat	Water services regulation authority
ONS	Office for National Statistics
Opex	Operating expenditure
PCC	Per capita consumption

PHC	Per household consumption
PR	Price review
SAM	Small area monitor
SDB	Supply demand balance
SIC	Standard industrial classification
uHH	Unmeasured household
UKWIR	UK Water Industry Research
uPCC	Unmeasured per capita consumption
uPHC	Unmeasured per household consumption
USPL	Underground supply pipe leakage
WAFU	Water available for use
WEFF	Water efficiency saving
WRMP	Water resources management plan
WRZ	Water resource zone

Glossary

The following terms may be used as part of this report and have the following meanings.

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 at least 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 (WRMP) under the Water Industry Act 1991. 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 Guideline (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 by the household sector is the largest 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, South Staffs 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.

This report describes the initial development of the demand forecasts for households for Cambridge Water, in support of the regional forecast for Water Resources East.

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)[2].

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 household demand over the planning period. The planning period is a minimum of 25 years.

The guidance sets out the methodology water companies should follow, with reference to further relevant technical guidance.

- *UKWIR (2016) WRMP19 Methods – Household Consumption Forecasting*
- *UKWIR (2016) Population, Household Property and Occupancy Forecasting*
- *UKWIR (2006) Peak Water Demand Forecasting Methodology*

The latest draft guidance also states, "*You should also refer to other relevant reports such as the water industry project on 'Water Demand Insights from 2018 (Artesia 2020)*".

The broad needs of the regulators are:

- Clearly explain the assumptions, risks and uncertainties associated with the results.
- State why a particular method has been chosen, the assumptions made, and the uncertainty associated with the demand forecast.
- Show how uncertainty is allocated in the rest of the plan.
- Consider the impacts of prolonged dry weather and droughts and the resulting high demand where it affects the supply-demand balance.
- Consider whether there are alternative methods to define dry year demand.
- Consider the results of water industry project on 'Water Demand Insights from 2018 (Artesia 2020).
- If the plan includes a critical period of high demand, it should be informed by recent peak demand years, including 2018 and 2020. It should include weather dependent demand, seasonal population changes and other factors as appropriate.
- Clearly describe the assumptions and supporting information used to develop population, property and occupancy forecasts, and any uncertainties. Demonstrate the incorporation of local council information in England.
- Explain the methods used to forecast property figures after the planning period used by local councils.

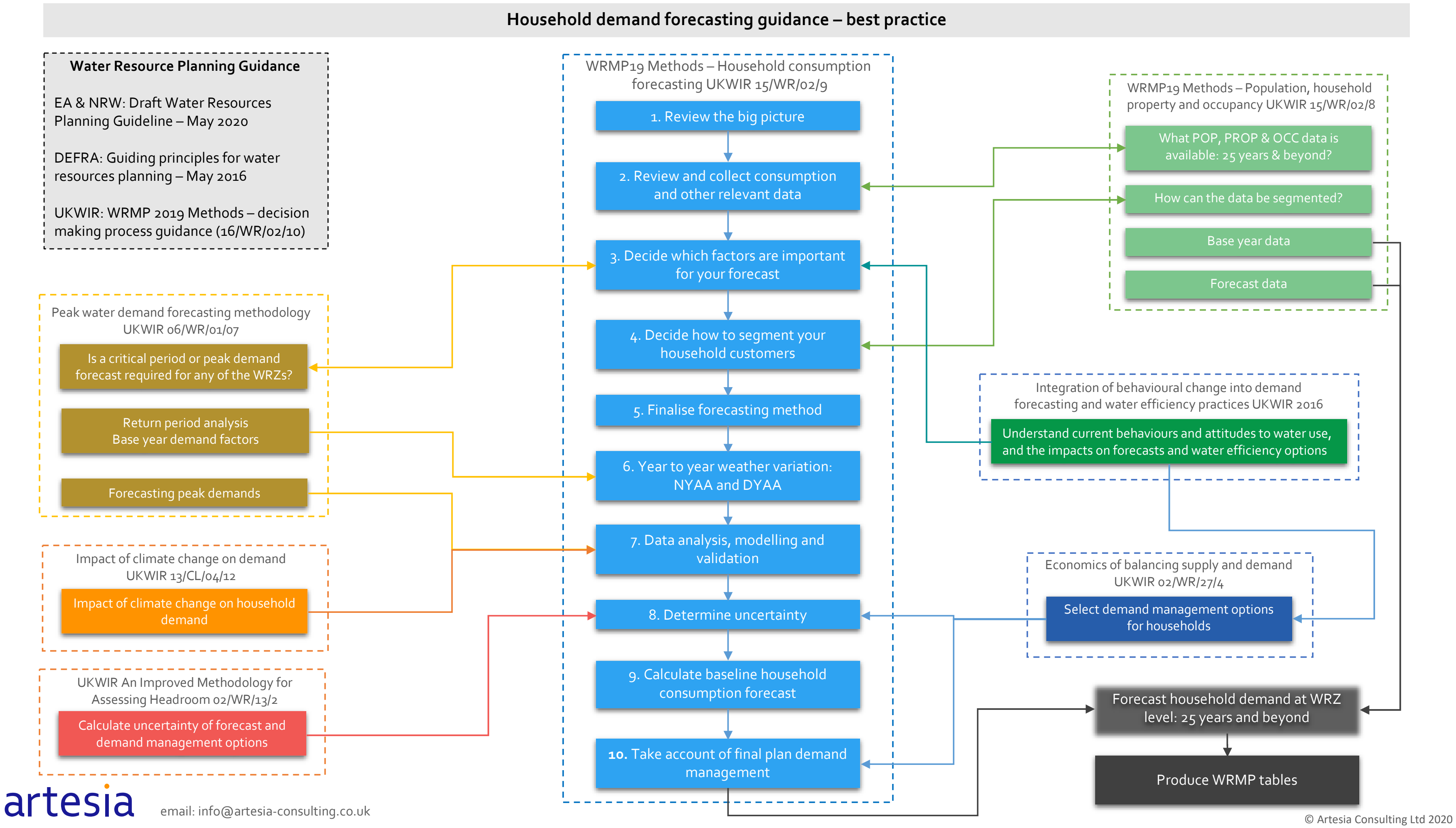
[1] There is no regional plan to cover Wales.
[2] Water Resource Planning Guideline, draft for consultation July 2020. Environment Agency.

- Demonstrate how other information sources have been included, and amended the forecast accordingly
- Clearly describe any limitations in the forecast
- Clearly describe how you have worked with regional groups (where applicable), neighbouring companies and those involved with strategic water resource solutions to align your forecasts.
- Explain the assumptions about how unaccounted populations have been derived.
- Describe how populations have been allocated to the geographically different water resource zones (such as using neighbourhood plans or census data to further subdivide the populations).
- Take account of local council local plans and supporting neighbourhood plans to understand future demands.
- Use improved and updated population and household data in the final WRMP if it is available and describe how this will be done in the draft plan. This should be consistent with that used in the business plan.

1.3 Best practice for developing 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 1.

Figure 1 Household demand forecasting best practice overview



1.4 Household consumption forecasting methods

Household consumption forecasts need to take into account factors such as population growth, climate change impacts, the effect of year-to-year weather variation, and peak demands which occur within years. Such plans have been required for about 20 years.

Household demand can be derived at the property level (per household consumption – PHC) or at the individual level (per capita consumption – PCC). The PHC or PCC household consumption values are then multiplied by either the number of households (for PHC) or the number of people (PCC) in a region to obtain total household demand, which is measured in megalitres per day (Ml/d). Artesia's preference is to produce household-based forecasts to reduce the error of occupancy being introduced into the forecasts.

The process by which household demand is determined and forecasts produced, are generally based one of two modelling approaches:

1. **Micro-component (MC) models**
2. **Multiple linear regression (MLR) models.**

MC models have been used for water demand forecasting in England and Wales from the late 1990s. They quantify the water used for specific activities (e.g. showering, bathing, toilet flushing, dishwashing, garden watering, etc.) by combining values for ownership (O), volume per use (V) and frequency of use (F). For example, per-capita (PCC) or per household consumption (PHC) can be modelled as:

$$PCC\ or\ PHC = \sum_i (O_i \times V_i \times F_i) + pcr$$

Where:

O is the proportion of household occupants or households using the appliance or activity for micro-component i,

V is the volume per use for i,

F is the frequency per use by household occupants or households for i,

pcr is per capita residual demand.

MLR models use standard statistical processes to develop relationships between historic demand and the explanatory factors that influence demand, typically including household occupancy, property type/size and some measure of socio-demographics. The resulting model has a number of model parameters and each has a coefficient that is derived from the model, and there is residual error term. The residual is essentially the consumption component that cannot be explained by the model parameters. Residuals are used for estimating error and developing further modelling refinements.

Some of model parameters will vary over time, whilst some are static over time.

Depending on the data available, problem characterisation, challenges that already exist and length of forecast required, either the MLR or MC models may be more appropriate.

No matter which method is selected, an overall modelling framework has been developed by Artesia which outlines the steps needed to develop the forecast. This is shown in Figure 2.

By producing a framework in this way, we ensure that:

- no step is omitted,
- there is full transparency in the method,
- allows consistency between the company outputs
- the process can be streamlined for automation resulting in complete auditability and repeatability of the outputs.

Figure 2 Household consumption forecasting framework for MLR and MC models

Phase	Task No.	MLR	MC
A. Data collection and formatting	1	Discuss the project requirements, finalise scope and produce a data specification	
	2	Collect and organise the data, considering data management protocols	
	3	Data formatting and submit data queries	
	4	Quality assurance of the data	
B. Population and property separation and exploratory analysis	5	Finalise model segmentation (e.g. umHH, mHH, etc)	
	6	Split the property and population forecasts into defined segmentations	
	7	Select and agree the modelling method following risk assessment	
	8	EDA of consumption data, explanatory factors and weather	-
	9	Outlier removal and gap analysis for each variable	-
C. Model build and testing	10	Undertake variable selection and develop the base year HHCF model	Apply ownership, volume and frequency (OVF) values to forecast
	11	Test the model	
	12	Calibrate the model to the base year per area/zone	
D. Model refinement and forecast	13	Residual modelling and testing (spatially and temporally)	-
	14	Select final model	-
	15	Apply normal year correction	-
	16	Forecast the model	
	17	Apply agreed trends to the forecast	
E. Weather modelling and peak factors	18	Compute dry year factors at required granularity	Compute normal year and dry year factors at required granularity
	19	Select return period and peak factor duration	
	20	Compute critical period factors per area/company, as required	
F. Scenarios, climate change and uncertainty	21	Collate outputs to company level	
	22	Apply climate change factors	
	23	Undertake uncertainty analysis	
	24	Run appropriate steps from 5-23 again, for any agreed scenarios to be tested	
G. Baseline outputs	25	Micro-component split outputs and EA table	
	26	Output forecast in a format specific to original requirements	
	27	Audit reporting	

1.5 Cambridge Water specific requirements

Water companies are required to use methods for supply and demand analysis that are appropriate to the level of planning concern in their water resource zones (WRZs), as given in the Water Resources Planning Guideline[2].

The UKWIR Household consumption forecasting guidance identifies the following methods for forecasting household consumption (in approximate order of complexity):

- Use existing study data;
- Trend based models;
- Per-capita methods;
- Variable flow methods;
- Macro-components (referred to as 'major consumption groups' hereafter);
- Micro-components;
- Regression models;
- Proxies of consumption; and
- Micro-simulation.

The criteria presented in Table 1 were developed in the UKWIR consumption forecasting guidance to assess the forecasting methods.

Table 1 Criteria for evaluating consumption forecasting methods

Criteria	Comment
Acceptance by stakeholders	The method should stand up to scrutiny from the regulators, and other external stakeholders, including customers.
Explicit treatment of uncertainty	The method should recognise that there will be uncertainty around the forecast and should quantify the level of uncertainty.
Underpinned by valid data	The method should be based on data that is valid for the area under consideration.
Transparency and clarity	The method needs to be understood and should be able to be replicated by others.
Appropriate to level of risk	The method should be appropriate in terms of cost and data requirements for the planning problem being addressed; i.e. the degree of vulnerability to a supply demand deficit.
Logical and theoretical approach	The method should command confidence to practitioners and decision makers. It should address those factors that people believe drive water demand, and it should be relevant to historical trends.

Empirical validation	The method should enable comparison to outturns or past projections. It should be possible to test the method on past data to predict demand, and predict any explanatory factors used in the forecast.
Explicit treatment of factors that explain HH consumption	The method should be able to take account of the different factors which drive household demand, and different segments of consumers with respect to household water use.
Flexibility to cope with new scenarios	The method should be method flexible enough to run different household consumption forecasts.

The overall problem characterisation for Cambridge Water is 'high'. An assessment of suitable household consumption forecasting (HHCF) methods was carried out based on this characterisation. This indicated that regression modelling would be the preferred forecasting approach for this level of concern.

However, Cambridge Water do not have sufficient data/information on individual household consumption/property characteristics to enable MLR modelling. Micro-component forecasting scored second overall and would be a suitable alternative in the circumstances.

After discussions with Cambridge Water and following a review of the big picture, the decision was made to produce an **MC based model** for WRMP24 HHCF, and this report discusses the methodology, results and conclusions from this work.

The RAG matrix scores produced for this analysis are given in Appendix section 6.1.

2 Methodology

Cambridge Water have selected an MC model for their household consumption forecast based on the available data, and their problem characterisation. This section provides an explanation of the complete HHCF method, including any assumptions made, split by the phases in the modelling framework.

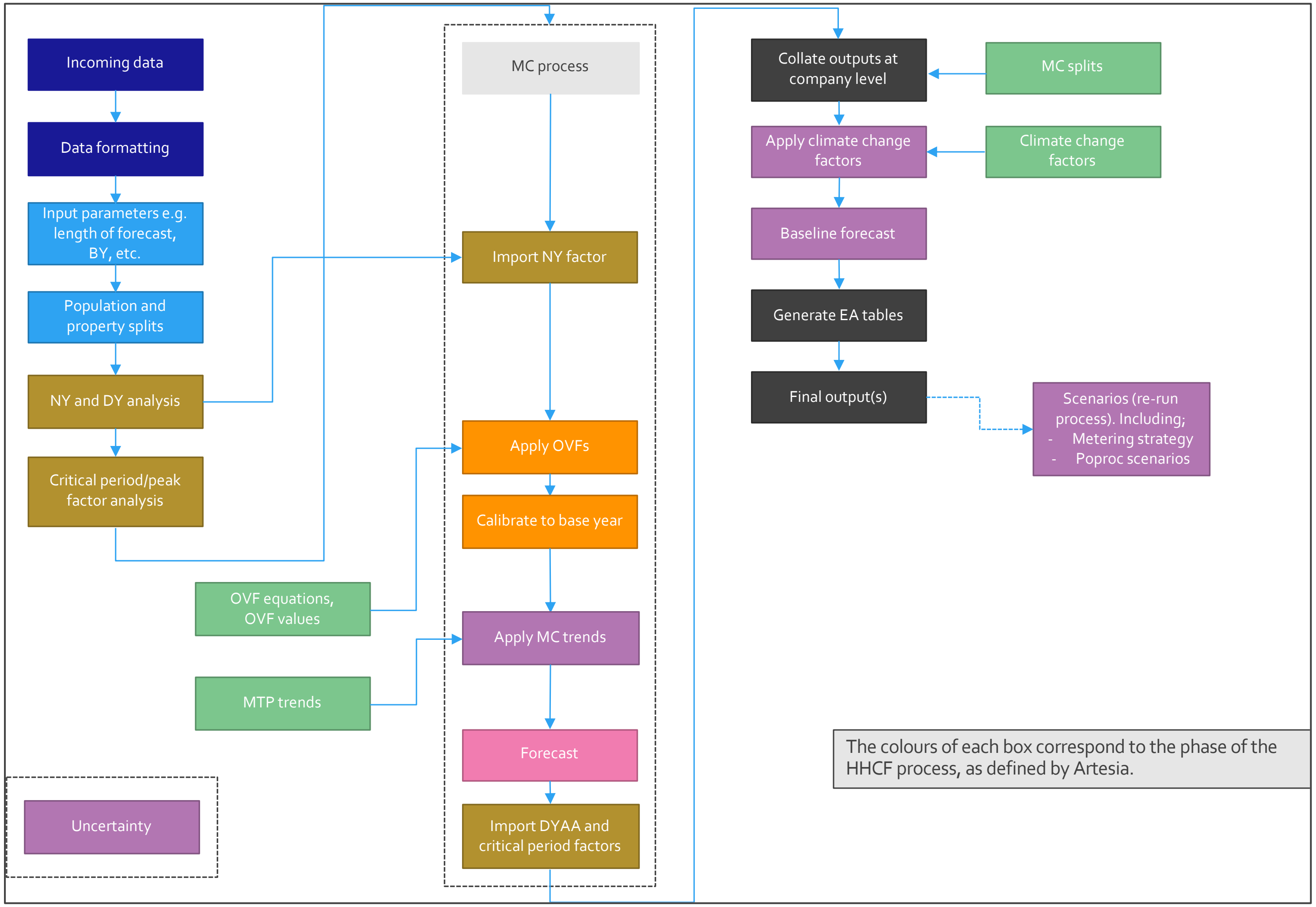
Each subsection (phase) starts with the relevant steps from the modelling process to provide clarity. Note that this is given for both MLR and MC modelling for transparency, though the detail will only be relevant to the MC method used for Cambridge Water.

The results of this process will be presented in section 3.

The MC model largely follows the process described in Figure 3. This is colour coded by the phases of the HHCF process, and so it shows that the steps are not entirely chronological. Therefore, although the phases of the process will be discussed in this section in the order given in Figure 2, this is sometimes not the order that is used in reality.

