artesia

Anglian Water

Project Discovery - Final Report

Segmentation and benchmarking of non-household properties

Project reference: 2579

Report number: AR1560

2024-03-13

Report title:	Project Discovery - Final Report
Project reference:	2579
Report number	AR1560
Date:	2024-03-13
Client:	Anglian Water

Version	Author(s)	Reviewed by:	Description	Date
1.0	Francesca Cecinati	Cristina Munilla Jamie Howe	Initial Draft	08/12/2023
2.0	Francesca Cecinati, Cristina Munilla Jamie Howe	Dave Gough	Draft	14/12/2023
3.0	Francesca Cecinati, Cristina Munilla Jamie Howe, Dave Gough	Dene Marshallsay	Final draft	21/12/2023
3.1	Francesca Cecinati, Cristina Munilla Jamie Howe	Dave Gough	Final Version	31/01/2024
3.2	Francesca Cecinati, Cristina Munilla Jamie Howe	Dave Gough	Added minor corrections to Sections 5.2.2 and 6.	07/02/2024
3.3	Francesca Cecinati	Dave Gough	Minor edits to Introduction	13/03/2024

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

Project Discovery has developed new approaches to segment non-household (NHH) customers with the aim of benchmarking water consumption, improving demand forecasting, achieving Outcome Delivery Incentives (ODIs) targets and regulatory performance targets, assisting in responses to water stress, and to help identify water efficiency opportunities. This includes understanding what is driving demand in NHH properties and exploring how granular consumption data could be used to augment the segmentation and targeting.

Conducting a global literature review at the start of the project proved insightful and revealed that previous attempts to benchmark NHH customers had often resulted in time consuming and data heavy processes. These typically consisted of conducting customer surveys to understand site specifics including water-using appliance details, estimated usage, and occupancy. While we recommend seriously considering regular onboarding and on-going surveys as a mean to improve understanding and data quality, we anticipated that we would not have the luxury of such detail within the timescale of this project, and needed to build a desktop solution which had a national reach, with a key focus on accessibility and ease of use.

Armed with water consumption data provided by the water companies in the project steering group, including granular consumption data (captured at 15 minute or 1-hour intervals) for over 45,000 NHH properties, we have built a schema which combines Standard Industry Classifications (SIC) with expert knowledge on water consumption and consumption data from smart metering systems. We have named this schema the Commercial Consumption Analysis (COCOA) Schema.

The COCOA schema consists of the following components:

- **The COCOA Classification**, which aims to classify NHH properties based on water usage behaviours. It is built on a functional water use classification and a data-driven classification, developed by studying consumption profiles from smart metering data.

- **The COCOA Benchmarking**, an estimate of expected consumption for properties in each of the COCOA Classification groups, for each calendar month.

These have been combined through modelling to allow the expected monthly consumption for individual properties to be estimated (within defined uncertainty bands).

The COCOA Schema can be consulted through lookup tables, linking SIC Codes to the COCOA Classification and Benchmarking. However, to make the tool more widely accessible and more efficient, we have produced the COCOA Excel tool, which can be used to output monthly and annual consumption estimates based on the industry grouping and the floor area, if current consumption is known the tool will identify if this is an outlier (using an excessive amount of water compared to COCOA Benchmarking estimations).

With a small amount of upfront information, the COCOA Schema and tool will provide users with the estimated consumption of a building and an indication of whether the building is using an excess level of water. This is unprecedented and constitutes a significant advancement in the industry.

We have validated the COCOA Schema using a set of water efficiency visits conducted on NHH properties, for which data was provided by one of the wholesalers on the project steering group. Although the selection of properties for these visits is not fully representative of the wider NHH population and the sample size was very small, the result is that the COCOA Classification proved accurate for all cases we could analyse, when SIC Codes are correct (73% of cases), and that the Benchmarking is reasonable in estimating the order of magnitude of consumption ($R2 = 0.93$ with $p\text{-value} < 2.2\text{e-}16$).

However, the Schema comes with many limitations due to underlying data accuracy and gaps and should be considered as a first resource to pinpoint high users, followed by more accurate verifications (e.g. online lookup, phone calls, additional data…).

Throughout the project, Artesia and the project steering group have identified a significant number of recommendations, and these are documented in this report (Section 8). A substantial number of our recommendations concern the quality of data available, both internally and externally to water companies. When attempting to develop a benchmarking schema with universal application, the availability of data for modelling and testing is crucial.

We recommend a further exploration of external data sources that could not be used for accessibility, price, or other practicalities, including Address Base Premium and the Inter-Departmental Business Register (IDBR) from the Office for National Statistics (ONS), to gather better classification and scaling data. The coverage and quality of SIC Classification across the water industry also needs improvement for all sectors, which would enhance COCOA applicability and accuracy. This is currently a retailer responsibility but could benefit from MOSL support.

With more and more smart meters being fitted to non-household properties, the COCOA schema could be greatly improved, particularly if further research on the use of continuous flow (CF) in commercial properties could be incorporated. The use of CF was thoroughly investigated, but, in agreement with the project steering group, it was decided that the understanding of CF in NHH is not advanced enough to consider it within the boundaries of this project. Finally, any further water efficiency site visits and other direct communication with NHH customers will help verify the type of businesses and their water-usage characteristics.

We anticipate that the COCOA Schema will be a valuable analytical tool to support companies more effectively target their business demand reduction activities. The COCOA Schema, the associated documentation, and an Excel tool for its easy application will be made openly available. Following a project dissemination, we plan on facilitating user groups to continue testing and enhancing the COCOA Schema. Extensive testing from all industry stakeholders is crucial and will allow model evolution to improve usefulness as a tool across users.

Contents

Figures

Tables

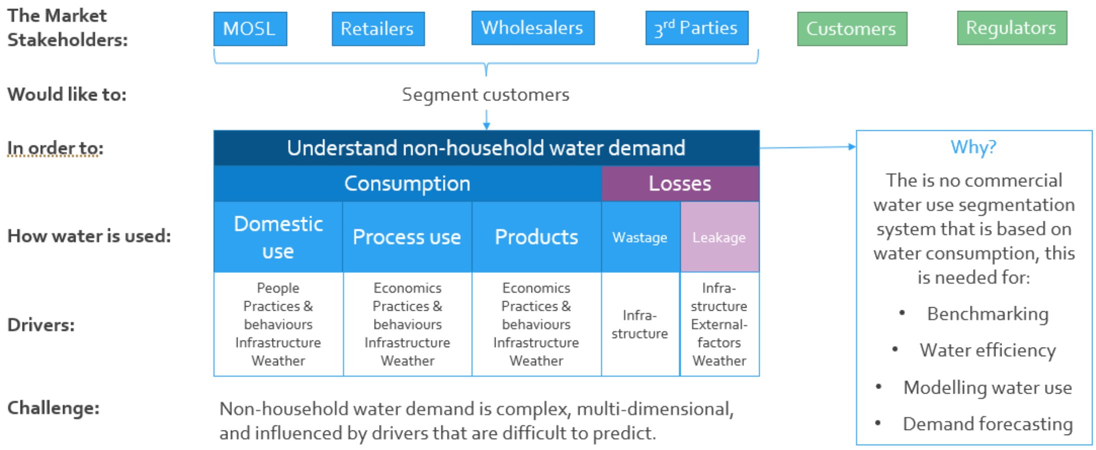
1 Introduction

Project Discovery is funded by the Market Improvement Fund (MIF). The MIF was set up to fund innovative projects that will benefit the non-household water market and its customers. The fund is overseen by the Strategic Panel (including project selection, funding allocation and progress of work) and administered by MOSL. A steering group was formed in summer 2022 consisting of several stakeholders, prior to the project starting in February 2023.

Project Discovery has developed new approaches to help segment non-household (NHH) customers with the aims of benchmarking water consumption, improving demand forecasting in the short term (peak and seasonality), longer term (sector growth), and in responses to water stress (resilience and criticality), and targeting water savings. This includes understanding what is driving water demand in NHH properties and exploring how granular water consumption data could be used to augment the segmentation and targeting.

Figure 1 visually depicts how a better understanding of water demand allows identification of consumption and losses and helps benchmark NHH customers for all water industry stakeholders.

Figure 1. Project Discovery overview



To achieve this, the Commercial Consumption Analysis (COCOA) Schema was developed, consisting of a water-consumption driven commercial classification and corresponding consumption benchmarking.

The anticipated benefits from Project Discovery include:

- Improved understanding of water consumption across NHH customers by industry sector.
- Better understanding of what is driving demand in NHH properties.
- Improving demand forecasting and planning in the short term (peak and seasonality), longer term (sector growth), and in responses to water stress (resilience and criticality).

- Reduced NHH demand through identification of targeted and repeatable demand management options.
- NHH customers and retailers being better equipped to identify and address leakage or wastage and compare consumption to other similar businesses.

Improved understanding of water demand in different sectors will assist wholesalers by improving resilience against regional demand and supply deficit imbalances.

The project was divided in three stages, and technical reports were produced at the end of Stage 1 and 2. This report is intended to summarise the rationale, outcomes, and recommendations from the project, for the benefit of a wider audience. For technical details, the reader is redirected to the project's technical reports.

Alongside this report and the technical reports, the COCOA Tool is provided, a spreadsheet that can be used to easily apply the COCOA Schema, both in terms of classification and benchmarking.

2 Preliminary Investigation

2.1 Literature review - external

Non household water benchmarking studies are scarce, so the literature review encompassed eight studies from 2003 to 2021 of which half were UK based.

Where benchmarking values were produced, specific property infrastructure and water use information was required for benchmarking, such as the star rating of hotels, whether schools had swimming pools, or premises had laundries. While this type of information is important to estimate water consumption, it is also difficult and costly to obtain at large scale.

As well as the characteristics of the premises, scaling factors are also needed for benchmarking (i.e., quantities that can inform us on the scale of the business). The literature review showed that various scaling factors were used within the same schema: floor area, visitors per year, employees/pupils/residents, tonnes of goods processed, building energy consumption, and other variables. Such data is likely to be well correlated to water consumption but is unlikely to be available at large scale.

The studies reviewed focussed on benchmarking consumption and potential savings for very specific sectors. A gap analysis of the studies showed that the following components were lacking:

- Analysing flow data in terms of continuous and variable use components.
- Temporal (e.g. event-based, diurnal and seasonal) profiling as a method for classifying and benchmarking properties.
- Effectiveness and coverage of existing scaling and clustering methods in the UK.

These findings were the first warning that classifying properties for a benchmarking schema would have been poorly supported by data availability, especially if ambition extended to national level for all NHH sectors, considering savings potential, seasonality, and criticality.

2.2 Literature review- internal

A review of Artesia client projects highlighted some useful technical capabilities in terms of benchmarking, like breaking down consumption into variable and continuous use, as well as clustering in terms of seasonality.

We found evidence that individual properties may be characterised by their consumption patterns; this is an important step, in that granular flow data can potentially help define the type of business by observing the consumption pattern.

A finding repeated in multiple projects was that it was difficult to benchmark very large users with confidence. Non household users (NHH) are more varied in scale and type of water use, with a more skewed distribution than households (HH). We would therefore require a larger sample size to model (or benchmark) NHH than HH, but we usually have a smaller sample.

Through this project, we collected a very large dataset (45,000+ smart metered properties) to try and overcome this issue.

Key findings of internal literature search:

- Existing classification schemas (like the Standard Industry Classification – SIC) are insufficient to characterise NHH properties with respect to their consumption behaviours.
- The larger high users are very few. Although these are usually logged, the small numbers make them difficult to characterise, and errors in modelling their use are not averaged out over a large sample. Modelling is greatly improved by removing these from the sample and monitoring them individually.
- Clustering NHH by seasonality is viable where granular data is available.
- Existing projects have not identified viable universal scaling factors for benchmarking.

2.3 Data investigation

As part of the preliminary investigations, we also received, pre-processed, and quality assured the data needed to carry out the project, provided by four wholesalers (Anglian Water, Affinity Water, Thames Water, and Yorkshire Water) and MOSL.

We received:

- Granular data from the wholesalers for ~46,000 properties, reporting meter data at 15-min or hourly resolution.
- Enriched data from the Market Operator Services Limited (MOSL), for all the NHH properties covered by the wholesalers. This dataset linked each supply point ID (SPID) to business details, including the Standard Industry Classification (SIC), version 2007, and the Unique Property Reference Number (UPRN), which can be used to link other property information.
- Information on ~12,000 business visits, which collected business and consumption information, as well as water saving information details. This information was precious for validation.