Note that the boxes in Figure 3 that are coloured in green are not specifically related to a particular phase but represent external data sources or analyses which are used in the corresponding process. For example, the “MC splits” which are used to separate the resulting consumption predictions into the components required for the EA tables were derived from a previous piece of work by Artesia to map from one to the other. Similarly, the “OVF equations and OVF values” form the basis of the micro-component model with the data used to generate the OVFs coming from a combination of studies by UKWIR and WRc.

Figure 3 Flowchart showing the stages of the MC model build coloured by the stages in the HHCF framework



2.1 Data collection and formatting

Task No.	MLR	MC
1	Discuss the project requirements, finalise scope and produce a data specification	
2	Collect and organise the data, considering data management protocols	
3	Data formatting and submit data queries	
4	Quality assurance of the data	

The amount of data required to build or update a household consumption forecast is vast, regardless of whether an MLR or MC model is used. The premise of a forecast is to collect enough historic data to understand the relationships between different factors and extrapolate this forward with confidence.

To streamline this process, the data requirements table provided in Figure 4 was used to accurately capture all necessary information. This list is colour coded according to the phase in which the data is required and is split into both the MLR requirements on the left, and MC requirements on the right.

Since MC based models are based upon assumptions of the ownership, volume and frequency of use of each of the micro-components, there are much fewer data sets required to build the model (orange phase in Figure 4). This is a key factor in determining if a regression-based model is possible during the problem characterisation. Aside from the model build, the data requests are the same.

Figure 4 Data requirements for MLR and MC methodologies

<table>
<tr><th>MLR Data requirements</th><th>MC Data requirements</th></tr>
<tr><td colspan="2">All household property and population forecasts, split into the same granularity as the forecast requires (e.g. zonally, company, regionally, etc).</td></tr>
<tr><td colspan="2">Metering strategy property forecasts. E.g. optant and compulsory metering forecasts split into the same granularity as the forecast requires.</td></tr>
<tr><td colspan="2">Base year property and population data, split into the forecast granularity (e.g. WRZ) as well as split into the forecast segmentation (e.g. measured, optants, unmeasured).</td></tr>
<tr><td colspan="2">Historic population and property data split into the forecast granularity (e.g. WRZ) as well as split into the forecast segmentation (e.g. measured, optants, unmeasured).</td></tr>
<tr><td colspan="2">Different population and property forecast scenarios, if applicable. This should be at the same granularity/segmentation as the baseline poproc forecast.</td></tr>
<tr><td>Consumption monitor data e.g. IHM, or area level. Data needs to be collected at least at annually, preferably monthly. This data should be as up to date as possible, with at least 5 years historic data. If this is not available, at least 12 months is necessary.</td><td>-</td></tr>
<tr><td>Property level demographics that can be attached to the consumption monitor data, preferably from the same period as the consumption data. This should include as a minimum; occupancy, meter status (linking to the forecast segmentation), property type and ACORN/Mosaic. Ideally, metrics about the occupants, the property and the area.</td><td>-</td></tr>
<tr><td>Demographic data for each area (WRZ, region, etc) for the base year, for each segment. E.g. the proportion of property types, ACORN and occupancy per segment for each zone.</td><td>-</td></tr>
<tr><td>Demographic data for each area for historic years if available, for each segment. E.g. the proportion of property types, ACORN and occupancy per segment for each zone.</td><td>-</td></tr>
<tr><td>Forecast of demographic data, if available, for each area, for each segment.</td><td></td></tr>
<tr><td colspan="2">Annual return consumption data (PCC, PHC and Ml/d) for the base year, split into the required segmentation at the forecast granularity.</td></tr>
<tr><td colspan="2">Annual return consumption data (PCC, PHC and Ml/d) for historic years, split into the required segmentation at the forecast granularity.</td></tr>
<tr><td colspan="2">Weather data, including as a minimum; temperature, rainfall and sunshine using at least monthly granularity.</td></tr>
<tr><td colspan="2">Historic DI data, preferably after the removal of leakage and non-household usage, to leave domestic consumption. This should be using the same granularity as the forecast.</td></tr>
<tr><td colspan="2">Base year for forecast</td></tr>
<tr><td colspan="2">Length of forecast</td></tr>
<tr><td colspan="2">Granularity for model</td></tr>
<tr><td colspan="2">Model segmentation</td></tr>
<tr><td colspan="2">Output format</td></tr>
</table>

In addition to the data given in Figure 4, it may sometimes appropriate for us to collect additional data from open source locations, such as the Office for National Statistics (ONS) or the Met Office. This may be necessary if company specific weather data is unavailable, or if there is still a high level of uncertainty in the forecast which may be explained using external data sources. If this is the case, this will be explicitly stated.

To adhere to the fully transparent and auditable process that the framework offers, an input template has been put together to collate all of the data required in Figure 4 to allow a simple way to sense check the outputs, as well as ensuring that all of the data units are consistent and visible. Figure 5 shows an extract of this template with tabs specifically for the following data:

- Annual return
- Metering strategy forecast
- Population, property, occupancy (POPROC) forecast
- Forecast trends
- Historic meter strategy data
- Weather
- DI.

Figure 5 Extract of the data input template

	A	B	C	D	E	F	G	H	I	J	K
1	Company/WRZ	Area	FY	Measured/Unmeasured	Method	Population	Properties	Consumption	Occupancy	PCC	PHC
2	Company		1992-93	Measured	Other	117	46.03	#N/A	2.541820552	#N/A	#N/A
3	Company		1992-93	Unmeasured	Other	6496	2535.45	867.560288	2.56206985	133.553	342.1721146
4	Company		1993-94	Measured	Other	199.792	78.413	24.78999157	2.547944856	124.079	316.1464498
5	Company		1993-94	Unmeasured	Other	6443.751	2528.997	869.4488787	2.547947269	134.929	343.7919771
6	Company		1994-95	Measured	Other	269.5	106.4	31.7883335	2.532894737	117.953	298.7625329
7	Company		1994-95	Unmeasured	Other	6395.8	2524.3	870.4619842	2.533692509	136.099	344.8330168
8	Company		1995-96	Measured	Other	342.21	136.221	39.80997372	2.512167727	116.332	292.2454961
9	Company		1995-96	Unmeasured	Other	6335.46	2521.916	910.6843622	2.512161388	143.744	361.1081266
10	Company		1996-97	Measured	Other	388.47	158.302	42.3665382	2.453980367	109.06	267.6310988
11	Company		1996-97	Unmeasured	Other	6340.39	2511.205	875.6776033	2.524839669	138.111	348.7081315
12	Company		1997-98	Measured	Other	399.1	184.502	53.55922	2.163120183	134.2	290.2907286
13	Company		1997-98	Unmeasured	Other	6340.28	2496.786	894.930522	2.539376623	141.15	358.4330103
14	Company		1998-99	Measured	Other	471.58	219.094	62.3334444	2.152409468	132.18	284.5054835
15	Company		1998-99	Unmeasured	Other	6256.57	2478.991	862.3430431	2.523837319	137.83	347.8604977
16	Company		1999-00	Measured	Other	568.18	246.309	74.374762	2.306777259	130.9	301.9571433
17	Company		1999-00	Unmeasured	Other	6160.98	2460.925	853.2341202	2.50352205	138.49	346.7127687
18	Company		2000-01	Measured	Other	593.13	280.237	80.1140691	2.116529937	135.07	285.8796986
19	Company		2000-01	Unmeasured	Other	6131.29	2439.71	863.1630062	2.513122461	140.78	353.7973801
20	Company		2001-02	Measured	Other	658.71	311.222	89.6043213	2.116527752	136.03	287.9112701
21	Company		2001-02	Unmeasured	Other	6055.48	2409.916	860.7259272	2.512734884	142.14	357.1601364
22	Company		2002-03	Measured	Other	813.88	357.428	104.1522236	2.277046006	127.97	291.3935774
23	Company		2002-03	Unmeasured	Other	5708.73	2375.347	850.3724208	2.403324651	148.96	357.99924
24	Company		2003-04	Measured	Other	928.33	416.333	122.446727	2.229777606	131.9	294.1076662
25	Company		2003-04	Unmeasured	Other	5635.1	2326.727	847.857146	2.421899948	150.46	364.3990662
26	Company		2004-05	Measured	Other	1053.98	478.467	136.332313	2.202826945	129.35	284.9356654
27	Company		2004-05	Unmeasured	Other	5548.14	2271.909	799.486974	2.44206084	144.1	351.900967
28	Company		2005-06	Measured	Other	1175.65	532.696	154.95067	2.206981092	131.8	290.880108
29	Company		2005-06	Unmeasured	Other	5468.48	2223.923	789.921936	2.458934055	144.45	355.1930242
30	WRZ	Area A	2005-06	Measured	Other	20.291	8.784	2.529093093	2.309995446	124.6411263	287.9204341
31	WRZ	Area A	2005-06	Unmeasured	Other	80.68	36.353	11.53562012	2.21934916	142.9799221	317.32237
32	WRZ	Area B	2005-06	Measured	Other	1130.798	513.493	149.3343346	2.202168287	132.0610176	290.8205849
33	WRZ	Area B	2005-06	Unmeasured	Other	5254.176	2129.383	758.8063818	2.467464049	144.4196734	356.3503521
34	WRZ	Area C	2005-06	Measured	Other	3.273	1.409	0.410699083	2.32292406	125.4809298	291.4826709
35	WRZ	Area C	2005-06	Unmeasured	Other	9.894	4.172	1.432093253	2.371524449	144.7436075	343.263004
36	WRZ	Area D	2005-06	Measured	Other	21.29	9.01	2.652647921	2.362930078	124.5959568	294.4115339
37	WRZ	Area D	2005-06	Unmeasured	Other	123.728	54.015	18.08942165	2.290622975	146.2031364	334.8962632
38	Company		2006-07	Measured	Other	1277.36	588.918	155.51858	2.168994665	121.75	264.0751004

README | Annual Return | Forecast Metering Strategies | Forecast POPROC | Trends | Cohort Historical Dat | +

As part of this project, Cambridge Water provided the following data, corresponding to the data requirements in Figure 4.

Annual Return/ DI

- Annual returns property and population both regions for DFs -Artesia
- CAM HHCF Model v1.4 - updated frontpage - Final Plan

Growth and metering forecasts.

- CAM HHCF Model v1.4 - updated frontpage - Final Plan
- Optant_forecast
- WRE & OxCam Forecasts - 01.07.2020
- Copy of meter profiles as at Oct 20.

Weather data

- Cambridge_weather_data_raw_20200929

In addition to the data provided by Cambridge Water, the following data was collected from publicly available sources to enhance the modelling:

- POPROC - Edge population and property forecasts from WRE.

Once this data was collated, it was subjected to quality assurance checks to ensure the following:

- The units were known and consistent
- No missing data was present
- The data format was as expected (e.g. if a numeric value is expected, this is not formatted as text or as an image).

Statistical quality assurance checks are conducted during the model build stage, and so are not appropriate here. The purpose of the initial checks is to verify that the data matches the requirements list, and there is no ambiguity in the meaning of the data or units.

Finally, the configurations given in Table 2 were provided by Cambridge Water to be used within the household consumption forecast and are therefore assumed throughout the remainder of the document.

Table 2 Model configurations for the Cambridge Water HHCF

Data requirement	Response
Forecast base year	2019-20
Length of forecast	Until 2100
Granularity of the model	Region (one single WRZ)
Model segmentation	measured and unmeasured, including new properties, optants, compulsory and progressive metering
Baseline growth forecast	Housing Plan P BY Rebase

2.2 Population and property separation and exploratory analysis

Task No.	MLR	MC
5	Finalise model segmentation (e.g. umHH, mHH, etc)	
6	Split the property and population forecasts into defined segmentations	
7	Select and agree the modelling method following risk assessment	
8	EDA of consumption data, explanatory factors and weather	-
9	Outlier removal and gap analysis for each variable	-

Now that the data has been received, and the configurations of the model selected, the next task of the framework is to split the property and population forecasts into the defined segmentations.

2.2.1 Population and property splits

Typically, population and property forecasts are supplied at total property level for each water resource zone. As Cambridge Water require the HHCF at meter status (measured and unmeasured) level, it is necessary to split the population and property (POPROC) forecast into the required segments. As the POPROC information supplied for this project contains multiple growth forecasts, this is complicated further as this is required for each version.

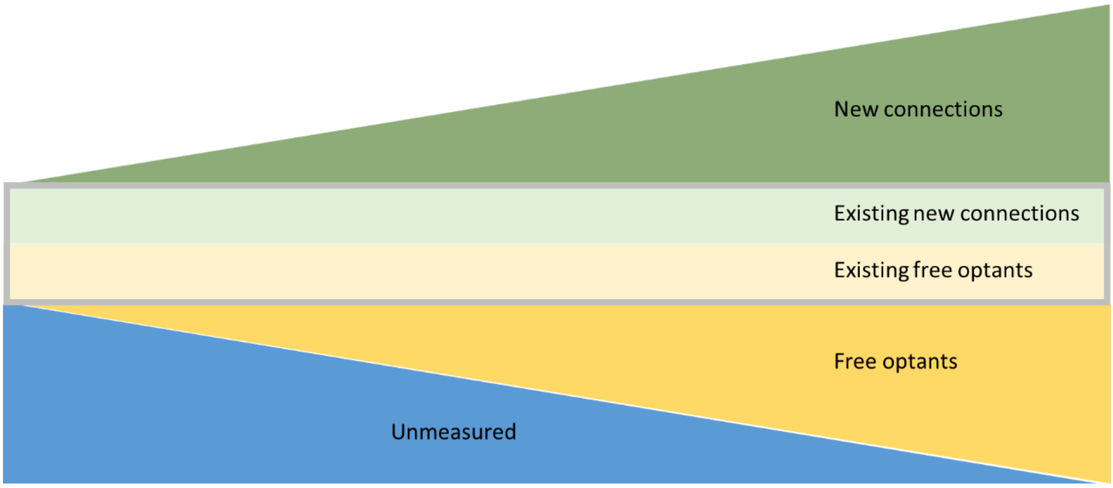
This is not a simple task, particularly for population and occupancy, due to the number of cohorts required (unmeasured, existing measured, compulsory measured, optants, new properties) as well as the complexity in the behaviors between these properties.

In order to split the forecasts, certain data is required, including:

- Data describing the company at the base year.
  - Total number of properties, and how many of these are measured/unmeasured.
  - The number of new properties that will join the companies water supply annually.
  - The occupancy of measured/unmeasured properties.
  - How the measured cohort is divided into new, compulsory and optant cohorts.
- Yearly forecast data. For each June return this must include:
  - The number of properties which will opt onto a meter (optants).
  - The number of properties which will be forced onto a meter (compulsory).
  - A global occupancy forecast.
  - A global property count forecast.
  - The number of properties which will be demolished.

As all of this data has been provided during the data collection stage, a method can be developed to segment the forecasts. The basis of the method is illustrated in Figure 6.

Figure 6 Illustration of splitting POPROC forecast into required cohorts, to the point of 100% meter penetration



In order to achieve this, certain logical assumptions have been made.

- New households will always be measured.
- Free optants move directly out of the unmeasured property segment.
- Voids are forecast to remain constant throughout the forecast period, in that there are no further voids added beyond the base year. Voids have not been included in the baseline forecast due to their negligible consumption.
- Demolitions are distributed evenly across the cohorts.

As well as mapping the properties into each of the segments, population must also be distributed, which is perhaps more complex. Figure 7 demonstrates that as meter penetration increases, the occupancy of the unmeasured and optant properties increase until 100%-meter penetration. Throughout the forecast the sum of the population for the optants plus unmeasured properties remains the same (this assumes that each year optants come from the unmeasured pool). Meanwhile the average occupancy of all the segments must follow the change in occupancy from the property and population forecasts.

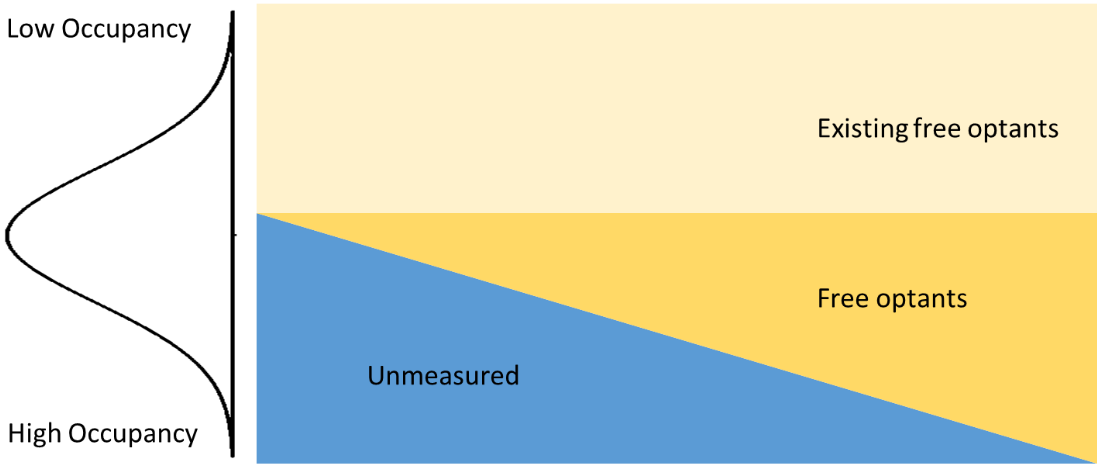
In summary, the assumptions in respect of splitting population are:

- Measured households have lower occupancy than unmeasured households.
- Optants have the lowest occupancy, on average.
- New properties are assumed to have the same occupancy as the average across all properties.
- Compulsory properties are assumed to have the same occupancy as unmeasured households.
- The optant households are taken from the lower end of the unmeasured occupancy distribution.
- As optants leave the unmeasured pool, the average occupancy of the households that remain will increase.

These assumptions provide an estimate of the change in occupancy within the household segments over time, which are applied in an iterative manner. There will of course be a complex movement of population within these segments, reflecting births, deaths, people moving into the region, people moving out of the region, and people moving within the region. However, the intra-cohort variation is not required for the forecast.

Finally, each year the segments are calibrated to consider the company and zonal level occupancy changes throughout the forecast period. To ensure the segmented households and populations sum to the company own forecast, various calibration steps and data validation checks are also included in the calculations.

Figure 7 Illustration of the change in occupancy as meter penetration tends towards 100%



2.2.2 Population and property rebasing

The final step in the separation of the population and property forecasts is the process of rebasing the outputs to match the company annual return (AR) data.

It is not uncommon that a large gap exists between the starting year of the POPROC forecasts, and the company's own annual return data for the same year. This often occurs due to the base year annual return data being unavailable at the point that the POPROC forecasts are provided by external providers. Therefore, a rebasing exercise is required.

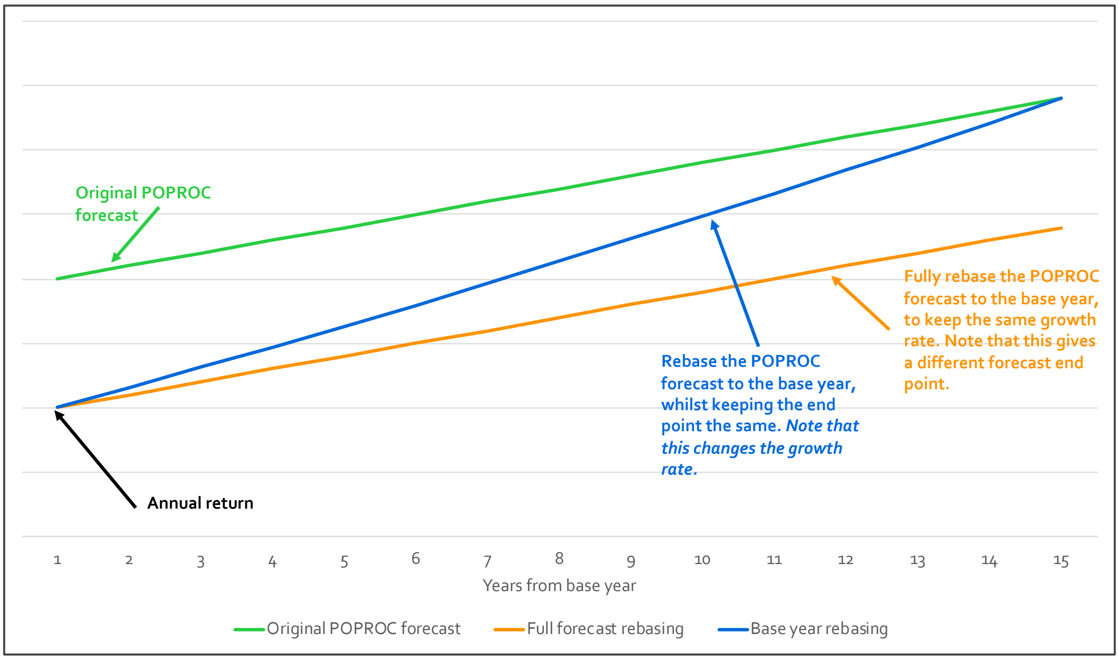
There are 3 main ways in which the population and property information can be rebased, which is shown in Figure 8 using an arbitrary example.

Firstly, the forecast need not be rebased, meaning that the POPROC data between the annual return and the forecast is mismatched, and is akin to the green line in Figure 8. This is the least advisable option as explaining a large difference in the base year is difficult.

The blue and orange lines in Figure 8 show two more reasonable rebasing options, with two main differences.

- **Fully rebasing** (orange) the forecast as in the orange line, ensures that the population and property growth rate remains as per the original data. However, the end point is often lower than the original data suggests. Note that in the case where the original POPROC forecast is *lower* than the annual return data, the "full rebase" option would result in a higher end point, not lower like the graph suggests.
- Conducting a **base year rebase** (blue) changes the original growth rate yet ensures that the end point of the forecast remains the same.

Figure 8 Different rebasing options for POPROC forecast



The selection of the different rebase options (no rebase – green, full rebase – orange or BY rebase – blue), is dependent upon the requirements of Cambridge Water. Following discussions with Cambridge Water it was decided to use the BY rebase option. Therefore, the results presented in section 3 will all be based upon this process, unless explicitly stated otherwise.

2.3 Model build and testing

Task No.	MLR	MC
10	Undertake variable selection and develop the base year HHCF model	Apply ownership, volume and frequency (OVF) values to forecast
11	Test the model	
12	Calibrate the model to the base year per area/zone	

This section explains the method and approach used to build the MC model required for the forecast.

As explained in section 1.4, MC models have been used for water demand forecasting in England and Wales from the late 1990s. They quantify the water used for specific activities (e.g. showering, bathing, toilet flushing, dishwashing, garden watering, etc.) by combining values for ownership (O), volume per use (V) and frequency of use (F). For example, per-capita (PCC) or per household consumption (PHC) can be modelled as:

$$PCC \text{ or } PHC = \sum_i (O_i \times V_i \times F_i) + pcr$$

Where:

> O is the proportion of household occupants or households using the appliance or activity for micro-component i,

V is the volume per use for i,

F is the frequency per use by household occupants or households for i,

pcr is per capita residual demand.

By applying this together with the population or property data, a water demand model can be formed. By forecasting changes in each of the variables (*O*, *V*, *F* or daily water use for each micro-component) over time, a water demand *forecast* can be created. Hence the micro-component forecast model requires estimates of changes in these variables, to reflect future changes in technology, policy, regulation, and behaviour.

This section describes how this modelling process has been applied, and how the inputs have been generated for:

- Base year micro-components from a micro-component occupancy model.
- Final year micro-components from an occupancy model. This allows a rate of change of micro-component daily water use to be derived due to the change in occupancy over the planning period. This is how the forecast is generated.

2.3.1 Selection of the modelling unit

Two commonly used methods of consumption forecasts are based on Per Capita Consumption (PCC) and Per Household Consumption (PHC).

In the case of PHC modelling, occupancy needs to be included as an explanatory variable, and PHC is composed of a consumption allotted to the house on the basis of its characteristics, and an additional consumption assigned to each occupant.

PCC modelling assigns a different consumption value per person on the basis of the characteristics of the property they inhabit.

In the former case, the model is property driven, which aligns with the data collection based on household meter reads.

The latter case introduces all the error associated with the household occupancy figure into the model at the very first step. If the model is based on PCC, the PCC is calculated from estimated occupancy (for which there is an error), so there is no part of the consumption modelling that is independent of occupancy error; all the error in population forecasting is propagated through the zonal forecast if it is based on PCC.

Modelling by PHC makes occupancy-driven household consumption components implicit in the model whereas PCC-driven modelling would need to incorporate a correction for changing occupancy rates in PCC forecasting.

For these reasons, PHC is used as the basis for modelling and aggregating up to a zonal/company-level consumption forecast.

2.3.2 MC occupancy modelling

Whilst the forecast is built at household level, there is an influence on a number of the micro-components from occupancy. For example, it is expected that dishwasher usage increases linearly with occupancy but washing machine use will not hold a linear relationship. Therefore, in calculating the base year and final year PHC values, we use a set of linear models that relate either daily use or frequency of use to occupancy in each year.

Because of the segmentation of the forecast required by Cambridge Water, the model is also used to provide the base and final year values for the different metered property types; existing metered, optants, new properties and compulsory metered.

Once the occupancy model is built, this forms the central part of the MC model, and when combined with the rates of change for each micro-component, a forecast can be generated.

Several national datasets have been used in building this model, to increase the understanding of historic and recent micro-component consumption. Historic micro-components are extracted from the WRc CP187 report (WRc, March 2005) and recent micro-components are extracted from an UKWIR study, (UKWIR, 2016).

This is micro-component data that has been collected by measuring the different micro-components used within the household (as opposed from survey questions and assumptions). This allows ownership (*O*), volume per use (*V*) and frequency of use (*F*), to be calculated for each micro-component. There were two main sources of data for this.

- 2015-16 data collected using the Siloette system:
  - A sample of measured billed households, with associated occupancies and demographic information on the households, collated during an UKWIR Study (UKWIR, 2016). This contains 62 households from around England and Wales.
  - A sample of unmeasured billed households, which do not have associated demographics (collated from other anonymous Siloette studies carried out by Artesia Consulting, from England and Wales).
- 2002 – 2004 *O*, *V*, and *F* data collected using the Identiflow system (a sample of unmeasured billed households, (WRc, March 2005)).

Both the Siloette and Identiflow systems measure the flow into a property and compute the individual micro-components through pattern recognition (although the detailed methodology of the two systems is different).

The UKWIR micro-component data for measured billed households were used for the modelling, because this dataset has a complete set of occupancy data for each household over the logging period. The total number of households in the sample was 62.

The following micro-components were used as part of this model:

- WC flushing
- Shower use
- Bath use
- Tap use
- Dishwasher use
- Washing machine use
- Water softener use
- External use, and
- Miscellaneous use (including internal plumbing losses).

Each of the micro-components were investigated to determine whether the daily volume per use, frequency of use or ownership varied significantly with occupancy. The following micro-components showed relationships where occupancy was a significant factor:

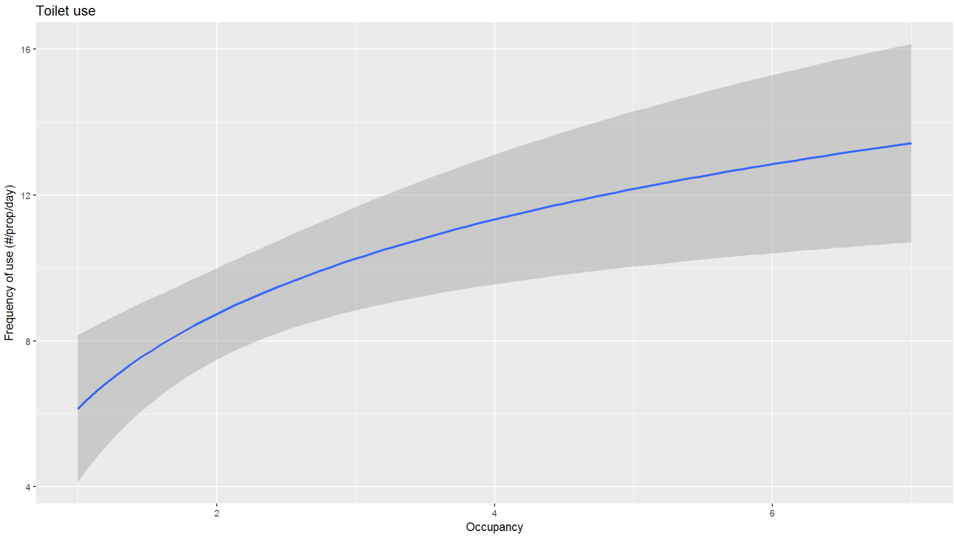
- WC flushing (toilets)
- Shower use
- Bath use
- Tap use
- Washing machine use.

For each of these micro-components (toilets, showers, baths, washing machines and taps) a linear model was developed using occupancy as the predictive factor.

To illustrate this, Figure 9 shows the variation of toilet flushing per day with occupancy, with the mean frequency of use per day plotted against occupancy. The model is a logarithmic relationship of frequency of use against occupancy with the following equation.

$$Frequency\ of\ use\ (uses\ per\ day) = 6.143 + 3.744 \times \ln\ (occupancy)$$

Figure 9 Variation of toilet flushing frequency (uses per day) with occupancy



This same exercise was repeated for showers, baths, washing machines and taps to generate frequency of use equations (or total daily volume equations) for the MC model, which are shown in Table 3.

Table 3 Use equations using occupancy driven micro-components

Micro-component	Use/Volume equations	Equation reference
Toilet	$Uses\ per\ day = 6.143 + 3.744 \times \ln\ (occupancy)$	1
Shower	$Volume\ per\ day = 15.47 + 57.47 \times \ln\ (occupancy)$	2
Bath	$Volume\ per\ day = 7.181 + 7.378 \times \ln\ (occupancy)$	3

Washing machine	$Uses\ per\ day = 0.3242 + 0.43705 \times \ln(occupancy)$	4
Tap	$Volume\ per\ day = 27.92 + 62.89 \times \ln(occupancy)$	5

The final step is to separate out the relationships between the micro-components and the metering status of the property, based on the cohorts being modelled. Table 4 shows the variations of the toilet, washing machine, dishwasher and plumbing losses micro-component volumes with meter cohort type. Toilets contain the largest variation, with new builds having the smallest flush volumes, consistent with new build regulations. Unsurprisingly, unmeasured properties have the highest toilet flush volumes, which by default causes compulsory metered properties to have the same value (as compulsory metered properties are taken from the unmeasured pool).

Table 4 Micro-component volumes dependent on meter status

Property type	Toilet flush volume (mean l/flush)	Washing machine volume/use (mean l/use)	Dishwasher volume/use (mean l/use)	Wastage / plumbing losses (frequency of occurrence)
Unmeasured household	7.58	54.19	16.7	0.825
Existing measured	7.26	54.19	16.7	1.55
Optant	6.0	54.19	16.7	0.275
New build	5.5	50.0	15.0	0.275
Compulsory metered	7.58	54.19	16.7	0.275

Bringing all of this information together, Table 5 shows the final ownership (O), volume (V) and frequency (F) values for each micro-component, and these are combined to give daily use per micro-component in the model. This is sometimes referred to as the "OVF" model.

Table 5 MC occupancy model parameters

Micro-component	Weighted Ownership 'O'	Volume per use 'V' (l/use)	Frequency of use 'F' (uses/day)	Daily use (l/prop/day)
Toilets	1	See Table 4	See Equation 1	$O \times V \times F$
Showers	-	-	-	See Equation 2
Baths	-	-	-	See Equation 3
Taps	-	-	-	See Equation 5
Dishwashers	0.42	See Table 4	0.5	$O \times V \times F$

Washing machines	0.95	See Table 4	See Equation 4	$O \times V \times F$
Water softeners	0.02	52.06	0.97	$O \times V \times F$
External use	0.18	285.18	0.07	$O \times V \times F$
Plumbing losses	0.22	37.2	See Table 4	$O \times V \times F$
Miscellaneous	0.95	1.63	3.74	$O \times V \times F$

These values can be used to define an MC model to calculate the micro-component daily use (and hence the per household consumption, PHC) for the following property types based on the occupancy assigned to each property type, in the base year and in the final year of the forecast:

- Unmeasured households
- Existing metered billed households
- Optant households
- New build metered households
- Compulsory metered billed households.

Using the base year and final year PHC values, a rate of change in PHC due to occupancy change can be calculated for each household metered status. This is what enables the forecast to be generated These are in addition to any technology and behaviour trends described in section 2.4.2.

However, before the forecast is created, the data requires calibration to the base year, to ensure that there are not any large gaps or deviations from the annual return data in the selected base year, 2019-20.

2.3.3 Base year calibration

At this point, the base year and final year PHC values have been generated from the occupancy model. This model relates each micro-component to known household behaviours using occupancy as a variable. For each of the household segments, the OVF models are applied using the base year occupancy values. However, it is entirely possible that the annual return data for Cambridge Water does not match the base year PHC values generated by the model. Therefore, a calibration is required before the rates of change are computed and a forecast generated.

There are two approaches that can be taken to calibrate the base year, and these are either before or after the application of the normal year factors. The normal year factors are values (typically around 1) that are designed to remove any influence of abnormal weather from the base year PHC/PCC values. This kind of normalisation is required so that the forecast does not contain any additional weather-related influences, making future scenarios difficult to apply.

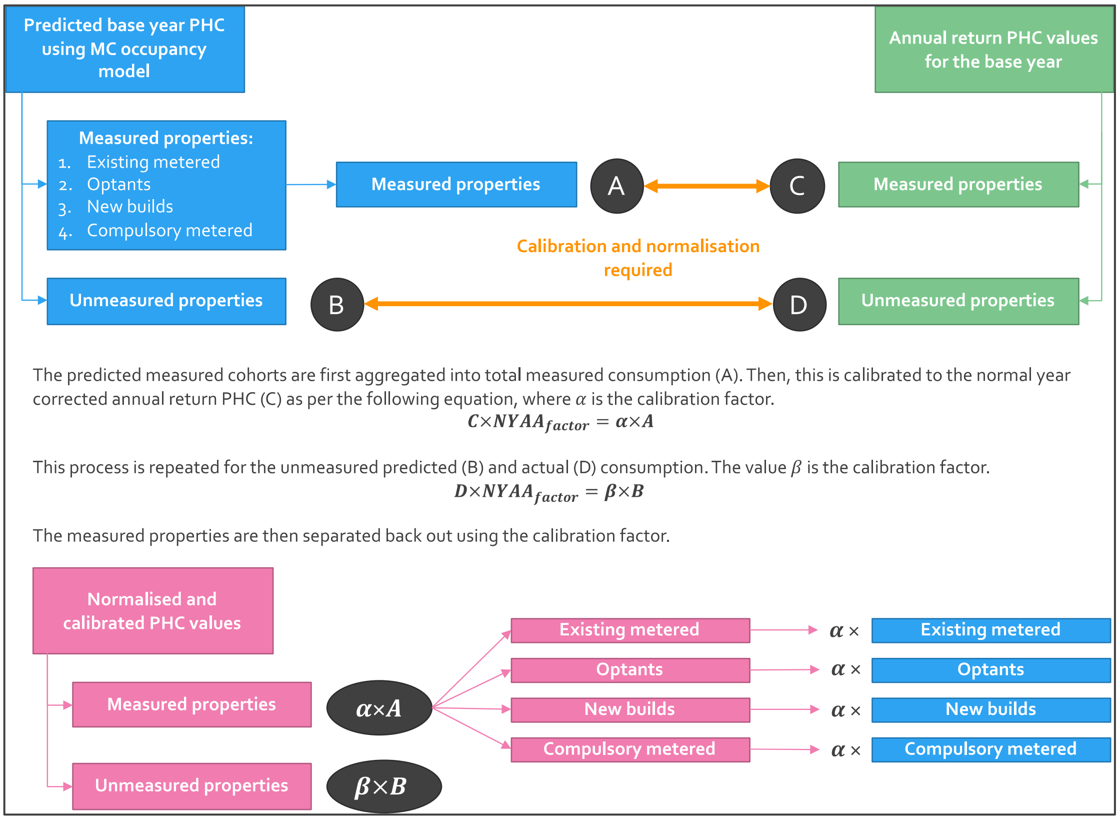
Therefore, it is important that the NYAA factor is applied within the base year calibration to ensure that the subsequent rates of change over time for each component is not affected by annual variation that might by contained within the base year.