A brief review of the SIC 20007 classification, widely used in the UK, showed that only the finest granularity (i.e., SIC Code – 1188 codes split between classes and sub-classes) contains information about **functional water use**, i.e. how water is used, and for what purpose. At this fine classification scale, we can gather information about the scale of use, temporal profile, and drivers. Note that SIC codes are 5-digit codes that represent the finest available classification, either class or sub-class.

For example, an office for an energy utility company versus a power plant has the same SIC Section, Division, and Group, and only differentiate in terms of Class, which is the finest level. However very different water use patterns might be expected between those two premises.

For this reason, we excluded the use of other schemas that did not reach that level of detail, like the Valuation Office Agency (VOA) schema.

While the SIC classification proved to be the most detailed available and is widely used and recognised in the UK, unfortunately its coverage in CMOS is limited, with only approximately 30% of properties having a SIC classification. MOSL does not plan to dedicate resources to improve it. Additionally, the SIC classification has 4 schemas (1980, 1992, 2003, 2007), which may create confusion and may require some re-mapping.

On this basis, we considered other datasets for possible use, including the Inter-Departmental Business Register (IDBR) dataset from the Office for National Statistics, private datasets, like the Ordnance Survey Address Base Premium and the Verisk's PropX datasets, or alternative data sources like the Energy Performance Certificate Data from the Department for Levelling Up, Housing & Communities.

However, we found that none of the datasets could be used in this project, for various reasons, including open access, price, completeness, or practical accessibility issues, but recommendations have been collated to help progress their use in the future.

2.4 Re-framing the problem

If we wish to develop a new classification for NHH properties based on water consumption, there is a wide range of distinctions we can make:

- We can consider functional use (process-driven, people-driven, hybrid).
- We can consider seasonality (natural seasons, holidays, financial seasons).
- We can consider consumption profiles at different scales (daily, weekly, yearly).
- We can consider properties' total, seasonal, variable, and continuous water use.
- We could consider specific characteristics, e.g. use of filling tanks, or private supplies.
- We could consider a business criticality (e.g. hospitals).

Even if we were able to consider all the factors above, the number of combinations would soon become too large to model, not having sufficient data to represent each combination meaningfully.

Therefore, we focussed the project's attention on the key applications for the schema (as defined by the project steering group) and consider practical constraints such as data availability and uncertainty.

In terms of applications, the steering group identified the following:

1. Identify premises with higher-than-expected consumption to deliver targeted water efficiency interventions.
2. Understand genuine continuous flow (CF) to improve nightline modelling and target leakage.
3. Understand how risk profiles varied across sectors.
4. Enable NHH consumption modelling by sector for Water Resource Management Plans (WRMP's).
5. Support drought order scenarios (non-essential use bans).

Based on the above and on technical considerations, we focused on total consumption on longer time scales (monthly, yearly), considering seasonality.

The use of continuous flow was carefully considered. However, the poor understanding of continuous flow in NHH premises, due to the difficulty of discerning between continuous use (e.g. for processes or for overlapping intermittent flow), wastage (e.g. leaky loos), and leakage, made it impossible to tackle the challenge within the boundaries of this project. On the other hand, identifying higher-than-expected flows will lead to saving potential, regardless of the cause. Based on this, it must be considered that when we refer to "consumption" in this document, we often refer to flows that can contain consumption and leakage.

The understanding reached by the end of the preliminary stage was that developing a schema is a delicate balance between using supporting data (flow data and property data), accessibility of the required information (e.g. SIC Codes), and robustness of the methodology (and therefore accuracy of outputs and conclusions).

The SIC codes contain useful information on functional water use and are the best available tool in that regard. For this reason, the COCOA Schema has been developed using SIC codes as the starting point. Regarding scaling data, no ideal dataset could be found, but Address Base Premium's *building area* was identified as the best balance between accessibility and meaningfulness. This field refers to the footprint area of the building and does not consider the number of floors, which represents a limitation. Other useful variables like the number of floors, building height, or number of employees were considered but were not universally available at accessible costs.

Therefore, the way forward was identified as building the COCOA classification in two layers:

- a logical new classification system based on functional water use clustering was developed, manually clustering SIC codes based on our expertise.
- Using the functional classification was further refined using seasonal profiles and a data-driven clustering approach.

3 Development of the COCOA Schema

This Section provides a high-level overview of how the COCOA Schema has been developed. For more details about the exploratory data analysis, the methodology, the testing, and the model performance the Stage 1 and Stage 2 technical reports for Project Discovery can be consulted[1,2].

3.1 Classification

Project discovery culminated with the development of the COCOA Schema. The development of this unified NHH classification was based on merging theoretical knowledge with a data-driven approach. We derived an optimal classification that can be used across the industry. Hence, this schema relies on two layers:

- Functional use clustering.
- Data-driven clustering.

3.1.1 Functional use classification

Following preliminary investigation, SIC codes were manually grouped based on their nature and functional water use (e.g. whether use is driven by people, processes, production). We used our experience in NHH analysis, modelling and field work to leverage information about the SIC description itself. We defined 32 functional use categories and allocated each SIC code in the most appropriate category. Table A-1, in Appendix A, specifies the 32 categories and their description and rationale, and their water functional use.

The categories were developed considering:

1. The type of business
2. The functional water use, including the following categories:
    a. **Domestic:** person driven demand, including washing and bathing, laundry, and food preparation.
    b. **Office:** person driven as office, but bathing, laundry, and food preparation absent or reduced.
    c. **Washing:** Cleaning of equipment or materials outside the scale of domestic use as baselined by occupancy of premises.
    d. **Embodied:** Water that is retained in the final product, such as soft drinks or cement.
    e. **Leisure:** water features, swimming pools, spas, theme park rides, etc…
    f. **Processing:** water used to change the nature of something. E.g. dying, boiling, electroplating, case hardening, diluting.
    g. **Coolant:** water used for the exchange or dissipation of energy E.g. steam generation, cooling.

[1] Artesia, 2023, *Project Discovery – Stage 1 Report*, Project 2579, report number AR1529

[2] Artesia, 2023, *Project Discovery – Stage 2 Report*, Project 2579, report number AR1551

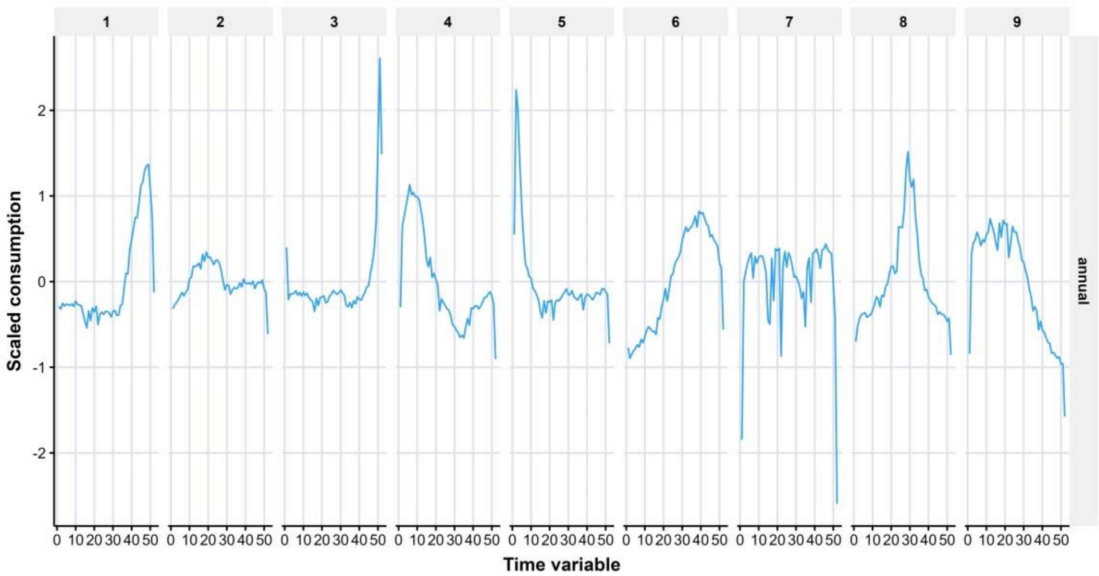
h. **Dust suppression:** sprays and mists used to condense airborne dust or prevent its formation.
i. **Irrigation:** agricultural use of water.
j. Unknown: likely that another type of functional water use may be present. For instance, a clinic may include hydrotherapy.

3. Seasonality: we are expecting the businesses in each category to be driven by similar seasonal drivers (weather, holidays, growing cycles, no seasonality)
4. Diurnal cycles: we are expecting the businesses in each category to be driven by similar daily cycle drivers (e.g. domestic or office routines)
5. Weekly cycles: we are expecting the businesses in each category to be driven by similar weekly cycle drivers (Monday-to-Friday, Monday-to-Saturday, weekend peak, continuous, etc...)

3.1.2 Data-driven classification

To refine the functional use classification, NHH consumption data from ~9000 properties with AMI meters at hourly or 15-minute resolution, was used to identify and characterise common annual water consumption patterns across commercial properties. To do this, an unsupervised machine learning technique was employed. The optimal number of clusters and the optimal temporal extent and resolutions were studied. This culminated with the identification of 9 common profiles, that constitute the COCOA data-driven classification. Figure 2 illustrates the 9 profiles, with the week number on the x axis (1 to 52) and the consumption normalised by its mean on the y axis (to allow comparison of properties with different total consumption).

Figure 2. Data-driven annual profiles at weekly resolution

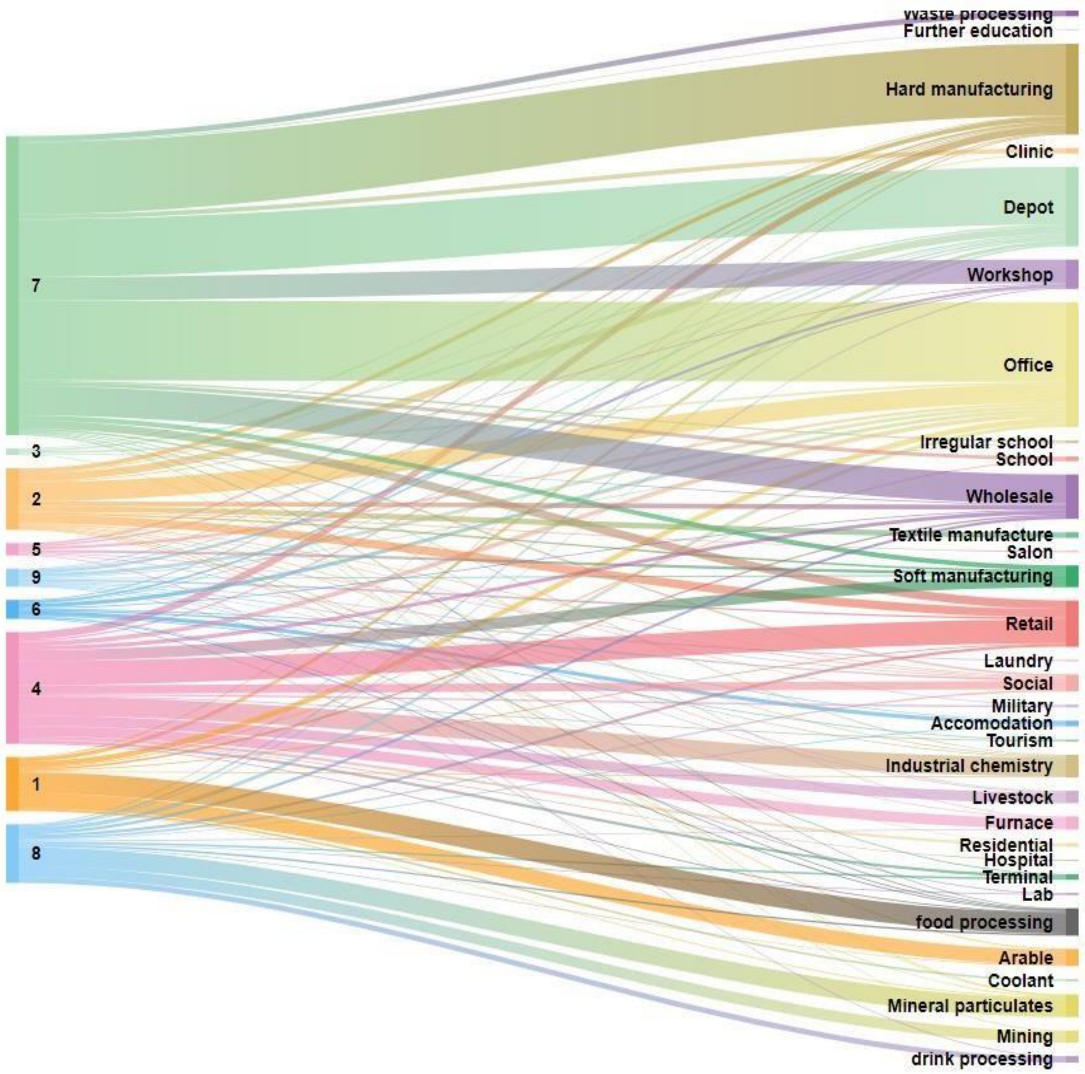


Each of the 9 profiles represent a different annual pattern: for example, Profile 7 represents businesses that close over Christmas, bank holidays, and summer holidays, while Profile 8 has a summer driven peak.