The calculation of the weather correction factors is explained in detail in section 2.5. So, instead of calibrating the predicted base year PHC values to the annual return data and applying the normal year correction afterward, the AR data is normalised and then the calibration takes place. This is the approach that has been taken in this model.

Since the AR data is only given at measured and unmeasured granularities, the first stage is to combine the predicted measured PHC values to "total measured" before the calibration takes place. The PHC values for the non-reported figures; existing measured, new builds, optants and compulsory metered, are calculated proportionally based on the NYAA measured calibration factor, using the OVF values in each segment.

This is illustrated in Figure 10.

Figure 10 Illustration of the base year normalisation method



2.4 Model refinement and forecast

Task No.	MLR	MC
13	Residual modelling and testing (spatially and temporally)	-
14	Select final model	-
15	Apply normal year correction	-
16	Forecast the model	
17	Apply agreed trends to the forecast	

Now that the MC model has been produced, the final step is to compute the baseline micro-component trends (rates of change) to apply on top of the PHC values from the occupancy model and generate the forecast. Note that this forms the basis of the *baseline scenario.* It is possible to alter these rates of change based on differences in technological and behaviour trends as touched on in the next section, but these are added separately and are explained in more detail in section 2.4.2.

2.4.1 Micro-component trends

The baseline micro-components trends due to technology change, policies and regulation, and behaviour change, have been computed using the same data sets from the UKWIR and WRc studies, (UKWIR, 2016) (WRc, March 2005) as used in the occupancy model. However, we also use the data from Defra's Market Transformation Programme (MTP)[3].

The MTP produced predictions of water use for different water using appliances in 2030 for three different scenarios:

- Reference scenario (equivalent to the baseline scenario)
- Policy scenario (assuming more effective implementation and accelerated take-up of more sustainable products)
- Early best practice (EBP) which assumes a more positive impact than the policy scenario and an early take up of innovative water efficient products.

We focus on the "reference scenario" to define the baseline trends. This has been done for all of the micro-components, though this is just provided for toilet flushing here, to give an example of the process used.

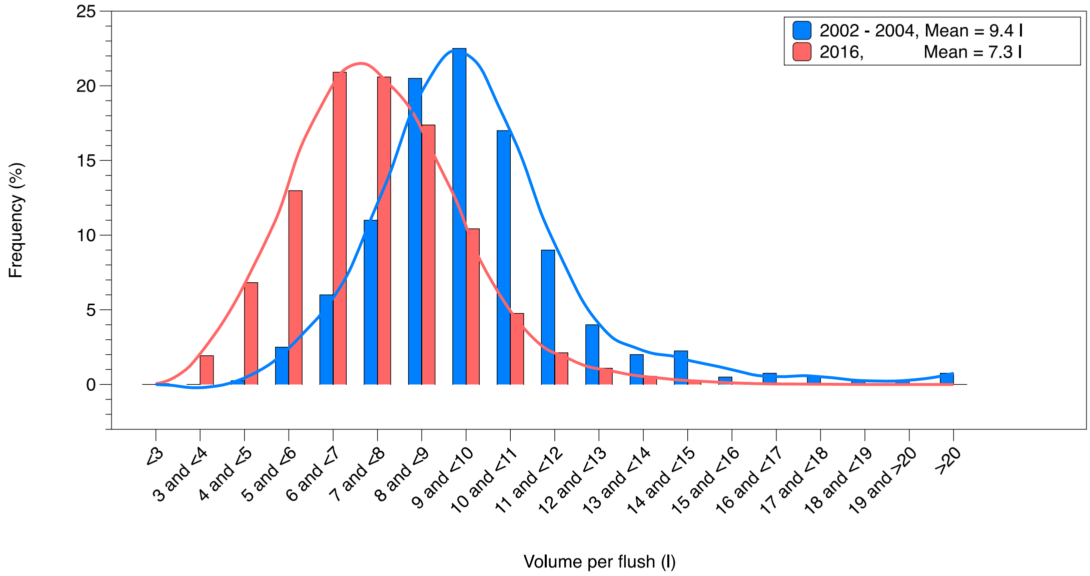
2.4.1.1 Toilet flush volumes

For the toilet flush volume trend, we assume that ownership and frequency of use remains constant, with the volume per use changing due to market transformation.

Using the available data, we created a histogram of the volumes per flush. These are shown in Figure 11 and Figure 12. This shows that for 2002/04 the mean flush volume was 9.4 litres per flush, with a range of flush volumes from 5 litres to more than 15 litres. In 2015/16 the mean flush volume had reduced to around 7.3 litres with a range from 3 litres to about 13 litres per flush.

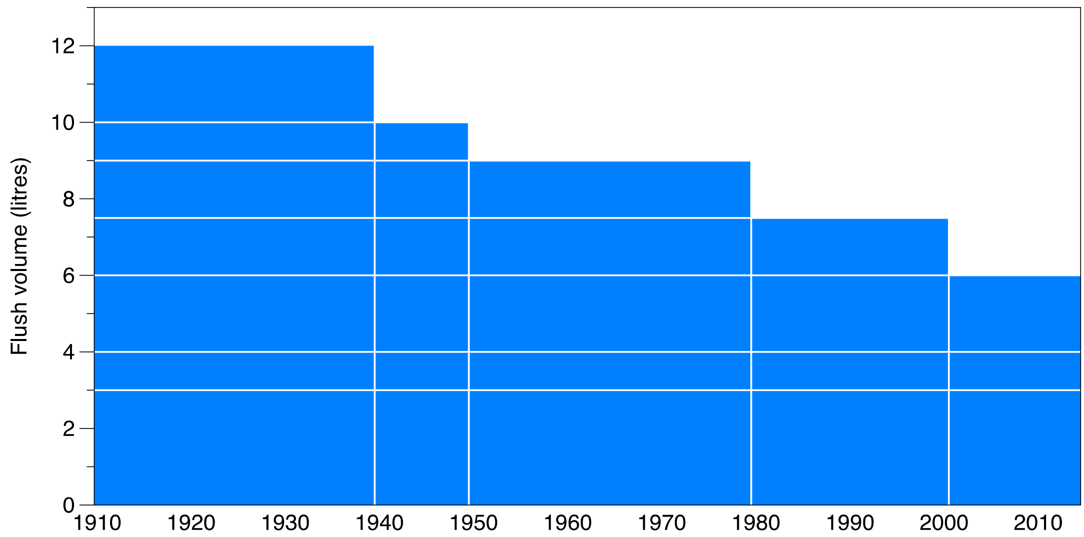
[3] For example, Defra (2011) BNWAT01 WCs: market projections and product details. Note that the MTP reports do not appear to be available online anymore

Figure 11 Histogram of historic flush volumes



The reason for this reduction in flush volumes is due to the replacement of larger volume toilet cisterns with smaller volume cisterns, due to market transformation based on regulatory policies. The schematic in Figure 12 shows the change in maximum flush volumes over time due to changes in regulation. From 12 litres in 1910 to a 6-litre single flush (or 6/4 or 6/3 litre dual flush) in 2000 to date. The reason we see larger flush volumes in the histogram is due to incorrectly setting up the fill height or over filling during the flush period.

Figure 12 Regulatory changes in flush volumes



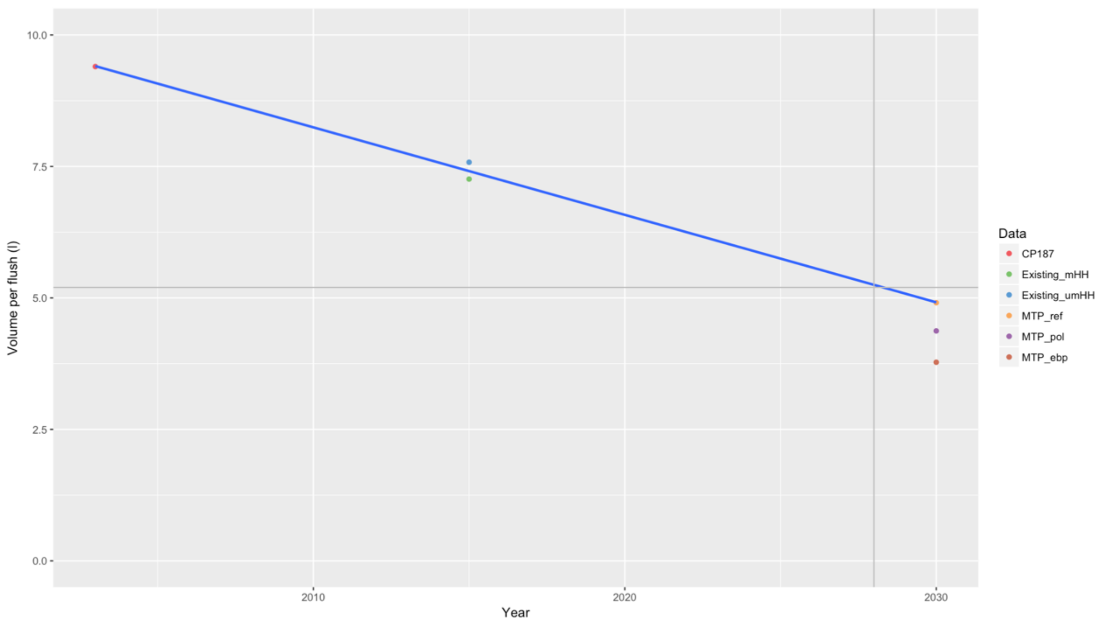
The latest projections for toilet flush volumes[4] in 2030 for the reference scenario is 4.8 litres/flush. Figure 13 shows the mean 2002/04 (CP187), the 2015/16 flush volumes and the

4 Source: http://efficient-products.ghkint.eu/spm/download/document/id/954.pdf

flush volume from the MTP scenarios in 2030. The blue line shows the linear fit from the 2002/04, 2015/16 and MTP Reference scenarios.

If we assume that the market transformation continues at the current rate (a reasonable assumption for baseline forecasts, as there are no planned regulatory changes in toilet flush volumes), then the flush volume in 2028 will be approximately 5.1 litres (shown by the intersect of the grey lines in Figure 13). This provides some confidence in the MTP reference scenario for toilet flush volumes.

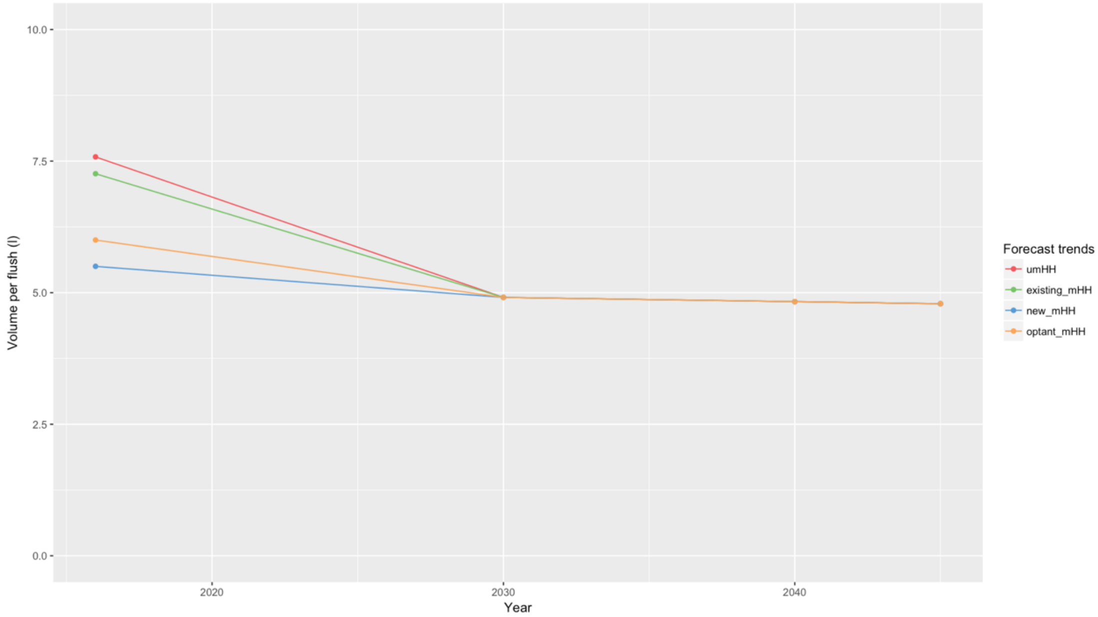
Figure 13 Historic, current and future flush volumes



We have therefore created future trends for toilet volumes per flush (see Figure 14) using:

- the base year volumes per flush in Table 4 for different property types,
- the 2030 projection for toilet flush volumes from the MTP reference scenario,
- an assumption that all property types will have achieved the MTP Reference scenario between the forecast base year and 2030 (for the baseline forecast assuming no change to current WC flush regulations),
- and an assumption that the volume per use will then remain relatively constant until 2050.

Figure 14 Trends for toilet flush volumes



From these trends, annual rates of change have been produced for each of the property types. The rates of change are then incorporated into the model to produce the forecast.

Note that since the final year of the forecast for Cambridge Water is 2100, these trends are held flat for all micro-components from 2050 until 2100. This is because there is a much higher level of uncertainty of these continued rates of change this far into the future.

2.4.2 Apply additional trends

The previous section describes the process used to determine the future micro-component trends which is required to produce the forecast. However, this is focused on the "reference scenario", (or the baseline scenario). Sometimes, it is necessary to include stricter assumptions about the micro-component trends to include within the baseline scenario. Or more likely, other trends are required for the generation of additional scenarios.

For Cambridge Water, the reference scenario is to be used for the baseline outputs, however time was spent producing additional trends using the alternative MTP values[5] for the scenario outputs.

These two additional trend scenarios based on micro-component trends to account for variations within the future predicted rate of change in consumption. These are:

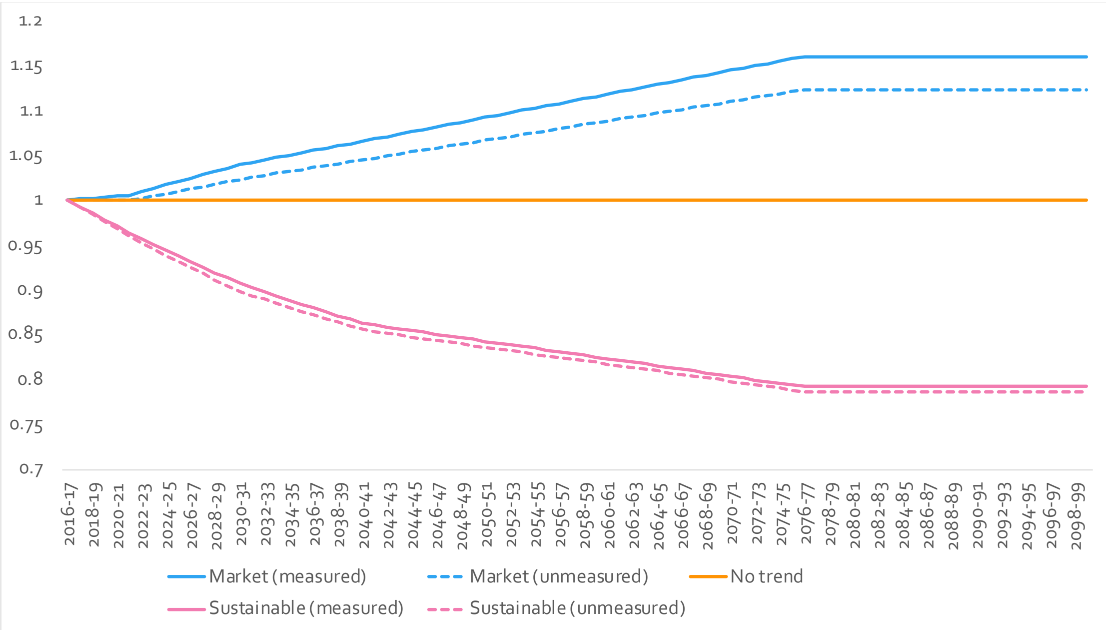
- **Sustainable Development:** This scenario assumes that the current paradigm of regulatory driven incremental technological efficiencies will continue past 2045 and arrive at an endpoint that is conceivable with existing technologies but currently not economically viable. Artesia consider that this represents the 10th percentile trend.
- **Market Forces:** This scenario assumes that the projected trend in micro components does not continue beyond 2022. This would require a situation such as

[5] For example, Defra (2011) BNWAT01 WCs: market projections and product details. Note that the MTP reports do not appear to be available online anymore

Brexit where UK building regulations may be decoupled from current standards and the logical decline in flush volumes is curtailed. The observed upward trend in showering continues to increase. Artesia consider that this represents the 95th percentile trend.

The variation in the trends are shown in Figure 15, for both measured and unmeasured, assuming a baseline of "no trend". As per the baseline trend, these trends are applied until 2050 (only in the scenario where they are selected) and held flat until the final year of the forecast, as the uncertainty is far greater that far into the future.

Figure 15 Variation in trends assuming a fixed baseline



The application of these trends is designed to be applied on top of the baseline micro-component rates of change, so they do not double count.

2.4.2.1 AMP7 PCC targets

It might be the case, that a water company has made a commitment to achieving certain PCC target by the end of the current AMP. Although hitting this target is not guaranteed, it may be required that the forecast should account for this target and rebase the forecast from this value, in the given year.

This process is known as "target PCC rebasing" and is an option to include within the HHCF process. The way in which this is achieved is simply to introduce an AMP-specific trend, to ensure that the end-of-AMP PCC value matches the company target.

Following a discussion with Cambridge Water, there are currently no requirements to include this target PCC rebasing option, and so the baseline outputs will use the reference trends only, with no additional trends applied on top. Therefore, the results presented in section 0 are given with this in mind.

2.5 Weather modelling and peak factors

Task No.	MLR	MC
18	Compute dry year factors at required granularity	Compute normal year and dry year factors at required granularity
19	Select return period and peak factor duration	
20	Compute critical period factors per area/company, as required	

Household consumption is dependent on a range of variables such as practices, behaviours or attitudes that need to be accounted for in order to develop reliable forecasts. Weather has proven to be a driver of consumption and the inter-annual variation in consumption due to its effect needs to be understood and accounted for in water resources planning. Historic demand forecasting methods deal with this by:

- Analysing historic data to determine how annual average consumption differs between typical 'normal' and 'dry years'.
- Comparing this to recent actual consumption; and
- Producing factors or uplift volumes based on this comparison which are then applied to the consumption forecast.

This enables a suitable consumption value to be determined for the first year of the forecast, and production of dry year forecasts from this starting point. In WRMPs demand should be calculated for a range of planning scenarios:

- Normal Year Annual Average (NYAA). The demand in a typical "normal" weather year.
- Dry Year Annual Average (DYAA) - represents the dry weather demand that is compared with water available for use (WAFU) in the supply-demand calculations, and thereby is used to identify whether any dry year deficits occur. DYAA is defined as: "The level of demand, which is just equal to the maximum annual average, which can be met at any time without introducing demand restrictions. This should be based on continuation of current demand management policies."
- Peak demand scenarios – for example summer peak week (often known as critical period or CP).

The application of the NY and DY factors are slightly different. The normal year factor is typically generated from the base year (BY) to convert this into a "normal year" without any weather influence. Therefore, sometimes the terminology "BY to NY" is used. In contrast, the dry year factors are applied to the already weather corrected normal year outputs, so sometimes this is named "NY to DY".

2.5.1 *Normal year and dry year factors*

The methodology used in generating both the NY and DY factors comes from the UKWIR guidance report on household consumption forecasting, (UKWIR, 2015). This presents a range of methodology options for the calculation of these factors, namely:

- Trend analysis of demand
- Comparison of summer and winter consumption
- Weather demand modelling.

The selection of the specific methodology has been motivated by the data availability and granularity and resolution required for Cambridge Water.

Cambridge Water indicated at the start of this project that company level NY and DY factors would be required for the forecast, which sets the resolution of the weather modelling.

Based on the data available, which consisted of zonal/company level PCC data segmented by measured and unmeasured properties, as well as daily/monthly weather data from a single weather station, the "trend analysis of demand" method was used.

Additionally, it was decided at this point that the NY factors would be computed for measured and unmeasured properties separately, while the DY would be for all properties. This follows the same approach that was used in WRMP19.

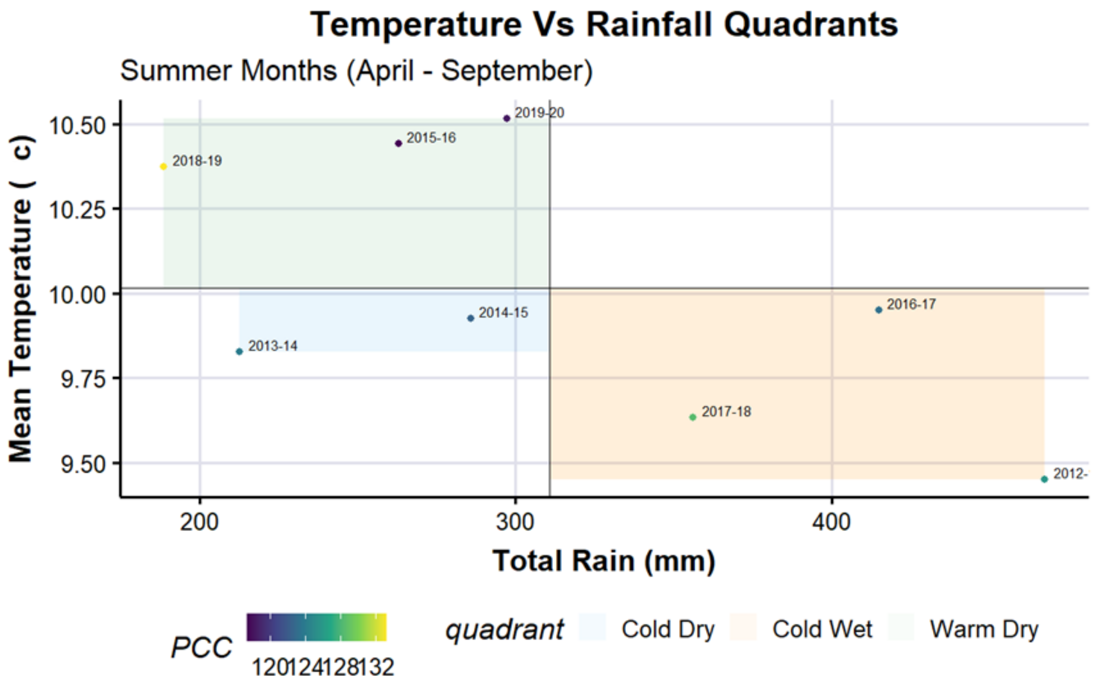
The normal and dry year factor calculation method follows the following process:

1. **Collation of the household demand data**, including mapping the PCC/PHC data to the weather data so that the weather variables can be compared with the resultant demand so that behaviours and patterns can be understood.
2. **Normalising the data**, where possible, to account for confounding factors such as meter penetration or water restrictions.
3. **Select dry years** using a rainfall-temperature quadrant plot which maps summer temperature to summer rainfall (April – September), coloured by the scale of consumption. This process is used to select the warmest and driest years with a large consumption increase as "dry years".
4. **Develop a regression model** to relate consumption with time. Using the outputs from the quadrant analysis, the dry years can be effectively removed from the trend line so that it does not affect the regression. From this, the actual consumption vs. predicted consumption can be assessed.
5. **Estimate the NY and DY factors** using the ratio between the predicted and actual consumption for the selected dry year (to generate the NY to DY factor), as well as the base year (to generate the BY to NY factor).

The first step of the process is to collate all of the household demand data. For Cambridge Water this was based on annual return data for PCC/PHC , as well as daily/monthly weather data including the variables temperature, rainfall and sunshine hours.

The most subjective part of the analysis is in the selection of the dry years using quadrant plots. An example of this plot is shown in Figure 16. The quadrants are divided along the mean lines of the weather variables. The candidate dry years are present in the top left-hand quadrant of the plot, though the final selection of the dry years is made only once consumption values are considered. In Figure 16, the year 2018-19 is the driest historic year, and it also has the brightest point, showing the scale of PCC. Therefore, 2018-19 would be selected as the dry year in this example.

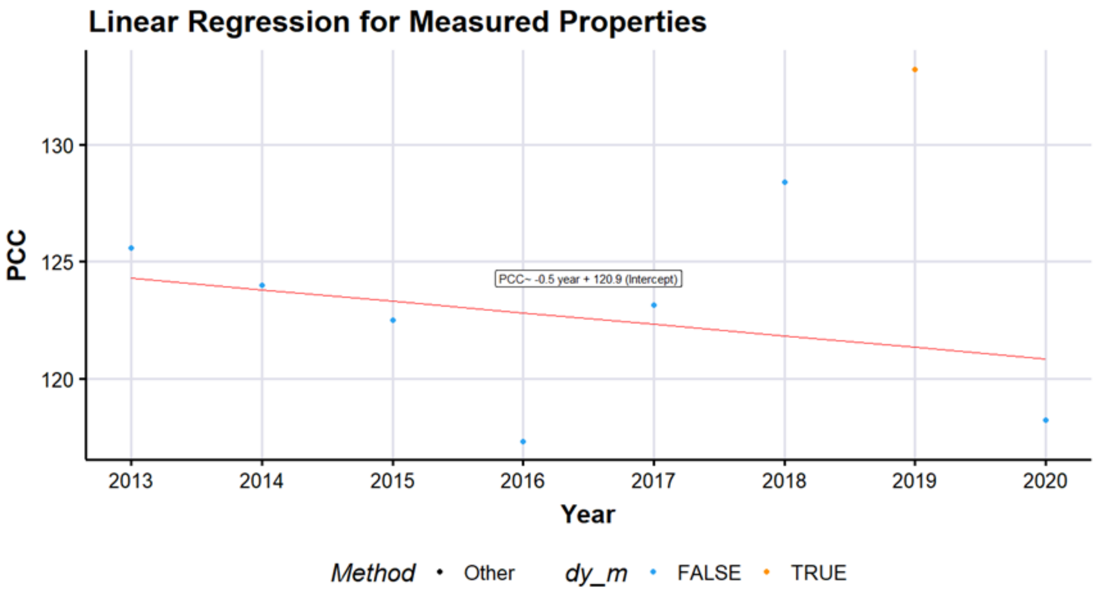
Figure 16 Example of a temperature/rainfall quadrant plot to select the dry years



The next stage is to create a linear regression between the historic PCC values, once the dry years have been removed. Where possible, this is done at meter status level, though this is not always possible.

Figure 17 shows an example of this linear regression. The blue points are years which have not been selected as "dry years", orange points are the selected "dry years". This process also allows account to be taken to different data collection methodologies. For example, with the new AMP7 consistency method, some companies have back-calculated PCC using the consistency method from 2017-18, but before this date the previous reporting method has been used. To account for any differences in consumption resulting from the methodology, this factor has been considered in the regression model.

Figure 17 Example of linear regression through PCC data



The following equations explain exactly how the NY and DY factors are computed.

1. First, simple linear regression using annual PCC values for measured and unmeasured households is computed.
   a. Slope

$$\alpha = \frac{n\sum(xy) - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

   b. Intercept

$$\beta = \frac{\sum y - \alpha \sum x}{n}$$

   c. Trend line

$$y = \alpha x + \beta$$

Where y represents all consumption records, excluding those in the dry year, and x is years.

1. BY to NY factor (NY factor):

$$BY\ to\ NY = \frac{predicted\ PCC\ in\ BY}{actual\ PCC\ in\ BY}$$

2. NY to DY factors (DY factor):

$$NY\ to\ DY = \frac{actual\ PCC\ in\ DY}{predicted\ PCC\ in\ DY}$$

The results of this analysis for Cambridge Water are presented in section 3.2.

2.5.2 Critical period calculation

As well as the normal year and dry year factors, water companies also consider a "critical period" planning scenario, in which water resource zone supply-demand balances are at their most constrained.

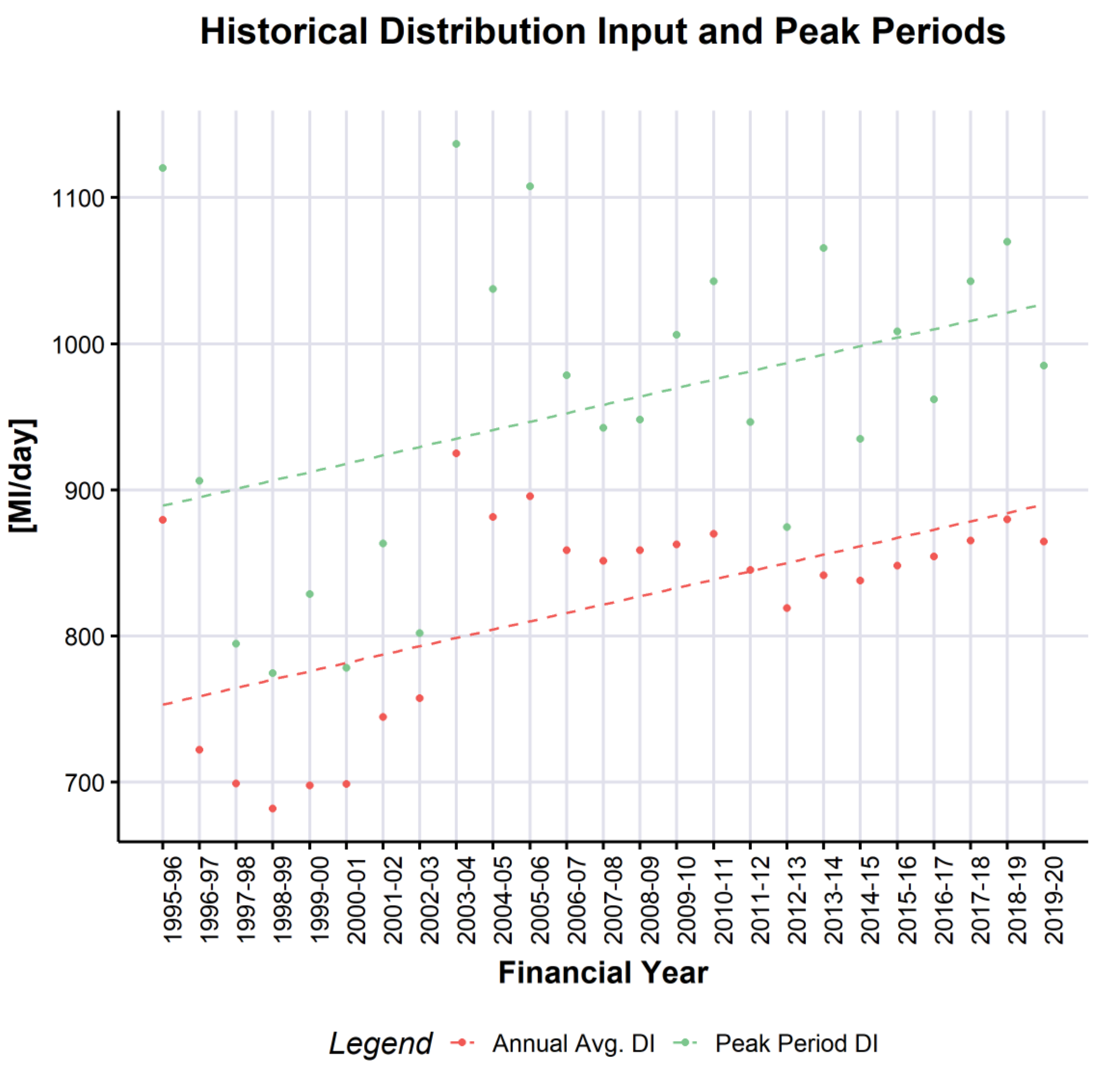
The method for computing these factors follows the UKWIR, Peak Demand Forecasting Methodology report, 06/WR/01/7 (UKWIR, 2006) and has the following steps:

1. **Data collection**. This includes distribution input data (DI) as a fine a resolution as possible.
2. **Determination of the peak period**. This is specific to Cambridge Water, but the recommendation is not to use a period of any less than one week.
3. **Disaggregation.** Where possible, it is preferable to remove the non-household demand and leakage from the DI data. However, this is not always possible and caution should be taken if disaggregation cannot occur.
4. **Rebasing and normalisation**. The aim of this task is to estimate the peak demand which would be experienced if the same conditions were to recur in the base year. This can be carried out using one of three measures of peak demand.
    a. Peaking factors: where changes to peak demand are linked to changes in annual average (e.g. change in number of customers rather than their characteristics)
    b. Peak volumes: where peak demand is related to activities which are independent of average demand change (e.g. tourism) or are considered to be a stable demand characteristic for each customer of a particular type (e.g. garden watering for each property with a garden)
    c. Absolute peak demand: where it is difficult to disaggregate reliably; demand characteristics and customer base are believed to have been relatively stable.

    For this project, we have considered either peak factors or peak volumes, which means that a normalization is required. The method for normalization should represent average demand and so could include using a long-term average or rolling average. For Cambridge Water we have used a rolling average methodology as this accounts for non-linear relationships in the historic data, that a long-term average will not do.
5. **Return period analysis**. Once the historical demand is normalised, the peak events can be compared. This allows companies to improve their understanding of the level of service that planning for a specific peak demand provides by assigning a probability to peak demands of different magnitudes.
    The method used here, is using fitted cumulative distribution functions (CDFs) to the normalised peak factors and/or peak volumes.
6. **Forecasting**. Finally, using the required return period, the critical period factor or critical period volume is determined using the probability from the fitted CDF applied to the factors and volumes, respectively.