The relationship between the data-driven classification and the functional use classification was evaluated and combined as shown in Figure 3. Figure 3 shows what proportion of properties (line/band width) falls in the data-driven (left) and functional (right) classifications. The thicker the line, the larger the proportion of properties. As expected, there is no unique relationship (no group on the left is uniquely linked to one group on the right, or vice versa). We would expect some functional categories to exhibit different profiles, or for different functional categories to have similar profiles. However, the fact that most of the functional categories on the right are almost uniquely linked to one of the data-driven clusters is very positive: it means that the data-driven classification is strongly coherent with the theoretical understanding and expert judgement we considered in the functional use classification.

It must be noted here that the two layers of the COCOA Classification are not attempting to classify properties in the same way: the functional classification aims at clustering properties based on how water is functionally used and the nature of the business, while the data-driven classification is clustering properties based on seasonality and annual profiles. For this reason it is expected to have non-unique relationships.

Figure 3. Relationship between data-driven (left) and functional use (right) classifications



3.1.3 COCOA Classification

Hence, the COCOA Classification combines the functional-use classification and the data-driven classification, both based on SIC Codes 2007. The COCOA Classification is illustrated in full in Table A-2, in Appendix A, but a short extract is reported below, in Table 1, reporting 3 of the 32 functional use categories.

Table 1. Extract of the COCOA Classification and SIC relationship

Functional Use Classification	Data Driven Classification	SIC Codes 2007
Accommodation	6	55100 55200 55201 55202 55300 97000
Accommodation	8	55900
Accommodation	9	55209
Arable	1	01110 01120 01130 01140 01150 01160 01210 01220 01230 01240 01250 01260 01270 01280 01300 01610
Arable	2	01190
Arable	8	01290
Clinic	2	86210
Clinic	7	75000 86220 86230 86900 96030

In the COCOA Classification table, each line is a specific category made up of the functional use classification and the data driven classification, e.g.: Accommodation-6 (the first line in Table 1). In the full COCOA Classification, there are 115 individual categories.

3.2 Benchmarking

Based on the COCOA Classification, we developed the COCOA Benchmarking, i.e. a tool that helps estimating expected consumption for properties in each category. The tool aims at being easily accessible and interpretable for all stakeholders.

To achieve this, we developed consumption models that can predict consumption for each of the COCOA Classification categories (combination of the functional and data-driven groups).

We conducted an exploratory data analysis, which resulted in some observations:

- The relationship between building area and consumption is significant ($R^2 = 0.37$, p-value < 2.2 e-16).
- The relationship between building area and consumption is approximately linear, although variable with categories and seasons.
- Temporal factors (such as Seasonality) are important in estimating consumption for each group.

Based on these observations, linear regression models were built for each one of the categories, for each calendar month, and formed the basis to develop the COCOA benchmarking table. This is a table that reports consumption in litres per day per property per square meter, for each COCOA classification group and each month of the year. It also reports the uncertainty on each coefficient (coefficient ± uncertainty), expressed in the same units.

The full table is reported in Table A-3, in Appendix A, but an extract regarding the first 3 out of 32 functional use categories is reported in Table 2.

Bold rows represent the standard data-driven classification, i.e. the most common profile for each functional group. If SIC Codes are not known, the functional classification can be estimated (e.g. from business name or from other classification), and the standard profile can be used.

Table 2. Extract of the COCOA Benchmarking: estimated consumption in l/prop/day/m², by functional and data driven classification and month.

Functional classification	Data driven class.	January	February	March	April	May	June
Accommodation	**6**	**11.54±5.16**	**13.23±5.76**	**13.03±6.54**	**14.49±7.01**	**13.13±6.48**	**13.42±6.5**
Accommodation	8	21.68±24.7	21.91±24.19	19.36±21.32	20.61±23.27	21.71±25.02	20.73±24.16
Accommodation	9	13.13±5.47	14.59±5.92	14.12±6.49	15.52±6.91	14.52±6.64	14.61±6.47
Arable	**1**	**1.3±2.4**	**1.68±4.14**	**1.17±4.61**	**1.36±3.04**	**1.82±1.85**	**3.53±3.02**
Arable	2	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2
Arable	8	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2
Clinic	2	2.18±1.11	2.2±1.02	1.9±1.41	1.18±0.92	2.43±1.4	2.42±1.37
Clinic	**7**	**2.18±1.11**	**2.2±1.02**	**1.9±1.41**	**1.18±0.92**	**2.43±1.4**	**2.42±1.37**
Functional classification	**Data driven class.**	**July**	**August**	**September**	**October**	**November**	**December**
Accommodation	**6**	**14.94±7.86**	**15.66±7.88**	**14.38±6.81**	**13.96±6.55**	**13.26±6.2**	**13.46±6.28**
Accommodation	8	21.37±24.42	20.2±22.4	23.31±27.31	24.99±29.45	24.55±28.93	25.67±30.05
Accommodation	9	15.98±7.43	16.39±7.33	15.69±6.81	15.58±6.76	14.9±6.43	15.24±6.54
Arable	**1**	**4.02±4.08**	**3.44±3.65**	**1.76±2.07**	**1.52±2.37**	**1.48±3.09**	**2.43±2.66**
Arable	2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Arable	8	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Clinic	2	2.58±1.45	3.12±1.74	2.77±1.45	2.58±1.28	2.41±1.18	2.42±1.49
Clinic	**7**	**2.58±1.45**	**3.12±1.74**	**2.77±1.45**	**2.58±1.28**	**2.41±1.18**	**2.42±1.49**

The uncertainty on the estimations is based on three components, quadratically combined: model uncertainty (which captures the inherent variability within the data), sample size uncertainty (which captures the representativeness of the sample), and model type uncertainty (which captures the modelling approximations). Note that some of the estimates are more uncertain than others, sometime with uncertainty bands larger than the estimate itself. This is still a valid estimation but highlights that we can expect great variability within the category.

3.3 Limitations

The development of a unified NHH classification was based on merging theoretical knowledge with a data-driven approach. We derived an optimal classification that can be used across the industry. However, there are some limitations that need to be considered when using the classification:

- One notable limitation encountered during this analysis is the availability of data: considering availability of SIC Codes, at least a year of smart meter data, and QA, only 9,027 properties could be used for clustering. Although this is not a small number, it must be considered that it is not necessarily representing the totality of NHH property in the UK, might be biased as AMI meters may not be evenly distributed across sectors, and that some of the 600+ available SIC codes were under-represented or not represented at all. More specifically over 33% of the SIC Classes described in the functional classification in Stage 1 were missing from the granular data sample. This could impact the accuracy of our findings, particularly when drawing conclusions about certain industries.
- We developed the classification based on the finest available SIC Code resolution. However, properties within each SIC Code may still exhibit different characteristics and water consumption behaviours.
- Location, business size, and other unaccounted factors might have an influence on the results and could cause variability within clusters that we could not account for.
- We have developed the classification using available historic granular data. It must be considered that water consumption habits have drastically changed over the past few years, as NHH properties were affected by COVID-19 and associated restrictions, a drought occurred in 2022 with temporary use bans in place, and water consumption was affected by an increase in inflation and costs of energy, materials, and services (cost-of-living crisis). Therefore, we cannot consider water consumption habits stationary and the conclusions if valid today may require a review in a near future.
- Our modelling assumes that SIC classifications are correct, but we are aware that this may not always be the case. At times, properties may be mis-labelled, or the classification could be out of date. We have no way of verifying this at the scale required for this project. We recommend reviewing the work in this project after MOSL concludes an ongoing data cleansing exercise, that will improve the quality and quantity of SIC classifications for NHH properties.

Based on the observations above, we recommend using the classification as an indication of how NHH properties are expected to behave on average, being aware that the variability compared to expectations could be large.

The COCOA Benchmarking is based on the COCOA Classification, with all the caveats above. Additionally, the Benchmarking was based on the use of granular data: for some models, sufficient data was available, allowing to build robust models, for others the data was scarce or the model performance not optimal.

The use of the building area as a scalar has been a pragmatic choice, and it helps estimating the order of magnitude of consumption, but it is not the ideal scalar. Consumption may not depend on building area only, and additional factors, like the number of employees, of onsite devices, the presence of private supplies, or specific characteristics of the building/business might have a strong effect too. Additionally, the building area as defined by the OS indicator does not account for multiple floors, or for non-building area (e.g. for arable land or golf courses), which could cause high variability.

For these reasons, we recommend using the benchmarking as follows:

- Use the estimations as an order of magnitude indicator.
- Properties that are classified as outliers (or possible outliers) should be reviewed before deciding on intervention targeting, to ensure that:
    - The classification and building area estimations are correct.
    - The property is not on multiple floors.
    - There isn't additional non-building, water-using land.
    - There are no simple explanations for the higher water consumption (e.g. presence of swimming pools)
- Consider the uncertainty in the predictions.
- The COCOA Benchmarking is the first tool of its kind and needs application and testing to lead to more refined tools in the future. Continue using it and providing feedback, when possible.

3.4 COCOA Schema

The COCOA Schema was therefore finalised, as the combination of the COCOA Classification and the COCOA Benchmarking, both available as lookup tables in the Appendix.

The COCOA Classification is composed of two layers: the functional use classification, which was manually defined based on logical grouping and functional use (what drives water consumption within the commercial property), and the data-driven classification, which represents the expected yearly consumption profile. The combination of the two layers results in 115 COCOA Classification categories.

The COCOA Benchmarking provides an estimation of consumption for each COCOA Classification category and each month of the year, using building area as a scaling factor. This means that consumption is provided in $m^3/day/m^2$. Additionally, an estimation of consumption uncertainty is provided, using the same units.

The COCOA Schema can be accessed through the lookup tables available in this document's Appendix A. However, to make the schema more accessible and efficient, a COCOA Excel tool has been developed. The tool allows the user to input SIC Code, building area, and consumption for one or more properties, and automatically provides the COCOA Classification and Benchmarking, together with an indication of whether the property is an outlier, i.e., consumes much more water than expected.

4 How to use the COCOA Schema

Throughout this report we presented the COCOA Schema. This section explains how to use the schema.

4.1 How to use the schema using lookup tables

Once a user is familiar with how the COCOA Schema has been developed and its limitations, the tables provided in Appendix A are all one needs to use the COCOA Schema.

To provide an example, imagine we would like to understand whether a given property is consuming water in line with expectations or not. We do have its SIC Code and Master Map building area.

Property A is a veterinary clinic, classified through SIC Code 75000, and has a building area of 100 m^2. We know in the past year it has consumed an average of 500 litres per day. Is this in line with expectations or should this be considered an outlier?

To start, we use the COCOA Classification table to review what is the COCOA Classification for Property A. Both the full COCOA Classification table in Appendix A, Table A-2, and the extract in Section 3.1.3, Table 1, can be used for this purpose. In Table 3 an extract is reported for the line of interest.

Table 3: Extract of the COCOA Classification table that concerns SIC Code 75000.

Functional classification	Data driven classification	SIC Code
Clinic	7	75000 86220 86230 86900 96030

Based on the COCOA Classification table, we know that Property A can be classified as "Clinic" with profile 7.