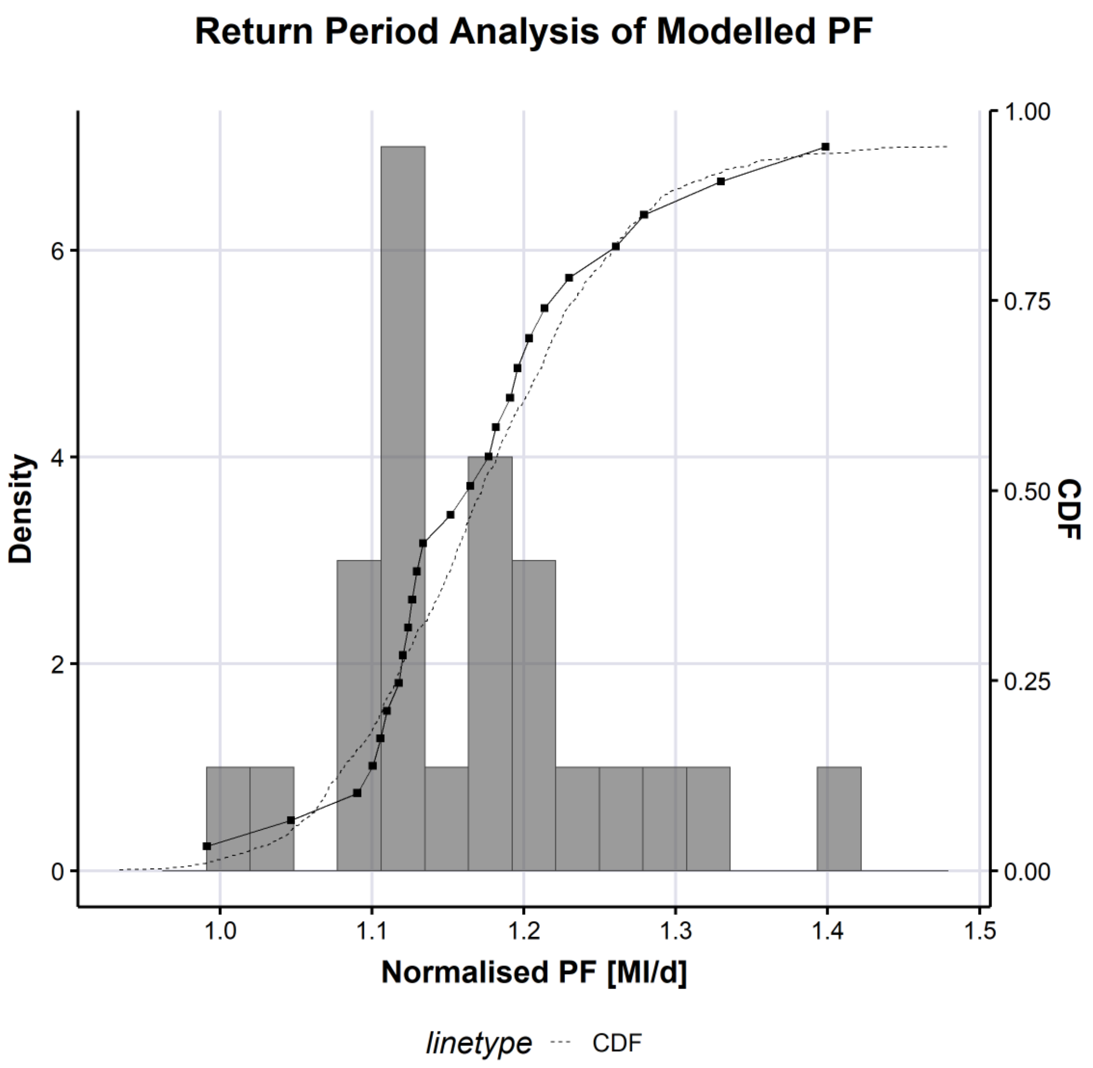
To illustrate some of these steps in more detail, Figure 18 provides a long-term plot of DI data, which has also had its peak period DI plotted in green. This is before the rebasing and normalisation process.

Figure 18 Example of historic DI data including peak period of 7 days



Next, Figure 19 shows an example of fitting the cumulative distribution function to the normalised and rebased peak factors. The fitted distribution is given as the dotted line, whereas the actual distribution is shown as the black squares joined by a solid line.

Figure 19 Example of return period analysis using peak factors



The method described in this section, has been applied to Cambridge Water's data using the selections below:

- Return period: 1 in 200
- Peak duration: 7 days
- Method: Peak factors

The results are presented in section 3.2.

2.6 Scenarios, climate change and uncertainty

Task No.	MLR	MC
21	Collate outputs to company level	
22	Apply climate change factors	
23	Undertake uncertainty analysis	
24	Run appropriate steps from 5-23 again, for any agreed scenarios to be tested	

Now that the HHCF model has been built, the POPROC data segmented and the weather modelling complete, the final stage is to apply the climate change adjustments, before running different scenarios and uncertainties.

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 household demand forecasts we separate the concepts through definitions:

- **Uncertainty** refers primarily to the variability we have in the forecasts due to data uncertainty and unexplainable variability uncertainty. Uncertainty is non-zero, even in the present, 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 being considered. Scenarios are usually represented by a discrete number of alternative forecasts.

We first discuss the method for applying the climate change factors.

2.6.1 Climate change

The household consumption forecasting guidance describes the requirement that all HHCFs should be provided with and without the addition of climate change impacts. To achieve this, we have used the methods and models provided in the UKWIR report, "Impact of climate change on water demand", (UKWIR, 2013). The aim of this project was to provide climate change demand factors to account for the impact of climate change to be used in the WRMP process.

More specifically, this report contains demand factors for each UKCP09 river basin, describing the percentage change in household demand for two case study relationships, Severn Trent and Thames, and three demand criteria (annual average, minimum deployable output and critical period). The demand factors are given for the 10th, 25th, 50th, 75th and 90th percentile to reflect the uncertainty in the climate projections.

The values provided as part of this project have been used to define the climate change factors for Cambridge Water.

The first step is to select the correct model for use. Based on proximity, the selected model for Cambridge Water is Anglian Water. The default percentiles selected are the 50th percentile, with the *annual average* values used for the normal year (NYAA) and dry year (DYAA) demand criteria, and *critical period* values being used for the peak demand (critical) demand criteria.

The selection of the correct river basin for Cambridge Water is the final step in determining the correct climate change factors. This selection has been made using the geographical distance between Cambridge Water and the river basin options and is shown in Table 6.

Table 6 Climate change factors and river basin selected for Cambridge Water

Area	Planning scenario	Company climate change figure	Climate change percentile	River basin	River Basin coverage	River basin climate change figures
CW	NYAA	0.75	p50	Anglian	100%	0.75
CW	DYAA	0.75	p50	Anglian	100%	0.75
CW	CP	2.05	p50	Anglian	100%	2.05

Once the climate change factors are selected, the final step is to generate the values by year. This is achieved by linearly interpolating the values from the base year point of zero, to the final climate change factor in Table 6 for 2045, and continuing this trend until the final year of the forecast.

2.6.2 Scenarios

As described at the start of this section, scenarios are defined as the variability in future projections due to foreseeable events. These are typically due to different growth forecasts in the POPROC data, or changes to the metering strategy (i.e. rates of optants or compulsory metering).

At the start of this project, discussions were had with Cambridge Water to determine which scenarios would be delivered in addition to the baseline forecast. Table 7 provides a summary of this information, specifically giving the growth forecast name, and metering strategy information for these scenario runs, as well as the same information for the baseline forecast.

Table 7 List of the different scenarios tested as part of this project

	Growth scenario	Optant metering strategy	Compulsory metering strategy
Baseline scenario	Housing Plan P BY-rebase	Std	
Scenario 1	ONS-18-Low-L	Std	
Scenario 2	oxcam – 1b-r-p	Std	
Scenario 3	oxcam – 2b-r-h	Std	

The results of these forecast outputs will be presented in section 3.5 of this report.

2.6.3 Uncertainty

In this context, the estimated uncertainty represents the variability within a given, foreseeable scenario. For each scenario, the uncertainty can be estimated and will be represented as buffer intervals around the central forecast, usually represented by quantiles (e.g. between the 5th and the 95th quantile or between the 25th and the 75th quantile). It is important to consider that the distributions of total consumption, PHC and PCC are unlikely to be symmetric, therefore the upper and lower thresholds of the buffer intervals may have a different distance from the central forecasts. Additionally, this means that the deterministic forecast may not correspond to the mean of the distribution.

Modelling the household demand uncertainty is a process that can be divided into three phases:

1. **Input uncertainty estimation**: as the household demand forecasts are estimated using a complex MLR or MC model that has a large number of inputs and parameters to consider, the uncertainty of the model results will depend on the uncertainty of the model inputs. So, we need to define how uncertain each of the inputs is and represent this uncertainty through probability distributions. In this context, we include the model uncertainty among the input uncertainties.
2. **Uncertainty propagation**: once all the input uncertainties are defined, we need to understand how they interact to define the resulting output uncertainty. For very simple models this can be attempted mathematically, but it is not the case for the household demand models which are made by many steps beyond the core of the model application. Therefore, we follow the guidelines and use an empirical approach using a Monte Carlo propagation. To improve the efficiency and reduce the number of samples, we opt for a Latin Hypercube Sampling (LHS) for the Monte Carlo.
3. **Output uncertainty summary**: using a Monte Carlo approach results in having a large number of possible alternative outputs. From these, we can derive the output probability distribution and summary statistics that represent the output uncertainty.

2.6.3.1 Input uncertainty estimation

Estimating the uncertainty on the inputs requires probability distributions to be defined for each of the model elements. These are:

- **Data:**
    - Annual Return (AR) data
    - Historic POPROC
    - Forecast POPROC
    - NY/DY factors
    - Peak factors
    - Climate change coefficient
    - MC trends
    - OVF values

- **Models:**
    - MC model
    - MC modelling assumptions
    - Residual model
    - Trend modelling

To simplify the process, the following assumptions are made:

- The uncertainty on past data is negligible compared to the uncertainty on future data.
- The uncertainty on residual models counterbalances the reduction on the main model introduced by using the residual model (the residual model is designed to improve the estimates of the main MC model).
- The uncertainty of the trend modelling is reflected in the uncertainty defined on the trends themselves.

Therefore, we evaluate the uncertainty on the following elements:

- Forecast POPROC
- NY/DY factors
- Peak factors
- Climate change coefficient
- MC trends
- OVF values
- MC model

Population, Properties and Occupancy (POPROC)

The uncertainty on population and properties is defined by the UKWIR guidelines (UKWIR and EA, 2015), while the occupancy is a derived value. The report indicates that a normal distribution should be used, and for each year an RMSE value is provided (Table 8 of the report) to be used as standard deviation. The mean is centred in the deterministic value. We also consider the uncertainty on the meter penetration, using the same definition.

Model

The way that the model uncertainty is defined depends on the type of model.

For MC models, the uncertainty is defined on each micro-component ownership, volume, and frequency, for each cohort. Where possible the distributions were estimated from previous studies; where the data was not available or applicable, distribution were estimated based on expert judgement and known limits. Some of the micro-components' ownership, volume and frequency values are not fixed, they are derived as a function of occupancy. In that case uncertainty is applied to the linear model factors. The selected distributions are normal, truncated normal, gamma and beta, depending on the known limits for each parameter.

Additionally, a truncated normal distribution is considered for the compulsory saving parameter, which defines how much water consumption is reduced when a property passes from unmeasured to measured.

Normal Year (NY), Dry Year (DY) and Critical Period (CP) factors

The NY, DY and CP factors are correction factors that rebase the forecasts to simulate a typical year, a dry year or a critically dry year. These are three real numbers, and their uncertainty can be modelled as:

- **NY**: a normal distribution centred by the deterministic value.
- **DY**: a truncated normal distribution centred by the deterministic value, limited by NY on the lower side.
- **CP**: a normal distribution centred by the deterministic value.

Although there are no other theoretical constraints, it is possible to use truncated normal distributions to avoid values that are unrealistically high or too low.

Trends

Trends can either be calculated from the time series or known realistic trends can be used. Either way, the uncertainty on the trends can increase in time and can be defined with a normal distribution centred by the deterministic value.

Climate Change

Climate change is modelled with an additional trend correction. In Artesia's model, this is represented by a linear trend, starting at zero and growing, quantified by the value it assumes in 2040. The 2040 value is derived from UKWIR guidelines (UKWIR, 2013) that reports probabilistic trend values given in Appendix 6. The values vary whether we consider an annual average (normal or dry year), or a critical period.

The UKWIR report describes the probabilistic nature of the climate change coefficients through percentiles. Observing the percentiles, they come from an almost uniform distribution, and we can extrapolate the extremes of the distribution from the given percentiles.

2.6.3.2 Uncertainty propagation

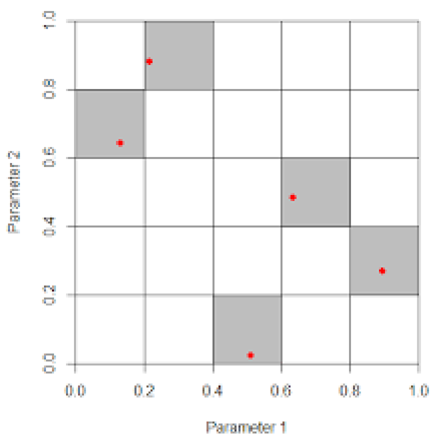
Given the complexity of the models used to estimate household demands, the guidelines (UKWIR, 2002) recommend using a Monte Carlo approach. The model needs to be run multiple times, each time using a different value of the uncertain inputs, drawn from the distributions defined in the previous section.

Traditionally, a Monte Carlo approach is applied by randomly sampling from the input probability distributions. This requires a large number of samples to define the output probability distribution with an acceptable accuracy, usually in the order of magnitude of 1000, requiring long computational times.

In this case we use the Latin Hypercube Sampling (LHS) technique, which is more optimised and requires a much smaller number of samples.

A Latin square is a square grid where there is only one sample in each row and each column, shown in Figure 20. Each dimension represents a parameter we need to sample from.

Figure 20 Latin square example



A Latin hypercube is the generalisation of this concept to an arbitrary number of dimensions, and therefore of parameters/variables we need to sample from, whereby each sample is the only one in each axis-aligned hyperplane containing it. This sampling technique covers the whole sampling domain with a lower number of samples. Published theoretical results (Aistleitner, Hofer, & Tichy, 2012) show that the sampling error of a random Monte Carlo sampling is $O\left(\frac{1}{\sqrt{N}}\right)$ whereas the sampling error for LHS is $O\left(\frac{1}{N}\right)$, quadratically faster for almost all distributions and statistics in common use. In simpler words, using an LHS you need square root the number of samples you would need in random sampling.

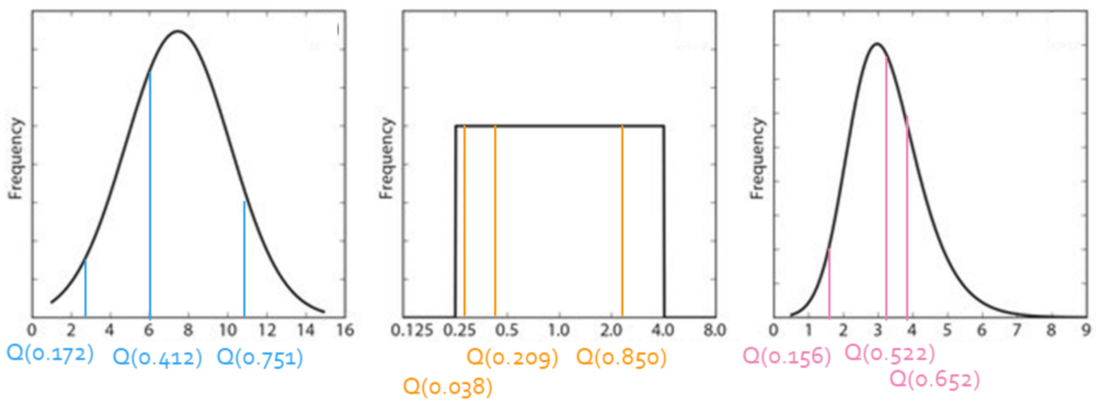
Operationally, if a random Monte Carlo sampling requires 1000 samples, the LHS can reach the same accuracy with approximately 32 samples.

In practice, an LHS samples a number x of near-random values from uniform distributions, between 0 and 1, knowing a priori how many parameters need to be sampled. Sampling from a uniform distribution can be converted to any different distribution using corresponding quantiles.

Figure 21 shows an example of sampling from different distributions using LHS.

Figure 21 Example of sampling from three different distributions using LHS with 3 samples

	param1	param2	param3
	<dbl>	<dbl>	<dbl>
1	0.412	0.209	0.522
2	0.751	0.0383	0.156
3	0.172	0.850	0.652



Once samples from the LHS are drawn from all the input parameters/variables' distributions, the model can be run multiple times to obtain multiple outputs.

2.6.3.3 Output uncertainty summary

The multiple outputs (each including estimates of Ml/d, PHC and PCC in time and for different areas/cohorts) represent an empirical probability distribution of the output. To interpret these values quantitatively, the distribution can be represented with percentiles and other summary statistics. We have used the following:

- Mean
- 10th percentile
- 25th percentile
- 50th percentile (Median)
- 75th percentile

- 90^{th} percentile
- 95^{th} percentile

As the distributions are likely to be asymmetric, it is not recommended to use the standard deviation or the variance to represent the distribution spread, as these statistics imply a symmetry in the distribution. Additionally, the median and the mean are likely to be different.

Note that the relationships between total consumption, PHC and PCC will not hold when comparing the percentiles. For example, dividing the 90^{th} percentile of total consumption by the 90^{th} percentile of number of properties will correspond in a relatively average PHC value, not the 90^{th} percentile.

2.7 Baseline forecast outputs

Task No.	MLR	MC
25	Micro-component outputs and EA table	
26	Output forecast in a format specific to original requirements	
27	Audit reporting	

The complete modelling process has now been completely described, with the only remaining step being putting all of the steps together, applying a company level collation and producing outputs suitable for the Environment Agency (EA), NRW and UKWIR templates and guidelines.

The method for separating the outputs into the macro-components specified by the EA is simply based upon combining the micro-components into the following categories based on a simple ratio approach.

- Toilet flushing
- Personal washing
- Clothes washing
- Dishwashing
- Miscellaneous internal use
- External use

2.7.1 *Baseline forecast selections*

In the interest of clarity, we now summarise the selections of each of the HHCF stages used in the generation of the baseline forecast. This is given in Table 8 and have been used in the results given in section 3 below.