Therefore, we can use the COCOA Benchmarking table to estimate its consumption profile throughout the year. Both the full table in Appendix A, Table A-3, and the extract in Section 3.2, Table 2, can be used. An extract of the line of interest is reported in Table 4.

Table 4: Extract of the COCOA Benchmarking that concerns Clinics with profile 7.

Functional classification	Data driven class.	January	February	March	April	May	June
Clinic	7	2.18±1.11	2.2±1.02	1.9±1.41	1.18±0.92	2.43±1.4	2.42±1.37
Functional classification	**Data driven class.**	**July**	**August**	**September**	**October**	**November**	**December**
Clinic	7	2.58±1.45	3.12±1.74	2.77±1.45	2.58±1.28	2.41±1.18	2.42±1.49

By using the MM building area for Property A, 100 m², we can convert the profile above in actual consumption, by multiplying coefficients by the property area in m². The result is reported in Table 5.

Table 5: Consumption profile for a 100 m² property classified as Clinic with profile 7

Functional classification	Data driven class.	January	February	March	April	May	June
Clinic	7	218 ± 111	220 ± 102	190 ± 141	118 ± 92	243 ± 140	242 ± 137
Functional classification	**Data driven class.**	**July**	**August**	**September**	**October**	**November**	**December**
Clinic	7	258 ± 145	312 ± 174	277 ± 145	258 ± 128	241 ± 118	242 ± 149

A property should be considered as a weak outlier if it exceeds the upper band of the COCOA estimation (i.e., estimation + uncertainty), for at least 6 months on a 12-month period, while it should be considered a strong outlier if it exceeds 1.5 x the upper band for at least 6 months on a 12-month period.

We do not have Property A consumption for each calendar month, but we can compare the average yearly value with each of the monthly values. Property A actual consumption is on average 500 l/day. Even considering the uncertainty (estimate + uncertainty), all COCOA estimates are below 500 l/month. However, the actual consumption does not exceed 1.5 times the monthly estimations.

This means that **Property A should be considered a possible outlier**.

4.2 How to use the schema using the COCOA tool

While using the lookup tables provided is effective, it may not be efficient if a larger number of properties need to be compared. For this reason, we have provided a spreadsheet that allows the user to consult the COCOA Schema in a much more efficient way.

The spreadsheet contains 7 visible tabs:

1. **Home** – Providing an introduction, instructions, and ancillary information.
2. **Context** – Providing more details about Project Discovery and the COCOA Schema.
3. **Lookup with SIC code** – This is the tab that needs to be used if a SIC code is known for the property(ies) to be investigated.
4. **Lookup with Functional** – This is the tab that needs to be used if a SIC code is not known for the property(ies) to be investigated.
5. **COCOA Classification** – For reference, reports Table A-1, Table A-2, and the data-driven profiles.
6. **COCOA Benchmarking** – For reference, reports Table A-3.
7. **SIC Classification 2007** – For reference, SIC Schema 2007 from the Office for National Statistics is reported.

Based on the above, only the tabs at point 3 and 4 can be modified.

Figure 4 shows how the first 6 columns of the tab *Lookup with SIC code* looks like, with some sample properties used as an example.

Figure 4. First 6 columns of the *Lookup with SIC Code* tab in the COCOA tool

INPUT				COCOA Classification	
Property Identifier	SIC Code (2007)	Area [m2]	Consumption [m3/day]	Functional Classification	Data-Driven Classification
1	55100	1000	20	Accommodation	6
2	01110	2000	11	Arable	1
3	42990	1000	1.5	Depot	7
4	01190	1000	5	Arable	2
5	75000	100	0.5	Clinic	7

Note that the last line reproduces the example used before, in Section 4.1.

Only the cells under INPUT need to be modified, by including the SIC Code and Area, while Property identifier and Consumption are not essential, but can be used to keep track of the properties and gather additional information respectively. By filling the SIC Code and Area, The Functional Classification and the Data Driven Classification are automatically populated.

Once this is done, the following columns are also automatically populated and provide the Benchmarking and other useful information (See Figure 5, which is cut to February, but the spreadsheet provides all values up to December).

If the *Consumption [m3/day]* column is populated, the spreadsheet compares the actual consumption with expectations:

- A record is marked as a non-outlier (*No*, in green) if the actual consumption is below the average upper uncertainty band.
- A record is marked as a weak outlier (*Maybe*, in yellow) if consumption is below 1.5 the average upper uncertainty band. This accounts for some additional uncertainty and variability between properties. These properties are lower priority than the ones labelled as outliers.
- A record is labelled as strong outlier (*Yes*, in red) if the actual consumption is above 1.5 the average upper uncertainty band.

As the development of the COCOA Benchmarking is based on various models, some more uncertain than others, the Uncertainty column highlights the models that have uncertainty bands larger than 75% of the coefficient itself.

The benchmarking is provided both as yearly average and as monthly values, and for each the lower, central, and upper estimation is provided. All benchmarking is provided as [m^3/day] already accounting for the Area of the building.

The spreadsheet has been designed to accommodate 10,000 properties.

Figure 5. First 17 columns of the *Lookup with SIC Code* tab in the COCOA tool

INPUT				COCOA Classification		OUTLIERS	UNCERTAINTY	COCOA Benchmarking [m3/day]								
								Low	Central	High	Low	Central	High	Low	Central	High
Property Identifier	SIC Code (2007)	Area [m2]	Consumption [m3/day]	Functional Classification	Data-Driven Classification	Is this an outlier?	Is this result highly uncetain?	Average	Average	Average	January	January	January	February	February	February
1	55100	1000	20	Accommodation	6	No	No	7.12	13.71	20.30	6.38	11.54	16.70	7.47	13.23	19.00
2	01110	2000	11	Arable	1	Maybe	Yes	0.09	4.25	10.42	0.00	2.61	7.41	0.00	3.36	11.64
3	42990	1000	1.5	Depot	7	Yes	No	0.34	0.49	0.64	0.35	0.47	0.60	0.28	0.45	0.61
4	01190	1000	5	Arable	2	Yes	No	0.97	1.15	1.34	0.94	1.10	1.26	0.96	1.12	1.29
5	75000	100	0.5	Clinic	7	Maybe	No	0.10	0.23	0.37	0.11	0.22	0.33	0.12	0.22	0.32

Following the previous example, represented in the last filled line, we can see that, by entering SIC Code, Area and consumption, the property got classified as Clinic with Profile 7, with a yearly average consumption between 0.10 and 0.37 m^3/day. The property got labelled as a possible outlier, as the consumption of 500 l/day = 0.5 m^3/day is above the upper threshold, but not above 1.5 the upper threshold. This means that the property consumes more than expected, but not order of magnitudes more.

It can be targeted for interventions, but higher priority should be given to properties with consumption much higher than the expected range.

4.3 How to use the Schema without SIC code

Both the lookup tables and the COCOA Excel tools account for the fact that SIC codes may not be known (we observed that this is the case for ~70% of properties in CMOS).

Example: Property B is called "The White Lion Hotel", has a floor area of 5000 m², and an average consumption of 50 m³/day. The exact SIC code is not known; however, from the name we understand it is a hotel, and therefore falls into the "Accommodation" functional classification.

Using Lookup tables, we can directly look at the COCOA Benchmarking table. Both the full table in Appendix A, Table A-3, and the extract in Section 3.2, Table 2, can be used. An extract of the line of interest is reported in Table 6.

Table 6. Extract of the COCOA Benchmarking that concerns Accommodation, in l/day/m².

Functional classification	Data driven class.	January	February	March	April	May	June
Accommodation	**6**	**11.54±5.16**	**13.23±5.76**	**13.03±6.54**	**14.49±7.01**	**13.13±6.48**	**13.42±6.5**
Accommodation	8	21.68±24.7	21.91±24.19	19.36±21.32	20.61±23.27	21.71±25.02	20.73±24.16
Accommodation	9	13.13±5.47	14.59±5.92	14.12±6.49	15.52±6.91	14.52±6.64	14.61±6.47
Functional classification	**Data driven class.**	**July**	**August**	**September**	**October**	**November**	**December**
Accommodation	**6**	**14.94±7.86**	**15.66±7.88**	**14.38±6.81**	**13.96±6.55**	**13.26±6.2**	**13.46±6.28**
Accommodation	8	21.37±24.42	20.2±22.4	23.31±27.31	24.99±29.45	24.55±28.93	25.67±30.05
Accommodation	9	15.98±7.43	16.39±7.33	15.69±6.81	15.58±6.76	14.9±6.43	15.24±6.54

As we do not know the best data-driven classification, we can use the standard one, corresponding to the most common profile for Accommodation properties, highlighted in bold (profile 6 in this case).

Then we can multiply the coefficients by the area of 5000 m², as reported in Table 7. Note that, given the order of magnitude, consumption in Table 7 is reported in m³/day, which can be converted to l/day multiplying it by 1000.

Table 7. Extract of the COCOA Benchmarking that concerns Accommodation-6. Note that in this case m³/day are used.

Functional classification	Data driven class.	January	February	March	April	May	June
Accommodat.	**6**	**57.71±25.81**	**66.17±28.81**	**65.15±32.71**	**72.47±35.07**	**65.63±32.39**	**67.11±32.51**
Functional classification	**Data driven class.**	**July**	**August**	**September**	**October**	**November**	**December**
Accommodat.	**6**	**74.69±39.31**	**78.29±39.42**	**71.89±34.05**	**69.82±32.75**	**66.28±31.00**	**67.31±31.40**

We know that the average consumption for Property B is 50 m3/day. This is below the central estimation for all months, meaning that the property is not an outlier.

This calculation can be confirmed by using the Excel tool. Instead of the Lookup with SIC Code, we need to use the Lookup with Functional tab. The result is shown in Figure 6, confirming that the standard profile for Accommodation is 6 and that the property is not marked as an outlier (note that the COCOA consumption estimation columns are not reported in Figure 6, but report the same values as Table 7).

Figure 6. Application of the COCOA Schema through the COCOA Excel tool when SIC Codes are not available

INPUT				COCOA Classification		OUTLIERS
Property Identifier	Functional Classification	Area [m2]	Consumption [m3/day]	Functional Classification	Data-Driven Classification	Is this an outlier?
1	Accommodation	5000	50	Accommodation	6	No

4.4 Applications of the COCOA Schema

As discussed in Section 2.4, the COCOA Schema has been developed to address 5 objectives identified by the PSG. This section looks at how the COCOA Schema can be used for these purposes.

4.4.1 *Identify premises with higher-than-expected consumption to deliver targeted water efficiency interventions*

Water companies are interested in reducing NHH consumption, as a large proportion of the total demand. To do so, they can carry out water efficiency interventions in NHH premises. Given budget and practical constraints, they usually need to prioritise a limited number of premises.

The COCOA Schema can be used for this purpose, by inputting in the COCOA tool consumption, SIC code, and floor area for all candidate NHH properties. The *Outlier* flag can be used as a first tool to shortlist a limited number of properties, which can then be scored based on the difference from expectations.

Once a preliminary prioritisation is done, it is recommended to validate the input data and the COCOA outcome through alternative data sources, for example verifying that the SIC code is correct and that the floor area is reasonable, considering the possible presence of multiple floors or outdoor areas.

4.4.2 Understand genuine CF to improve nightline modelling and target leakage

This point could not be addressed through the project as it is focused on a different temporal scale (sub-daily) compared to the other points (seasonal), and a significant amount of investigation would be needed to robustly model the differences between wastage, leakage, and genuine continuous use in NHH.

However, the COCOA classification can support future investigations on NHH CF, by providing a segmentation that is better targeted to water consumption.

4.4.3 Understand how different risk profiles varied across different sectors

The COCOA classification can help identify which sectors are more likely to provide outliers (i.e. NHH customers with excess water consumption). As the COCOA schema is tested more widely in the industry, it will be possible to investigate which sectors are more likely to exceed the COCOA consumption benchmark.

This information, in conjunction with a better understanding of seasonal profiles, can also identify supply risks in a peak consumption scenario.

4.4.4 Enable NHH consumption modelling by sector for Water Resource Management Plans (WRMP's)

Currently WRMP use a very high-level segmentation, usually simply dividing NHH in service and non-service businesses, at times using a few more economic-based categories.