Table 8 Baseline household consumption forecast selections in the framework

Metric	Format	Report reference
Base year for the forecast	2019-20	Section 2.1
Final year of the forecast	2099-00	Section 2.1
Forecast granularity	WRZ level	Section 2.1
POPROC rebasing option	Base year rebase	Section 2.2.2

Trend selected	Central	Section 2.4.2
Peak duration	7 days	Section 2.5.2
Return period for peak analysis	1 in 200	Section 2.5.2
Target PCC (l/head/day)	-	Section 2.4.2.1
Target year	-	Section 2.4.2.1
DY grouping	All properties	Section 2.5.1
NY grouping	Measured and unmeasured	Section 2.5.1
NY/DY resolution	WRZ level	Section 2.5
Baseline metering strategy name	Standard	Section 2.6.2
Baseline growth forecast name	Housing-plan-p	Section 2.6.2
NY climate change figure for 2045	As per Table 6	Section 2.6.1
DY climate change figure for 2045	As per Table 6	Section 2.6.1
CP climate change figure for 2045	As per Table 6	Section 2.6.1
MC compulsory saving (from unmeasured)	10%	Section 2.3.2

3 Results

The following section presents the results of applying the full HHCF methodology as per the framework. Unless explicitly stated, the outputs have been generated according to the selections presented in Table 8.

3.1 Population and property forecasts

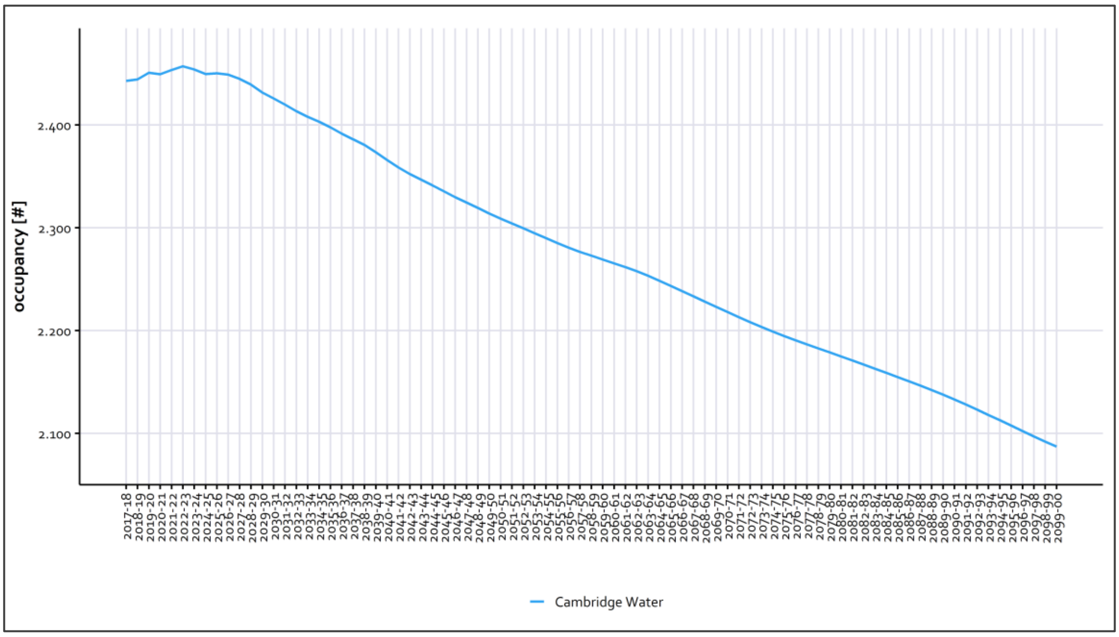
We first start with the population and property forecasts for the baseline scenario, generated as per the method given in section 2.2. A base year rebase was applied to the forecasts provided by Edge as part of the forecasts for the Water Resources East (WRE) region. This means that the population and property forecasts are calibrated to the annual return values for 2019/20 and then are extrapolated to meet the end point forecast by Edge in 2100.

The Edge forecasts for population and properties for 2019/20 were 322,895 and 143,258 respectively. These were calibrated to the annual return value of 326,922 and 133,404.

Figure 22 shows the input occupancy data from the forecast provided by Edge for the Cambridge Water region. This shows that the overall occupancy initially increases and then is steadily decreasing over the forecast period. These forecasts will be influenced by national and regional population and property forecasts, and for this region will take account of the specific forecasts for the Cambridge Water region.

The company occupancy values for all properties reduces from 2.44 to 2.02.

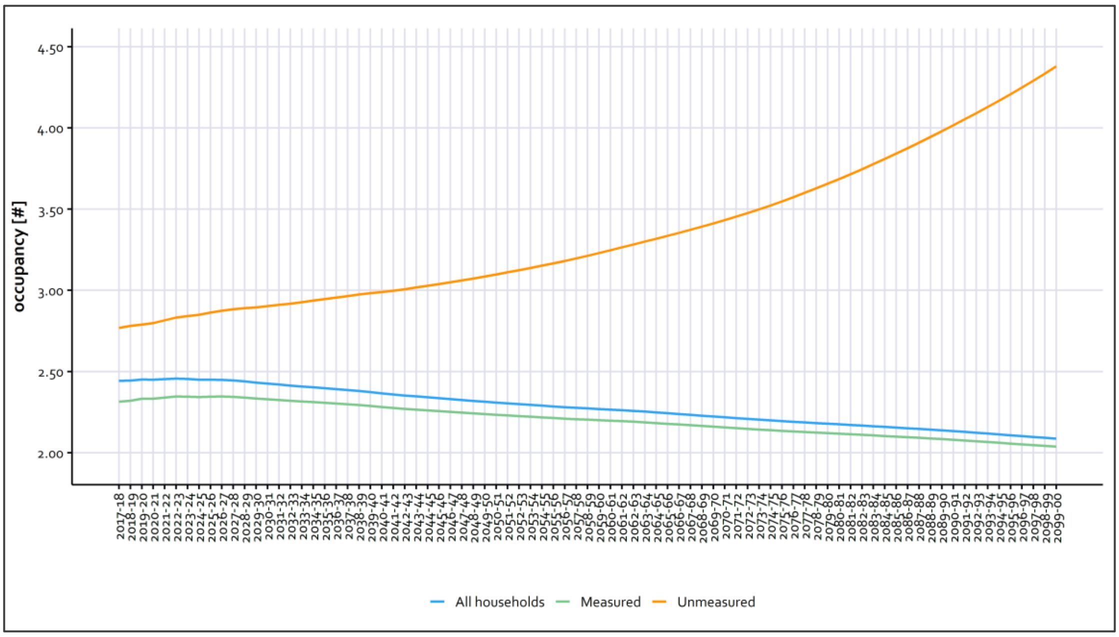
Figure 22 Occupancy forecast for Cambridge Water



The next stage is to separate these into the individual meter cohorts.

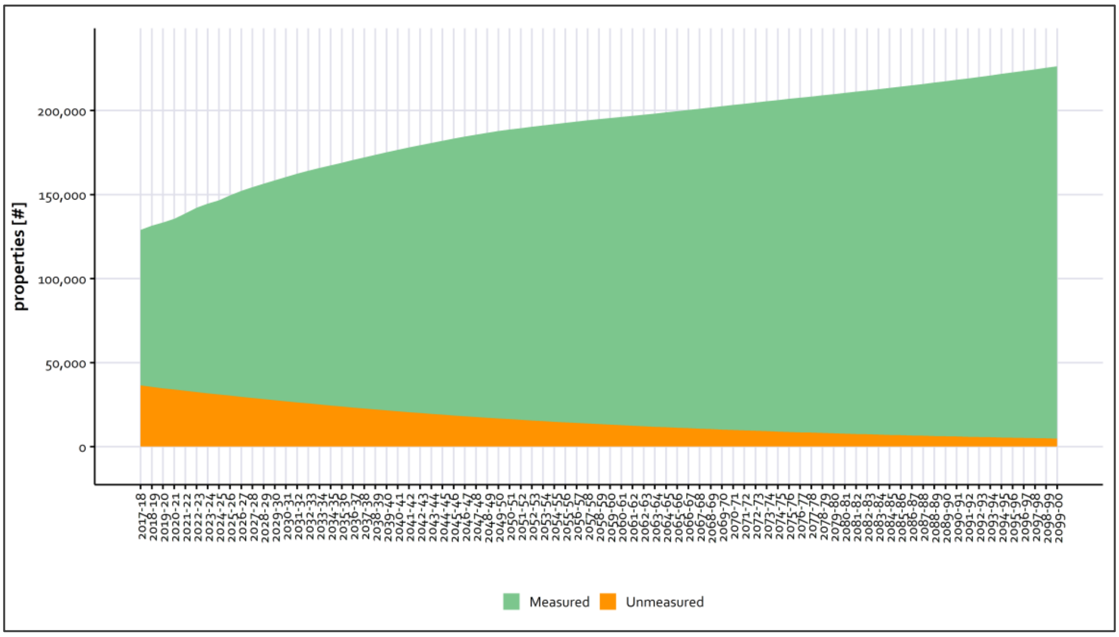
Figure 23 shows how the occupancy values have been separated into the meter status categories, unmeasured, measured and all. We can see that measured occupancy is much lower than unmeasured, which is what we would typically expect based on the measured category comprising of optant properties who have chosen to have a meter usually due to reduced household occupants. As the forecast extends, we see that unmeasured occupancy increases rapidly. This is expected. As more properties move from the unmeasured housing pool to measured through the free meter optant programme, it is natural that the properties that move have a lower occupancy, causing the average occupancy of the unmeasured group to steadily rise.

Figure 23 Occupancy forecast for Cambridge Water split by meter status



Finally, the property forecast for Cambridge Water shows that the number of properties from the base year of 2019-20 to the final year in 2100 increases from 128,981 to 226,278, an increase of 75.44%. The meter penetration changes from 72% to 91%. This is shown in Figure 24.

Figure 24 Property forecast for Cambridge Water split by meter status



3.2 NY, DY and CP factors

Before presenting the baseline consumption forecast outputs, we present the DY and CP factors used within the analysis.

Table 9 Final NY, DY and CP factors

Area	Meter status	NY	DY	CP
Cambridge Water	Measured	1.022	1.049	1.278
Cambridge Water	Unmeasured	1.066	1.049	1.278

For comparison, the WRMP19 factors were:

- Normal year – 1.026 for measured and 0.946 for unmeasured,
- Dry year - 1.045,
- Critical period 1.224.

3.3 Baseline household consumption forecast results

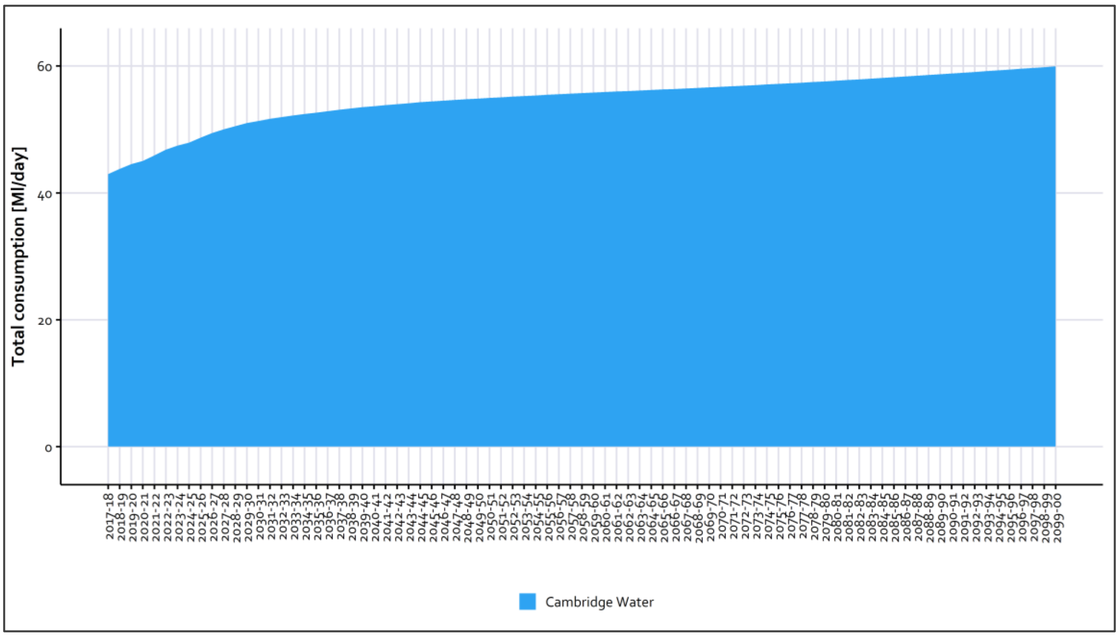
The following outputs have been generated for the baseline scenario. The plots that follow are based upon the normal year planning scenario. The DY and CP scenarios are achieved after applying the simple uplift factors given in **Error! Reference source not found.** so are not scrutinised in any great detail.

Finally, the upcoming plots are all inclusive of climate change, unless explicitly stated.

We first look at total consumption across the planning period, expressed in megalitres per day (Ml/d). Figure 25 shows total Ml/d coloured by the influence from each water resource zone. As the Integrated zone is the largest, this dominates the plot.

This shows total consumption start from 42.95 Ml/d, rising to 59.94 Ml/d, an increase of 16.99 Ml/d. From the plot, we can see an increase in consumption which is relatively steep for the first 30 years of the forecast, and starts to increase more slowly from 2050. This is simply because the trends applied are only applied until 2050 and held flat after this date. This is because the uncertainty of any long-term trends increased markedly and thus, they cannot be extrapolated indefinitely. From 2050, the rise in consumption is completely driven by the property forecast.

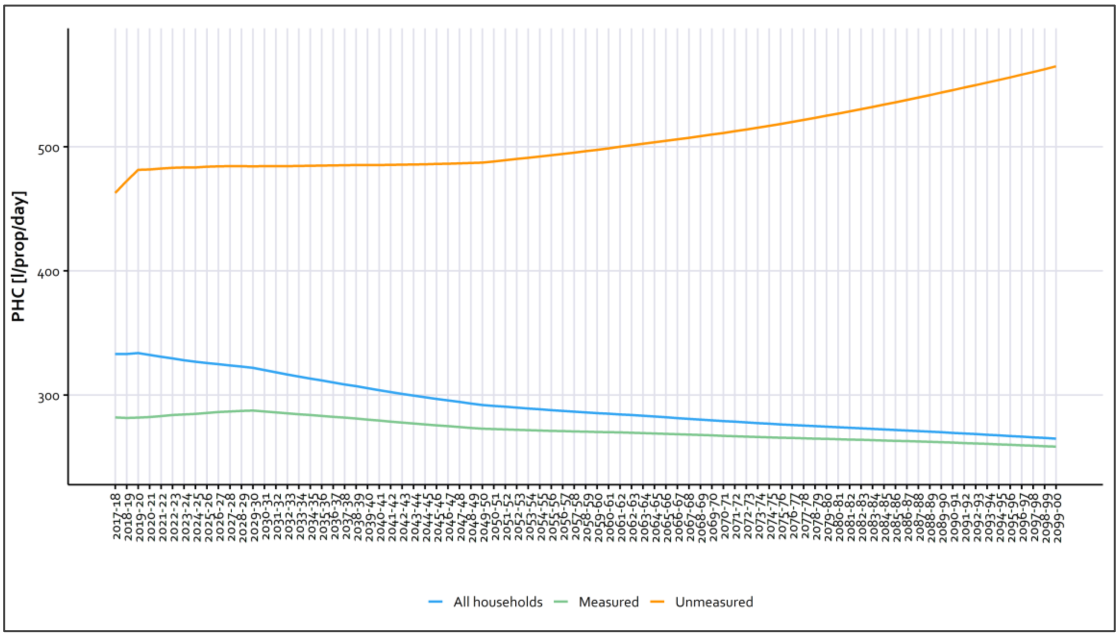
Figure 25 Total consumption (Ml/d) at company level across the forecast period



Splitting this into household level consumption outputs, Figure 26 provides PHC values for all, metered and unmetered properties. There is an uptick in unmeasured PHC in the first three years from 2017/18, which are reported values that have been normalised and uplifted by the dry year factor. The forecast values, which start in 2019/20 show unmeasured PHC at a steady level before increasing about halfway through the planning period.As expected, household consumption is declining for measured and "all" properties, as the increasing unmeasured PHC is due to the lower consumption properties moving from unmeasured into optant (measured) groups.

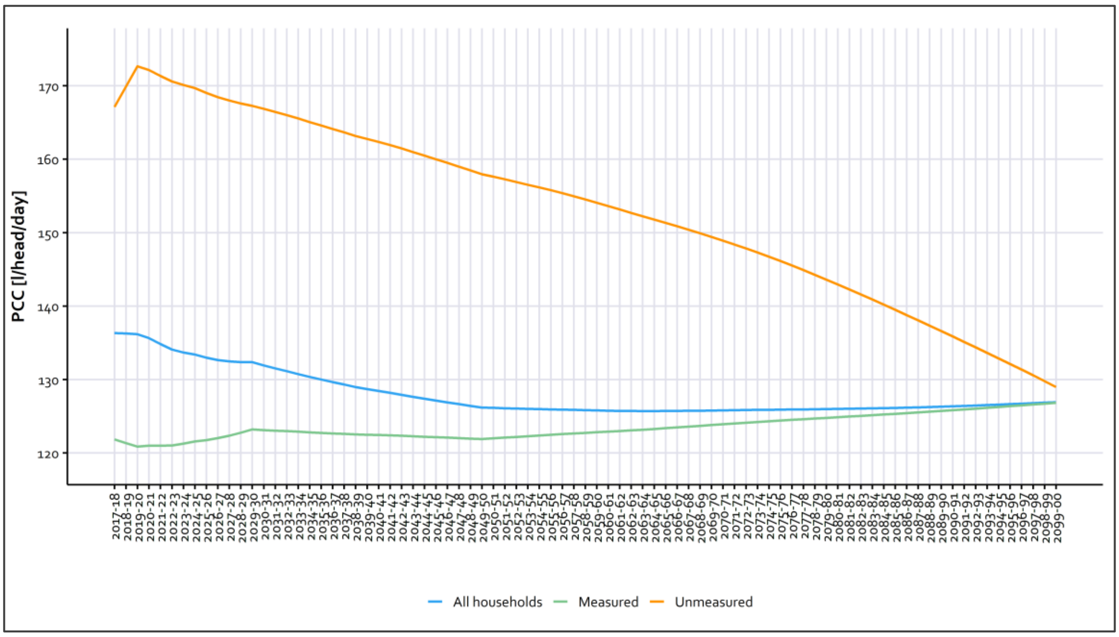
Total PHC reduces from 332.99 l/prop/day to 264.88 l/prop/day in 2100

Figure 26 Company level PHC (l/prop/day) by meter status



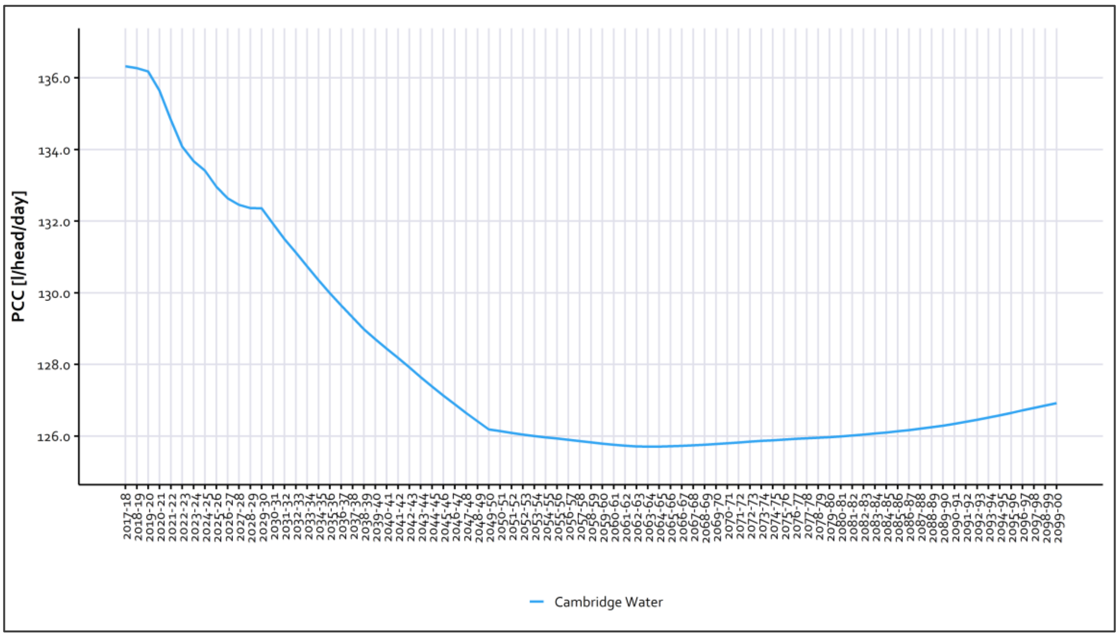
Finally, if we look at the results at PCC level in Figure 27 and Figure 28, we first see a downward trend in unmeasured PCC, after the first three years of reported values (which have been normlaised then uplifted as described for PHC). This is driven by the steady increase in occupancy. We would not expect the household consumption to increase linearly with occupancy, and so this results in PCC declining.

Figure 27 Company level PCC (l/head/day) by meter status



The trend observed using the zonal PCC values in Figure 28 is entirely driven by occupancy and the micro-component trend. First PCC falls due to the micro-component trend and the optant rate. At the same time, the occupancy steadily decreases, which counteracts the falling PCC due to micro-component trends. Once the micro-component trend is kept flat after 2050, the PCC increases as the occupancy keeps on dropping. It should be noted that Figure 28 is a "zoomed in" version of the blue line in Figure 27, so the pattern of PCC in Figure 28 looks more extreme than it otherwise might.

Figure 28 PCC (l/prop/day) for all households



3.3.1 Conclusions

Over the planning period of 2019-20 to 2100, total consumption for Cambridge Water increased by 39.55% to 59.94 Ml/d. This is considering a property increase of 75.44% over the same period.

In contrast, total PHC decreased by 20.45% over the forecast period and PCC showing a smaller decrease of 6.9%. The reason for this disparity is due to decreasing occupancy . If occupancy is forecast to decrease, then per household consumption will be more greatly affected than PCC, as the relationship between the two variables is not linear.

Table 10 summarises all of this information, for the company and for each zone. Note that these values are for the normal year, with climate change applied.

Table 10 Summary of the baseline HHCF outputs

Area	Metric	2019-20 base year	2100 final year	Percentage change
Cambridge Water	Total population	315,060	472,255	49.89%
	Total properties	128,981	226,278	75.44%
	Total consumption (Ml/d)	42.95	59.94	39.55%
	Total PHC (l/prop/hr)	332.99	264.88	-20.45%
	Total PCC (l/head/day)	136.32	126.92	-6.90%

3.4 Baseline uncertainty

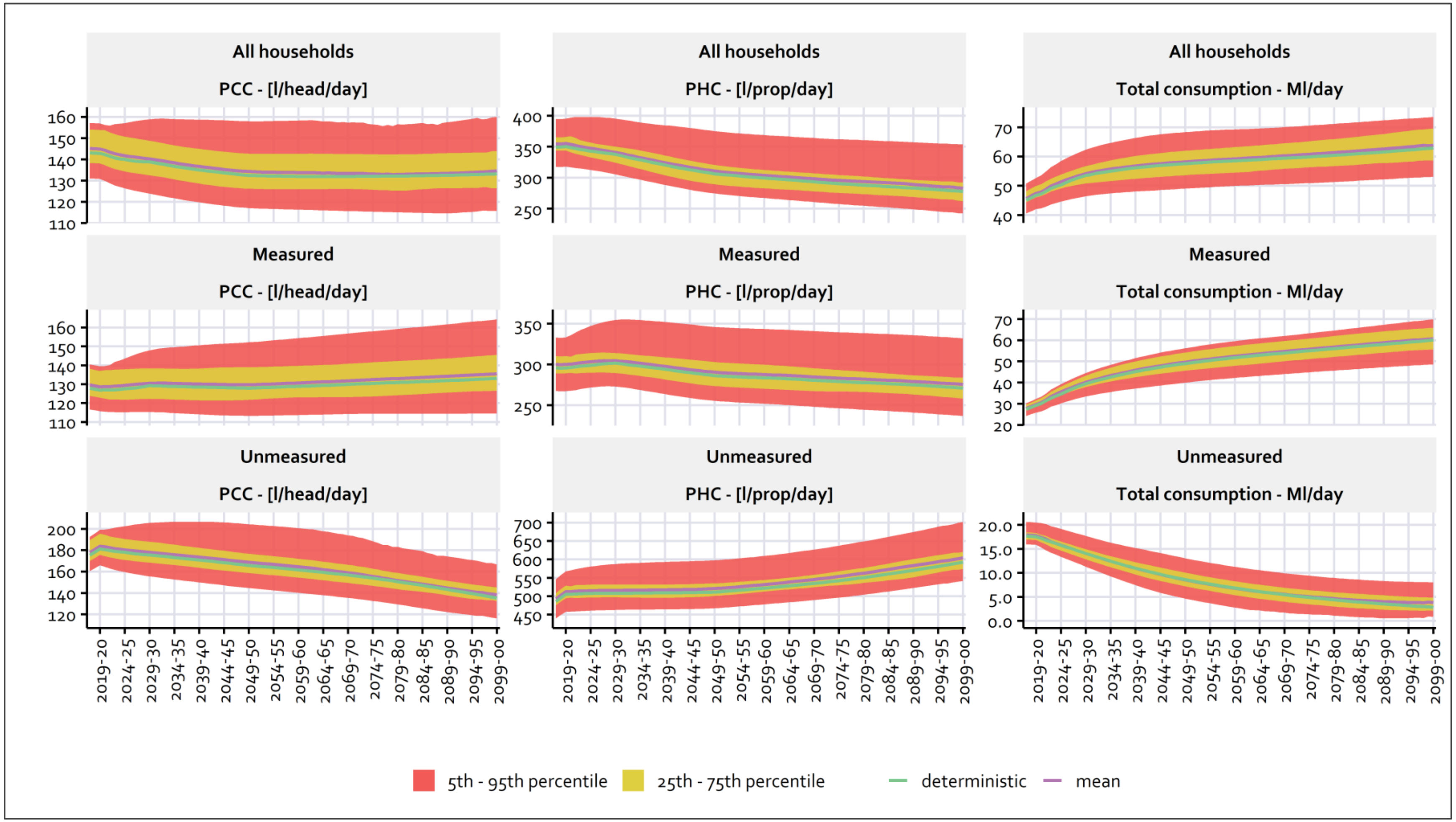
The baseline uncertainty for the dry year annual average is shown in Figure 29. The plots present the deterministic forecast, which is the is the value we get from the normal HHCF model, the mean forecast, which is the average of the uncertainty analysis, and percentile uncertainties, for a range of forecast variables.

This ranges presented in Figure 29 represent the uncertainty associated with the following model inputs:

- Forecast POPROC
- NY/DY factors
- Peak factors
- Climate change coefficient
- MC trends
- OVF values
- MC model

The total consumption results can therefore be used as an input for the D2 – demand forecast uncertainty component of target headroom.

Figure 29 Uncertainty plots for dry year annual average demand variables



3.5 Scenarios

We now present the outputs generated for the scenarios given in Table 7. As this provides different growth and metering forecasts, the other selections given in Table 8 are implied, unless specified otherwise.

Table 11 summarises the outputs to provide the high-level outputs in conjunction with the separately issued tables and plots.

Overall, the 2100 consumption (Ml/d) values using the different scenarios vary from a minimum of 39.04 Ml/d to 75.36 Ml/d. The former, represents a 3.9% reduction in baseline consumption, even though the properties are forecast to increase by 13.96% in this scenario.

For PHC, the range of outputs varies between 257.21 l/prop/day to 265.63 l/prop/day, showing that this metric is relatively stable.

For PCC, the range of outputs varies between 127.59 l/head/day to 128.64 l/head/day.

Table 11 Summary of scenario outputs for the company, NY with climate change

Scenario*	Company level metrics	2019-20 Base year	2100 final year	Percentage change
1	Total population	315,060	303,268	-3.74%
	Total properties	128,981	146,982	13.96%
	Total consumption (Ml/d)	42.95	39.04	-9.09%
	Total PHC (l/prop/hr)	332.99	265.63	-20.23%
	Total PCC (l/head/day)	136.32	128.64	-5.63%
2	Total population	315,060	491,944	56.14%
	Total properties	128,981	245,241	90.14%
	Total consumption (Ml/d)	42.95	63.08	46.87%
	Total PHC (l/prop/hr)	332.99	257.21	-22.76%
	Total PCC (l/head/day)	136.32	128.22	-5.94%
3	Total population	315,060	590,674	87.48%
	Total properties	128,981	292,496	126.77%
	Total consumption (Ml/d)	42.95	75.36	75.47%
	Total PHC (l/prop/hr)	332.99	257.65	-22.62%
	Total PCC (l/head/day)	136.32	127.59	-6.41%

*1 ONS-18-Low-L, 2 oxcam – 1b-r-p, 3 oxcam – 2b-r-h

4 Conclusions

Water companies in England and Wales have a statutory duty to develop Water Resource Management Plans (WRMPs) under the Water Industry Act 1991. Forecasting the demand for water is a key element of this plan, and household demand is, in turn a significant part of overall demand.

Companies are now working in a more extensive and co-ordinated way within the context of regional plans, which have been implemented across England in the run up to the next round of WRMPs, to be published in 2024 (WRMP24). Regional plans have been implemented to improve resilience and environmental protection, and to better understand how resources may be shared between companies.

This report sets out the initial development of household demand forecasts for Cambridge Water (CAM) for the Water Resources South East regional plan. This household demand forecast has been developed within the context of regulatory requirements and technical guidance. In addition, for this round of plans, Artesia has developed an updated and improved modelling framework which sets out the detailed steps required to develop the household demand forecast.

The forecast set out in this report has been developed based on micro-component modelling methods, which model household water use based on estimates of specific water using activities within the home. This is a well-established and extensively used approach to modelling and forecasting household water demand. This method is suitable for water resource zones with a low level of water resource planning concern.

This report describes the steps involved in producing a micro-component-based household demand forecast. A key step is to split population and property forecasts into metered segments, including unmeasured, existing measured, compulsory measured, optants and new properties. Assumptions are made about these segments in order to ensure consistency within and between the segments for key variables such as household occupancy. Calibration ensures consistency with zonal population, property and occupancy totals. These values are then rebased in an agreed way to match the base year values.

Micro-component modelling uses the most recent available data on micro-component use and occupancy to determine statistically significant relationships between these variables. A linear model has been developed for toilets, showers, baths, washing machines and taps based on this analysis. Trends are then added to the model to reflect likely technology developments, and to explore scenarios associated with these, over the planning period.

Weather modelling is then used to derive normal year, dry year, and (where needed) critical period factors. Scenarios have then been produced to reflect a range of potential variations in population, property and meter forecasts.

The results of the forecast give a 16.99 Ml/day increase in household consumption for normal year demand scenarios including the impact of climate change, over the planning period (2019/20 to 2099/00), this is an 39.55% increase for the company. This is largely driven by a 75.44% increase in the property forecast.

In contrast, total PHC decreased by 20.45% over the forecast period and PCC showing a smaller decrease of 6.9 The reason for this disparity is due to decreasing occupancy . If occupancy is forecast to decrease, then per household consumption will be more greatly affected than PCC, as the relationship between the two variables is not linear. This reflects the 'economies of scale' inherent in the occupancy model which means that the proportional increases in consumption reduce as more people live in a property.

5 Recommendations

Although the model is as robust as it can be using the data available, some recommendations are made.

Additional data collection

The micro-component data and consumption-occupancy models used in the MC model and forecast are based on national datasets, dating from 2002-04 and 2015. While these can be used effectively at a company level by calibrating to local data, as described in this report, more up to date and regionally representative data is likely to improve the model.

The additional data could include a small sample of properties monitored for a number of weeks for micro-component consumption, and also surveyed to collect data on occupancy as well as other variables that may be useful for the MLR model. The monitored properties could be a sub-set of a larger survey sample.

Improve MC estimates for new builds

Another specific area where more data would be useful is in the estimation of micro-components for new-build properties. This type of property starts to become dominant amongst metered households in later parts of the planning period, given the rate of meter optants is low. Therefore, estimating current and future MC trends for new households would bring greater confidence in the forecasts of metered consumption.

Reconsider the critical period

The dry summer of 2018 resulted in exceptionally prolonged levels of peak demand in many areas of England and Wales. Artesia delivered a project to assess these peak demands, and their implications, which was published in summer 2020. This looked at the magnitude of peak demand over different durations for different water companies. It found that for Cambridge Water, the peak factor for the day and the week of highest demand was comparable or less than the 2013 peak. This is shown to be the case in Figure 18, which shows that the 7-day peak for 2018 is similar to the 2013 peak. The 2018 peak stands out much more when the peak volume over longer periods are compared.

The April and May period of 2020 was also unusually dry and warm. This occurred at the same time as the first UK lockdown, in response to the Coronavirus pandemic. This also resulted in prolonged periods of very high levels of household demand. At the time of writing, work is ongoing to determine the relative effects of the 'stay at home' lockdown and weather on demand for water during this period. Results will be available prior to the development of water company water resource management plans.

The extended periods of hot and dry weather experienced in summer 2018 and late spring 2020 are generally different from the shorter peak periods that have previously been considered critical to water resource systems. It is unclear whether such prolonged events will become more frequent in the future, however it is recommended that companies consider the resilience of their systems to these longer spells of hot and dry weather.

6 Appendix

6.1 Problem characterisation – RAG matrix

Guidance on the selection of appropriate household consumption forecasting methods were developed by UKWIR, along with guidance on the application of these methods.

The UKWIR guidance identifies nine criteria (Table 1) and a weighting and scoring framework, set out in a 'RAG Matrix' .The guidance recommends that practitioners adapt the weightings and scores in this matrix to reflect their own situation, in order to identify the most appropriate methods for forecasting household consumption. In particular, the matrix should be amended to reflect the level of planning concern in a particular WRZ.

Cambridge Water have used the RAG matrix, with amendments to reflect the status of its single WRZ to shortlist preferred methods for household consumption forecasting.

The following tables show the result of this work.

Cambridge (Moderate)	Weighting	Regression models	Micro-component models	Macro-component models	Trend-based models	Variable flow methods
Acceptance by stakeholders	20	6	8	7	5	3
Explicit treatment of uncertainty	14	8	6	6	5	3
Underpinned by valid data	14	6	6	7	5	3
Transparency and clarity	10	6	8	7	4	4
Appropriate to level of risk	10	8	7	6	5	4
Logical and theoretical approach	4	8	8	8	6	4
Empirical validation	3	6	7	7	6	5
Explicit treatment of factors that explain HH consumption	2	6	7	7	6	4
Flexibility to cope with new scenarios	1	8	6	7	4	4
Weighted score		**526**	**551**	**526**	**388**	**267**
Rank		**2**	**1**	**2**	**4**	**5**

References

Aistleitner, C., Hofer, M., & Tichy, R. (2012). A central limit theorem for Latin Hypercube Sampling with dependence and application to exotic basket option pricing. *International Journal of Theoretical & Applied Finance*, 15(7).

UKWIR. (2002). *An improved methodology for assessing headroom, 02/WR/13/2.*

UKWIR. (2006). *Peak water demand forecasting methodology, 06/WR/01/07.*

UKWIR. (2013). *Impact of climate change on water demand - Main report. 13/CL/04/12, ISBN: 1 84057 684 7.*

UKWIR. (2015). *WRMP19 Methods: Household consumption forecasting, 15/WR/02/9.*

UKWIR. (2016). *Integration of behavioural change into demand forecasting and water efficiency practices, 16/WR/01/15.*

UKWIR and EA. (2015). *WRMP19 Methods - Population, household property and occupancy forecasting guidance manual. 15/WR/02/8. ISBN: 1 84057 804 1.*

WRc. (March 2005). *Increasing the value of domestic water use data for demand management.*