The COCOA Classification can support a better segmentation based on water functional use. The 32 functional categories can represent a starting point, that can then be grouped based on similar functional use (as per Table A-1) or common observed behaviours, rather than using economy-driven expectations.

4.4.5 Support drought order scenarios (non-essential use bans)

Currently drought measures are mostly targeted at HH consumption. This is because HH consumption represents a larger share of total demand, but also to the fact that it is more predictable and understood.

The COCOA Schema offers insight into:

- How water is used functionally within NHH premises, and therefore what type of drought measures could be implemented without impacting the business ability to work.
- What businesses are expected to align their consumption profiles with HH consumption peaks, i.e. what businesses are associated with data-driven profiles

with summer peaks (mostly profiles 6 and 8), but also what profiles may be more affected depending on the timing of the drought (early summer, late summer, etc...).

5 Validating the COCOA Schema

In the final stage of the project, the COCOA Schema was validated. For this purpose, Smart Business Visit (SBV) data was provided for a set of commercial properties identified by SPID references. SBV's help NHH customers fit water saving devices, identify and potentially fix leaking appliances and provide recommendations to reduce water consumption. This dataset contained data regarding specific interventions in individual properties such as install of water saving devices or fixes, as well as a record of the type of business. This makes this dataset particularly valuable to validate the COCOA Schema.

Although the dataset includes data for more than 12,000 visits, we had to restrict the database significantly, by considering only those properties for which we had smart meter data (only 217) and for which we had sufficient consumption records before and after the visit. This led us to select 75 properties. This is a small number. However, it must be considered that the models supporting the COCOA Schema were already tested on available granular data; this validation is an additional sense-check on properties where much more information is available.

We then proceeded in validating the COCOA Schema using these properties. We validated both the COCOA Classification and the COCOA Benchmarking and selected 3 case studies to illustrate the value of the Schema.

5.1 COCOA Classification

As previously illustrated, the COCOA Classification is made of two layers, the functional use clustering and the data driven clustering.

5.1.1 *Functional use clustering*

We manually compared the COCOA functional clustering to the SBV classification, based on the property visits, for each of the 75 selected properties. For each property, we also manually reviewed the business with online information, to understand whether the classifications were correct and, in case of disagreement, which one of the two classifications was right, and where the disagreement could come from.

Out of 75 properties, 41 were correctly matched. The remaining 34 were divided as follows:

- for 5 properties, both the COCOA and the SBV were right, although the classification was different. They were businesses that had two components (e.g., a production of a good and the corresponding shop).
- For 1 property, both the SBV and the COCOA were wrong.
- For 9 properties, the COCOA was correct, while the SBV had a definition that was too vague (e.g., *shop* was used to indicate both a *wholesaler* and a *retailer*)
- For 5 properties the COCOA was wrong because the underlying SIC code was too vague.
- For 14 properties the COCOA was wrong because the SIC code was wrongly assigned.

Overall, based on the figures above, the COCOA was correct in 55 out of 75 cases, i.e. 73% of cases.

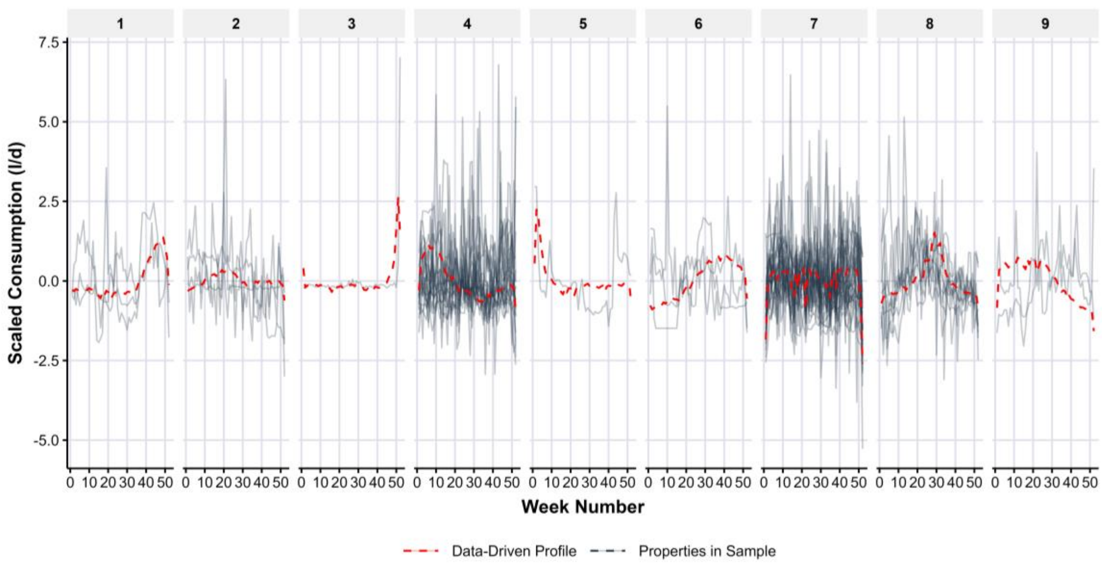
The remaining cases highlighted how the current SIC classification accuracy can be improved, by better working with business owners in understanding it. Most errors were due to the use of the general business industry instead of the building purpose (e.g. *Coffee manufacturer* instead of *Café*, or *Brewery* instead of *Pub*).

5.1.2 Data-driven clustering

The selected properties were then evaluated against the data-driven clustering (shown in Figure 2).

Considering properties in groups, they generally follow the expected patterns as defined by the COCOA Classification, with some deviations (Figure 7). It is expected for individual properties to deviate from the data-driven profiles to some extent, as the aim of the clustering is to simplify complex real-world data, which can contain different levels of variability.

Figure 7. Data-driven profiles (red dotted line) and granular data annual profiles for the 75 properties in the selected sample (grey).



The actual consumption profiles were then compared one by one to the assigned COCOA profiles, and it was decided whether the trend was in general aligned or not. The result is that 41 out of 75 properties had a profile that was well represented by the corresponding COCOA data-driven cluster (55%). While this might seem low, it must be considered that consumption profiles are highly volatile, as they can significantly vary from year to year due to occasional leaks, events like Covid-19 or summer droughts, and other business-specific factors. COCOA data-driven profiles only provide an estimate of the most likely profile for that type of business.

5.2 Benchmarking

5.2.1 *Validation of COCOA Benchmarking*

The COCOA Benchmarking could be compared to the observed consumption. We compared the COCOA Benchmarking to the actual consumption for each of the 75 property and each month, for the years 2022 and 2023.

Figure 8. Estimated consumption against actual consumption for the evaluated properties

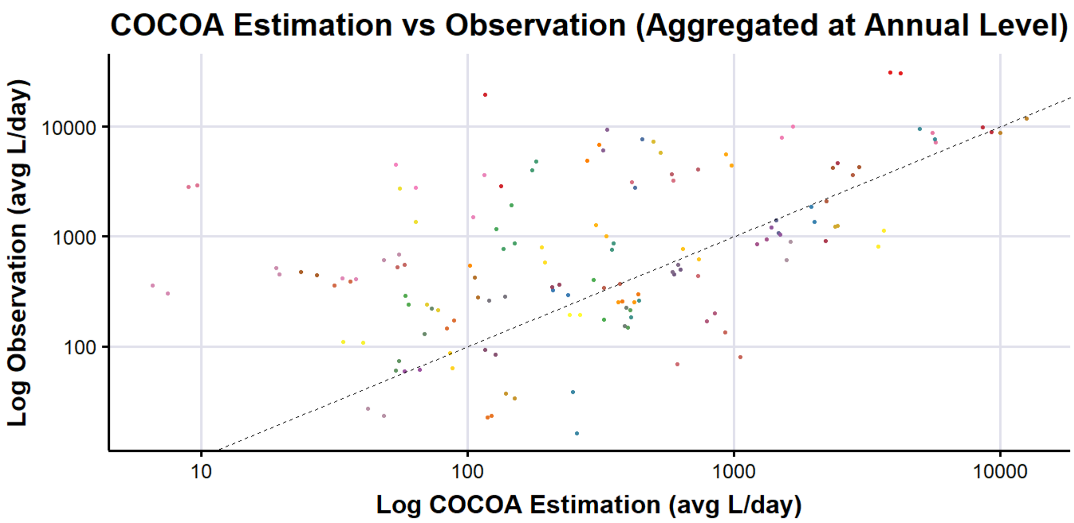


Figure 6 illustrates how the observations compare to the COCOA estimations, one colour for each property; the dashed line is the x=y line. Note that the logarithmic scale is used, to be able to compare properties with significantly different consumptions. Note that a regression proved highly significant (R^2=0.93, p-value < 2.2e-16).

Of the 75 properties, 30 are marked as not outliers, 9 as possible outliers, and 36 as outliers.

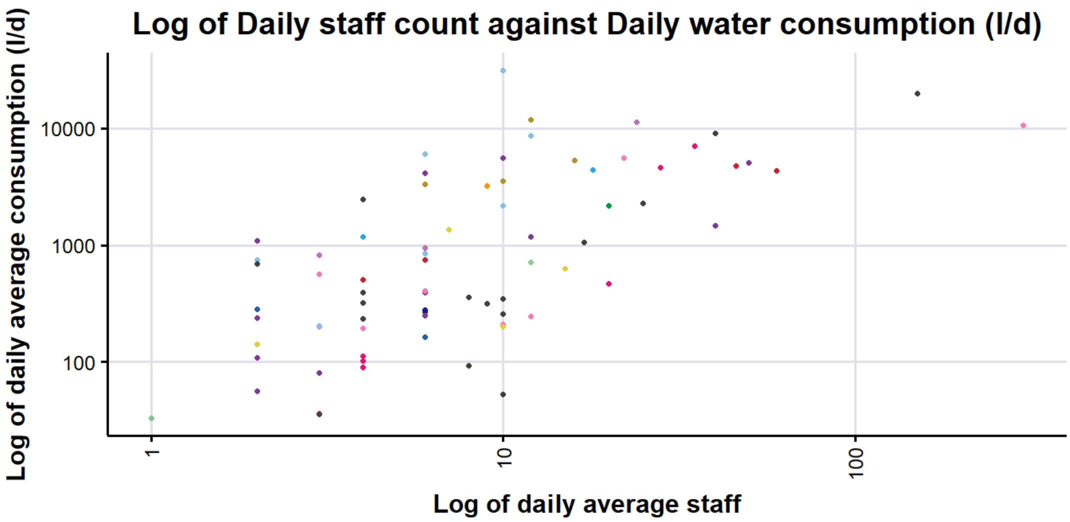
This investigation reveals that there is a general agreement between the expectations and the observations, although variations can be large at times. The COCOA Benchmarking is designed to provide an order of magnitude of the property expected consumption, and the analysis above seem to confirm its suitability for this purpose.

5.2.2 *Exploration of other SBV dataset*

The SBV dataset not only included details about interventions but also contained data on additional property information, such as the average number of staff or the number of water using devices (toilets, urinals, showers). This information could be valuable for enhancing the COCOA Schema. This section presents the most relevant findings of the data exploration.

We compared the available SBV variables to the average consumption. The result is that the number of staff has a good correlation with consumption (see Figure 9), as does information about the number and volume of showers and taps. While this may be biased by the selection of businesses for the SBVs, this finding suggests that these variables could help refine the COCOA Benchmarking, if such data was available widely for NHH properties.

Figure 9: Relationship between staff number and consumption, with logarithmic scales. One colour for each property



A regression between the staff number and average daily consumption in the logarithmic domain has a $R^2 = 0.44$ with p-value < e-10.

6 Case Studies

We have explored three case studies, selected from the 75 SBV used for validation, which provide a practical application of the COCOA schema, and water efficiency opportunities. These case studies help highlight how the COCOA schema can help targeting water interventions or highlight additional aspects that can be improved in the future.

The selection of case studies has been performed manually, not with the aim to have a representative sample of the population, but with the idea of highlighting key learnings.

Table 8 illustrates the result of applying the COCOA Schema to the three case studies.

Table 8: Case study selection, and how they have been classified and benchmarked by the COCOA Schema tool

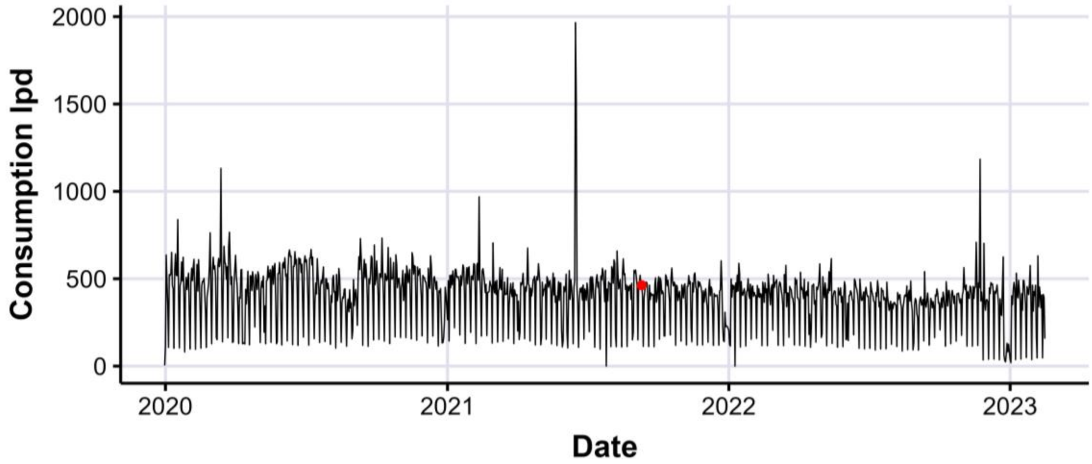
Case study	COCOA Functional Classification	COCOA data-driven classification	Outlier	Uncertain
Case Study 1	Depot	7	Yes	No
Case study 2	Office	7	Yes	No
Case Study 3	Soft manufacturing	4	No	Yes

6.1 Case study 1

The first case study is a 450 m^2 warehouse for a window manufacturer business, where approximately 6 staff work. The building has 5 toilets and 15 taps, including kitchen facilities, no urinals nor showers. The COCOA Classification identified this business as Depot, profile 7.

Figure 10 illustrates daily consumption for this property, and the SBV is highlighted as a red dot. Consumption seems to be consistent between 2020-23, with weekly cycles (regular, frequent drops corresponding to weekends), and a drop around Christmas time. The minimum daily consumption rarely reaches zero, even during weekends. This could mean that there is either continuous flow or a consistent level of activity on site every day. Although not very large, there is a slight reduction in consumption after the SBV.

Figure 10. Daily consumption profile for case study 1. The SBV is identified as a red dot.



The business visit did not find any leaks and only installed two save-a-flush bags, estimated to save approximately 22 l/d. In this case, the high consumption could be due to a specialised process (e.g., glass cutting), or to specific practices (e.g. daily cleaning). Interestingly, weekend consumption significantly dropped toward the end of 2022, suggesting a change in continuous flow processes / practices. However, as there is a peak before, it is also possible that a burst occurred with consequent repair, that also addressed previously undetected leakage. However, as we only have information about the SBV, what happened at the end of 2022 can only be speculation.

Figure 11. Monthly consumption profile for Case study 1, compared to the COCOA expectation

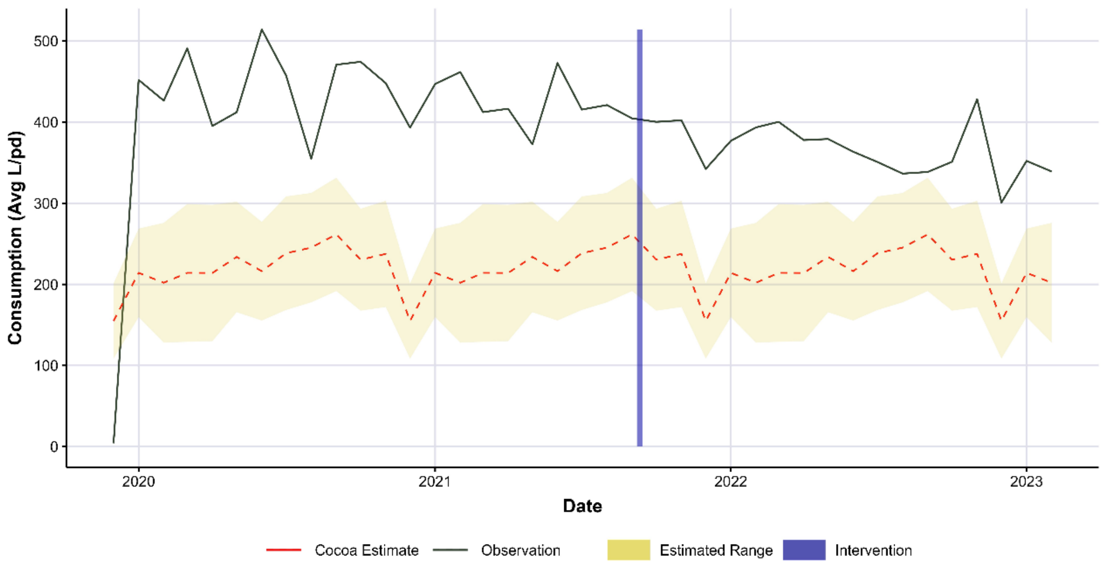


Figure 11 shows how the monthly consumption (black) compares to the COCOA Benchmarking estimations (red dashed line with yellow uncertainty band). The SBV is marked as a blue vertical line. At monthly resolution, it is even clearer that the SBV resulted in some consumption reduction.

Although the saving from the SBV did not bring consumption down to COCOA expectations, this case study showed that:

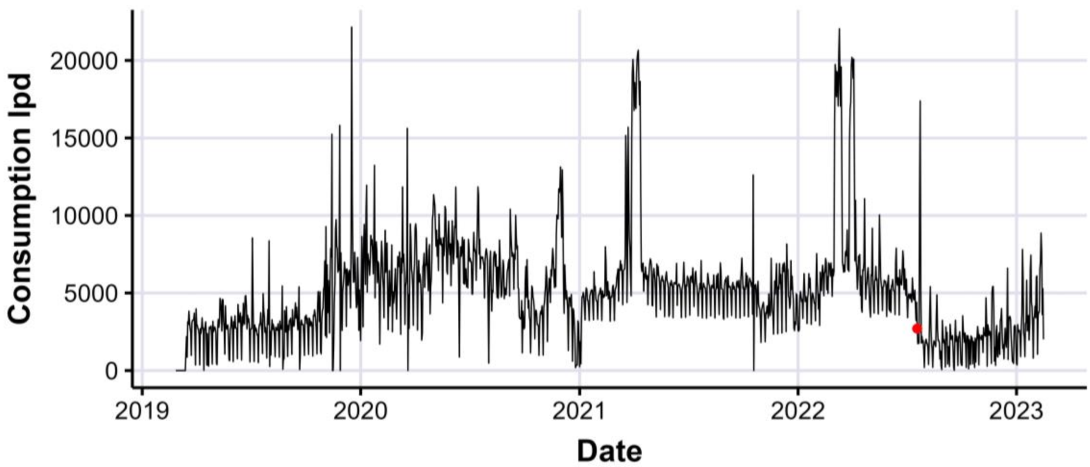
1. The COCOA benchmarking was able to capture the outlier, which had saving potential that the SBV could (at least partially) address.
2. The COCOA Benchmarking described the yearly profile very well, capturing end-of-year drops.
3. There is always an element of volatility in consumption estimation (it would be impossible to account for every single building or business characteristic) and a review of the selected NHH property before targeting intervention is recommended.

6.2 Case study 2

The second case study identified an office building dedicated to postal services. The COCOA classifies it as an Office, with profile 7. The office employs 50 staff, has an average of 50 visitors a day, and it is open 16 hours a day. There are 17 toilets, 2 urinals, 32 taps (including kitchen), and no showers.

Figure 12 illustrates daily consumption for this property, and the SBV is highlighted as a red dot.

Figure 12. Daily consumption profile for case study 2. The SBV is identified as a red dot.



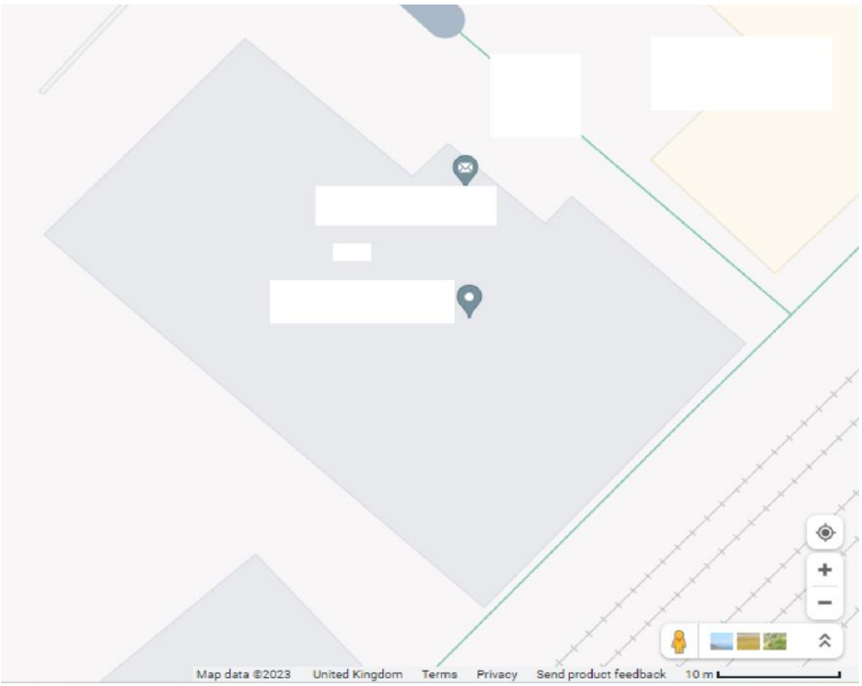
From Figure 12 we observe that the SBV resulted in a significant consumption drop, although then consumption seems to increase later in 2023 again. Consumption exhibits large peaks around easter time in 2021 and 2022, not present in 2020, suggesting that activities in this building were affected by Covid-19 in some form. Additionally, we see a much lower consumption before 2019. Overall, we cannot consider consumption for this building stationary.

At first, the COCOA benchmarking was far from the actual consumption. For this building, we had an ABP area of 152 m^2, which does not match with the SBV information available, resulting in average yearly consumption between 0.15 and 0.21 m^3/day. Before the SBV visit we observe a consumption of 5.52 m^3/day, approximately 20 times more than expected.

From an internet search, we found out that the building is much larger than what ABP states in this case (probably due to a mismatch of UPRN – business relationship).

Using Google Maps, we could estimate a floor area of approximately 800 m². Additionally, from Google Street view, we could observe that the building has 3-4 floors.

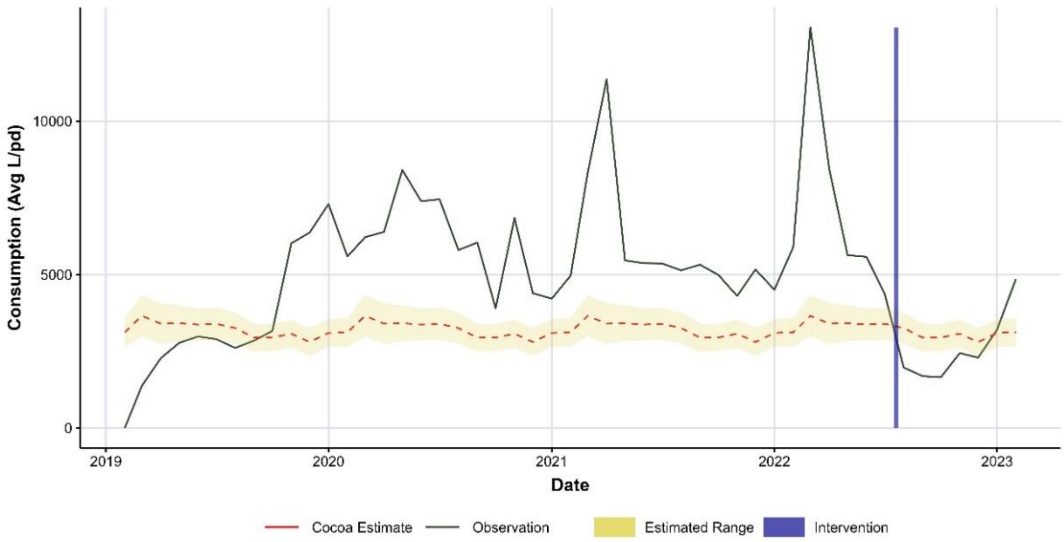
Figure 13. Google Maps view of the building in Case study 2. Using the scale on the bottom-right, we could estimate the building area.



So, we review the COCOA benchmarking using an area of 2,800 m². Using these settings, the building is still classified as a possible outlier, but ideal consumption is now estimated to be between 2.69 and 3.72 m³/day. Figure 14 shows how the monthly consumption (black) compares to the COCOA Benchmarking estimations (red dashed line with yellow uncertainty band). The SBV is marked as a blue vertical line.

During the SBV, multiple interventions were carried out: urinal control devices were fitted, a dripping tap was identified, and multiple leaking toilets were fixed. This resulted in the average consumption after the visit to fall to 2.5 m3/day, below COCOA expectations.

Figure 14. Monthly consumption profile for Case study 2, compared to the COCOA expectation



Some key findings from this case study are:

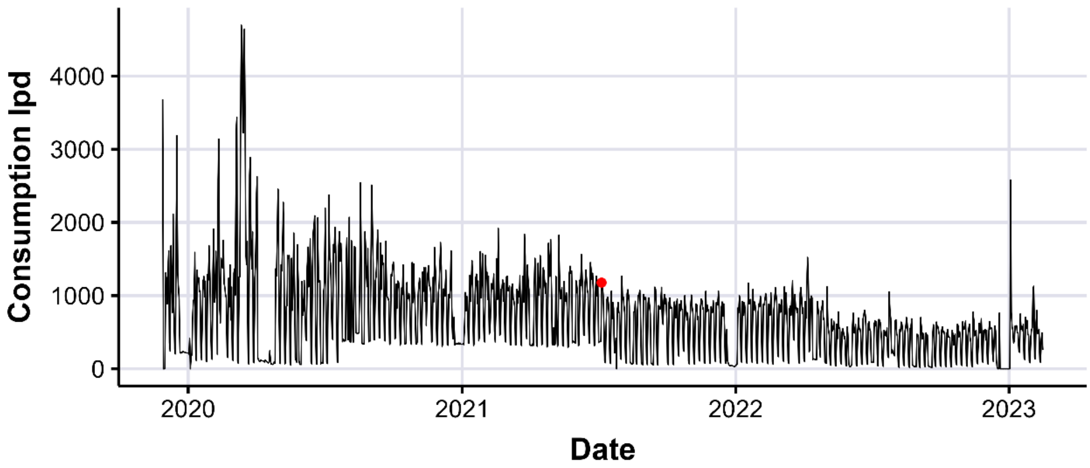
- The COCOA Classification is correct.
- The COCOA Benchmarking is reasonable, if supplied with correct data.
- The MasterMap floor area was incorrect in this case, highlighting that supporting information is key to use the COCOA Schema.
- Accuracy of the MasterMap floor area is untested and there is a risk in relying on one source. It is recommended to consider additional sources of data (e.g. EPC).
- Currently, the COCOA Schema does not account or multiple floors. This property had 3 floors, and after the floor area estimation was adjusted for correct footprint and number of floors, the COCOA Benchmarking could produce a much more realistic range of consumption.
- While this building was classified as an outlier before the SBV, it dropped below COCOA expectation after the SBV (at least for 6 months.
- After 6 months consumption increased again. We cannot know exactly the reasons, but this highlights that interventions are never permanent, and it is recommended to continuously monitor commercial properties.
- Although the scale was not right, the COCOA benchmarking identified the correct annual profile, with a peak in spring.

6.3 Case study 3

The third case study is a veneer manufacturer and retailer. It has an average of 12 employees and 10 visitors a day. COCOA has classified it into soft manufacturing, cluster 4. Note that this case study was not picked up as an outlier by the COCOA Benchmarking.

Average daily consumption for this property for the period 2020 up to 2023 is shown in Figure 16 (black line). This image shows that for the period June 2020 to June 2021 consumption raised by a continuous flow which lasted a year (approximately 0.31 m³/d). The plot also shows that baseline consumption for this property returned to lower levels following SBV water efficiency interventions (red dot).

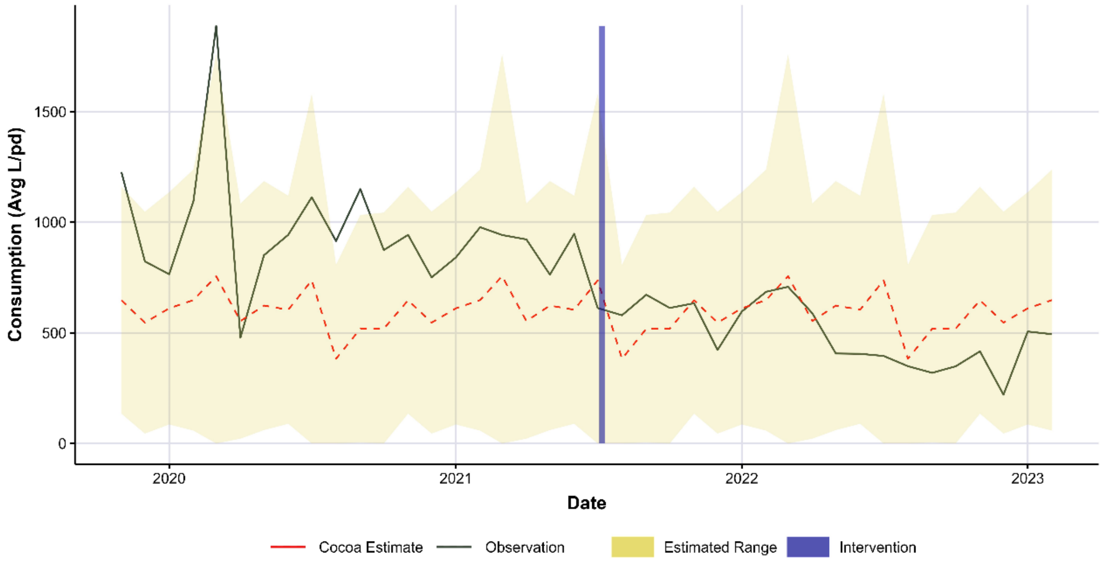
Figure 15. Average daily consumption for case study 1 for the period 2020 - 2023



During the smart business visit, it was observed that the business does not have any showers but has toilets and urinals in place. No leaks were found. The urinal flushing at the building was not controlled, and during the smarter business visits urinal controllers were fitted along with toilet cistern displacement devices. The cistern displacement devices are designed to reduce the amount of water used per flush by approximately 1.2 litres, while the urinal controls are designed to stop the continuous flushing when they are not used (e.g., at night). The installation of these devices had an immediate effect in reducing the continuous flow, but also in reducing the average daily flow.

Figure 16 compares the actual monthly average consumption for this property (black line) against the COCOA central estimated values (red dotted line). It can be observed here that the actual consumption of the property consistently remains above the COCOA central estimates until the intervention point (vertical blue line) when consumption falls below COCOA estimates. However, given the wide uncertainty bands, the property is not highlighted as an outlier by the COCOA Benchmarking tool (Green in Table 8).

Figure 16. Monthly average consumption for case study 1 compared to the COCOA estimate.



This case study would have not been picked up by the COCOA Benchmarking but highlights how the use of continuous flow may be a great indicator for water saving interventions in small businesses. However, the industry it is not at a point where continuous flow in NHH property is understood sufficiently well to be modelled reliably, as larger businesses are likely to have genuine continuous use. More research needs to be done and could potentially result in an improvement of the COCOA schema.

Interestingly, this case study shows that the SBV ha a material impact.

6.4 Case studies: learnings

By selecting these case studies, we could highlight some key learnings:

1. The COCOA Classification was correct for all case studies.
2. The COCOA Benchmarking can capture many outliers, which have saving potentials. Business visits can capitalise on this potential, at least partially.
3. The COCOA Benchmarking can describe yearly profiles well.
4. The COCOA Schema requires correct SIC and floor area data to be able to work properly. Adjusting floor area by the number of floors seems to result in positive improvements.
5. Both SIC Classification and MasterMap floor area showed some inaccuracies: it is important to verify the information before relying on the outcomes and it is recommended to use additional sources of data for validation.
6. There is always an element of volatility in consumption estimation (it would be impossible to account for every single building or business characteristic) and a review of the selected NHH property before targeting intervention is recommended.
7. Interventions are never permanent, and it is recommended to continuously monitor commercial properties.

8. The use of continuous flow may be a great indicator for water saving interventions in small businesses. However, the industry it is not at a point where continuous flow in NHH property is understood sufficiently well to be modelled reliably, as larger businesses are likely to have genuine continuous use. More research needs to be done and could potentially result in an improvement of the COCOA schema.
9. SBV had a material impact in all three case studies, although of variable magnitude. We observed that targeting larger businesses often results in larger saving potential.

7 Conclusions

This project has developed an industry specific segmentation schema for non-household (NHH) customers. The schema combines Standard Industry Classifications (SIC) with expert knowledge on water consumption and consumption data from smart metering systems.

The resulting Commercial Consumption Analysis (COCOA) Schema has been tested on smart meter data and further validated using data from 75 business water audits, which has demonstrated the schema's value in using water consumption benchmarking in identifying and driving targeted water efficiency savings.

Our global literature review identified that most datasets available for NHH clustering, such as SIC, are based on industry groupings and are not based on water consumption. Those that have used water consumption have required specific audits or questionnaire data from each NHH customer. This project has used industry available data plus a large sample of smart metering data to derive the segmentation schema, which is the first time this has been achieved.

The COCOA Schema is built from:

- a functional water use classification developed by grouping SIC codes, on the basis of their industry nature and functional water usage characteristics.
- a data-driven classification was developed by studying consumption profiles from smart metering data resulting in 9 unique clusters of NHH seasonal water use.
- building area data sourced from the Ordnance Survey (OS) Address Base Premium (ABP) and Master Map (MM) data package.

These have been combined through modelling to estimate the expected monthly consumption (within defined uncertainty bands) for individual NHH properties.

The COCOA Schema can be used to benchmark NHH properties' consumption, identify consumption outliers and to segment NHH properties for water demand modelling.

During the development of the schema, we identified that there was a lack of coverage of SIC classification in the NHH sector. Only approximately 30% of properties have a SIC Class assigned, while it is missing for 70% of properties. Despite this low coverage we concluded that SIC was the best classification for use in this study as it is the most widely available and detailed form of classification.

A potential alternative to SIC was identified in this project for classifying NHHs, and this should be investigated further to determine if it can increase the effectiveness of the COCOA Schema.

The COCOA Schema can be used without a specific SIC code, but with a greater uncertainty.

The COCOA Schema has been validated using data from smart business visits (water audits). The key conclusions from the validation are:

- The functional use classification was correct for 73% of the tests, which is a good result considering the inherent uncertainty in allocating industry codes to buildings.
- The consumption derived clustering was accurate for 55% of the tests, which illustrates the inherent variability in NHH consumption, but the clustering does improve the modelled estimate of consumption and provides a monthly profile.
- Continuous use could be a very valuable addition to the analysis of total consumption, but the more understanding of continuous use in the NHH sector is required.
- Most of the failures to correctly identify the functional use classification were due to incorrect classification of multi-use buildings, incorrect allocation of SIC codes or codes that were too vague. Data quality is crucial for COCOA and data quality improvement efforts and data validation can be key.
- Consumption estimates were challenging for properties with multiple floors.

8 Recommendations

During the project, we have documented a significant number of post-project recommendations. These include those discussed at monthly PSG meetings with all stakeholders. We have provided a succinct summary of these below, recommending the following:

- **A further investigation into external data sources.** During the project, we identified several external data sources that can hold key metrics to help with pertinent scalars. With any external dataset, there are costs and accessibility considerations, and the fact there would need to be unilateral commitment between stakeholders to use any recommended external dataset. We would like to highlight the ONS IDBR database as an example. This data source contains detailed information about businesses in the UK associated to either VAT or PAYE information. This include turnover or employee number, which could be extremely useful in evaluating business size and would likely improve drastically the use of the building area as pertinent scalar. However, the IDBR dataset is currently only accessible to Government or research entities and even in these cases access is limited. We recommend liaising with the Environment Agency (EA) and the Department for Environment, Food & Rural Affairs (DEFRA) to gather support in negotiating access to this database for the whole UK water industry in some form. Another option could be to consider mandatory data collection on business size information (e.g., number of employees) by retailers.

- **Improve classification and data matching by using addresses, postcodes, or coordinates.** Currently we base the matching on the use of Supply Point IDs (SPIDs), but these are often incomplete and result in a poor matching. Other forms of matching based on less unique identifiers would require more time but could potentially greatly improve the SIC classification.

- **Investigate how uniform or diverse SIC Codes are.** This project worked with SIC Codes under the assumption that properties within each SIC Code behave uniformly, but we are aware this is not the case.

- **Investigate most heterogeneous codes.** Some SIC Codes showed heterogeneous behaviours and matched with multiple functional and data-driven clusters. This suggests that water usage behaviour within the SIC Codes are highly variable and there may be a need to split SIC codes.

- **Investigate and improve accuracy of SIC Codes:** our modelling assumes that SIC classifications are correct, but we are aware that this may not always be the case and we observed this first hand through the validation of the COCOA Classification. At times, properties may be mis-labelled, or the classification could be out of date. We recommend MOSL and the retailers to invest in improving the quality and coverage of SIC Classification for all NHH properties.

- **Investigate further daily consumption profiles in NHH properties.** This work considered the use of daily profiles but decide not to use them as the uncertainty is wide and they are strongly driven by business size too. However, the use of daily profiles is very useful to better understand night use, continuous flow and potentially losses.

- **Better understand and use continuous flow in NHH water efficiency.** Continuous flow has proven to be a precious indicator of water saving potential but has not been used in this work due to the poor understanding of when genuine continuous use is expected (or not) in NHH properties. This is something that the industry should study more and capitalise on.

- **Split continuous flow into continuous use, wastage, and leakage.** At Artesia we have developed some advanced algorithms to separate continuous and variable flow, and then split the continuous flow into use and losses for household properties. However, the application to non-household properties is further complicated by the more variable size of consumption and greater flow variability. It is recommended to look at continuous flows to identify and better target properties at higher saving potential, but also understand how to split continuous flow into use or losses.

- **Investigate the relationship between granular and CMOS ADC data.** In this work MOSL ADC data has been used only as an indicator of the business size, to separate outliers. Currently MOSL ADC is poorly understood, as it is a mix of estimated and actual consumption readings, with readings made available any time between one and 12 months in the past. For this reason, it is difficult to understand the uncertainty associated to this information. However, working with MOSL and retailers to improve and better understand this data would be extremely beneficial to be able to apply methodologies similar to the ones presented here to properties that do not have granular data.

- **Investigate potential and limitations of building area as a pertinent scalar.** As detailed throughout this report, we have used building area from the MasterMap dataset as a pertinent scalar. Going forward, we recommend to study strengths and weaknesses of this metric. This work has shown that its accuracy may need some testing. The use of EPC data could help expanding the number of commercial properties with associated building area. However, we would also recommend studying those cases where a business is associated to multiple buildings, or where one building is associated to multiple businesses. Verisk support could be extremely beneficial in this. Potentially, the use of other data sources in conjunction with building area could improve the understanding of consumption in NHH properties.

- **Investigate criticality.** Criticality has been highlighted by the project steering group as an element of interest. While in the limits set by this project timeline and budget, we could not investigate this type of business classification, it is recommended to look at what form of data could help understanding the criticality of water supply to a business and the criticality of the business itself.

- **Investigate the interannual variability of NHH consumption.** Due to the limited granular data available, we could only use a limited number of years to develop the classification and the benchmarking. However, we are aware that the post-covid years have been affected by many non-stationary changes (e.g. drought, cost-of-living crisis, hybrid working) and we recommend using longer data series in the future to better understand consumption variability on a longer term.

- **Plan for a more regulated future.** From a wider industry perspective, NHH water consumption will have even more focus, with the emergence of ever tighter regulatory water use targets expected and businesses striving towards net zero commitment.

- **Plan for a future based on smart metering.** This work has been made possible because some industry leaders in the field of smart metering have come together and shared granular data for the benefit of the wider water industry. The benefits that smart meters can bring go much beyond what highlighted in this project and can help us understand customers' behaviour at different temporal scales in a responsive way. This makes wholesalers and retailers more resilient to transient events, better equipped target waste, reduce consumption, and improve efficiency.

- **Consider mandatory usage audits.** As site visits are ultimately recognised as an important source of truth, verifying the type of business, its size, its water usage drivers and practices, and the presence of losses, the water industry should consider a more radical approach to them. Mandatory usage audits could be proposed by the regulators, asking retailers to perform them for each new client, and at regular intervals, to ensure consumption expectations in every NHH property are verified from the start.

- **Testing the COCOA Schema and collect feedback.** As the COCOA Schema is an innovative tool, it will take some extensive use across the industry to evaluate strengths and weaknesses. By regularly collecting feedback, at least from water wholesalers and retailers, it will be possible to target the most critical points with future work.

- **Produce lookup tables for other classification schemas.** As SIC Codes are expected to cover only a 30% of NHH properties in the MOSL database, it is recommended to develop lookup tables that link different classification schemas (ABP, VOA, or SIC divisions) to the functional classification of the COCOA Schema. This will allow to use the COCOA schema with the standard profiles even when SIC Codes are not available (although at a lower accuracy).

Appendix A COCOA Schema Definition

The Commercial Consumption Analysis (COCOA) schema is formed of two elements: the COCOA Classification and the COCOA Benchmarking.

8.1 COCOA Classification

The COCOA Classification is made of two interacting classifications: the functional use classification and the data-driven classification. Table A-1 illustrates the 32 categories for the functional use classification, their definitions, and which SIC codes are included in each.

The data-driven classification identifies 9 yearly profiles that NHH properties can follow, and it further splits each functional use classification group into sub-groups based on different yearly profiles observed. The COCOA Classification, including both the functional use and the data-driven classifications, is summarised in Table A-2.

Table A-1: Functional classification produced as result of Stage 1

Functional use classification	Description and rationale	Functional Water Use
Accommodation	Domestic style consumption - hotels, guest houses	Domestic, Leisure, Washing
Arable	Plant-based agriculture with seasonal irrigation	Irrigation, Washing
Clinic	Medical day care	Domestic, Unknown
Coolant	Water used to cool processes - computer servers, power generation, possibly 24/24hr.	Cooling
Depot	Work is carried out primarily off site with low mid-day activity. Premises used for storage, maintenance of equipment, preparation of materials and office work, possibly washing of equipment.	Office, Washing
Drink processing	Physical processing brewing and packaging of drinks for distribution.	Office, Washing, Embodied
Food processing	Physical processing, fermenting, cooking, and packaging of food for distribution requiring high cleaning demands.	Office, Washing, Processing
Furnace	Dominant process involves heating and melting materials at high temp. Possibly low water use to energy consumption ratio.	Office, Cooling, Unknown
Further education	Education for over 18s with more erratic holiday and office hours than under 18s.	Office, Domestic, Leisure
Hard manufacturing	Manufacturing using insoluble materials that are not fibrous possibly resulting in lower water process demands.	Office, Unknown

Hospital	Residential or 24hr medical treatment facility with process augmented residential style use.	Domestic, Washing
Industrial chemistry	Where chemical processes are dominant, requiring high washing and possibly coolant demands.	Office, Washing, Processing, Coolant
Irregular school	Education that may not conform to school hours.	Domestic, Unknown
Lab	Small scale technical operations with unpredictable demand.	Office, Washing, Processing, Unknown
Laundry	Textile cleaning. High demand per sq. foot.	Washing
Livestock	Farming that is primarily livestock. Erratic water use that may not conform to weather signals, possibly high washing demands.	Washing, Irrigation, Embodied
Military	Manned depots with military or emergency response personnel. Expect erratic person-driven 24/24hr demand.	Domestic, Washing, Unknown
Mineral particulates	Processing of dry materials with water for manufacturing or dust suppression. Possibly high consumption per square foot and seasonal signal.	Washing, Dust Suppression, Processing, Embodied
Mining	Extraction of minerals from the ground. Supplies are prone to bursts from machinery damage, possibly seasonal dust suppression or washing.	Washing, Dust Suppression, Processing
Office	Indoor non-manufacturing workplace catch all. Domestic day time style person-driven consumption.	Office
Residential	Residential care. Possibly between hotel and hospital style consumption.	Domestic, Washing, Leisure
Retail	Shop and showrooms open to the public. Moderate cleaning and washroom demand.	Office, Washing
Salon	Hair and beauty salons. High use per square foot.	Domestic, Washing
School	School for children. Predictable seasonality and person driven consumption. Distinction between residential and day schools is not made in SIC codes.	Domestic, leisure
Social	People orientated business that also operate outside office hours. Restaurants, pubs, gyms, theatres with potential night use.	Office, Processing, Leisure
Soft manufacturing	Manufacturing with textiles with potentially high washing demands.	Washing, Processing
Terminal	Transport hubs open to the public, potentially 24/24hr seasonal people driven demand with additional cleaning demands.	Office, Washing

Textile manufacture	Manufacturing textiles not products. Potentially high process demand.	Washing, Processing
Tourism	Tourist attractions. Holiday related seasonality and unpredictable demand.	Office, Leisure
Waste processing	Potentially highly variable water use depending on processes used.	Processing, Washing, Dust Suppression
Wholesale	Warehousing and sale of goods. Potentially low activity levels per sq foot.	Office
Workshop	Bespoke, small-scale, repair and manual work. Unpredictable demand depending on processes	Office, Unknown

Table A-2: COCOA Classification: relationship between final data-driven classification from Stage 2, functional classification from Stage 1, and SIC codes and classes.

Functional Classification	Data Driven Classification	SIC Codes
Accommodation	6	55100 55200 55201 55202 55300 97000
Accommodation	8	55900
Accommodation	9	55209
Arable	1	01110 01120 01130 01140 01150 01160 01210 01220 01230 01240 01250 01260 01270 01280 01300 01610
Arable	2	01190
Arable	8	01290
Clinic	2	86210
Clinic	7	75000 86220 86230 86900 96030
Coolant	1	35110 63110
Depot	1	42110 42210 42220 52241
Depot	2	33170 42120 43310 49390 49420 51210 77110 81222
Depot	3	09100 35220
Depot	4	33160 81291
Depot	5	35300 81221
Depot	6	43120 49320
Depot	7	01700 02100 02200 02300 02400 03110 03120 03210 09900 33150 33200 35120 35130 38120 38310 41200 41201 41202 42130 42910 42990 43130 43210 43220 43290 43320 43330 43340 43341 43342 43390 43910 43990 43991 43999 49200 49410 49500 50200 50400 51220 52210 52211 52212 52213 52219 52220 52230 52240 52242 52243 52290 53100 53200 53201 53202 56210 77120 77210 77310 77320 77340 77341 77342 77350 77351 77352 77390 81100 81210 81220 81223 81229 81290 81299 85530 85600

Depot	8	38110 81300
Depot	9	43110
Drink processing	5	11070
Drink processing	8	10320 11010 11020 11030 11040 11050
Food processing	1	01630 01640 10120 10130 10310 10390 10410 10420 10510 10511 10512 10519 10520 10610 10611 10612 10620 10720 10730 10810 10820 10822 10830 10831 10840 10850 10860 10910 10920 11060
Food processing	2	10890
Food processing	4	10110
Food processing	6	12000
Food processing	7	10710
Food processing	8	10200 10832
Food processing	9	10821
Furnace	4	23510 23520 24100 24420 24430 24440 24450 24510 24520 24530 24540
Furnace	7	23130
Furnace	8	23110
Furnace	9	24410
Further education	7	85410
Hard manufacturing	1	27120
Hard manufacturing	2	16100 25500 30110 30920 32409
Hard manufacturing	3	29201
Hard manufacturing	4	22220 23190 27510 28220 28921 29100 29310 31020
Hard manufacturing	5	26309 31010
Hard manufacturing	6	25120
Hard manufacturing	7	16220 16240 18200 18201 18202 18203 22110 22190 22210 22230 22290 23120 23140 23200 24200 24310 24320 24330 24340 25110 25210 25290 25300 25400 25620 25710 25720 25730 25910 25920 25990 26110 26120 26200 26300 26301 26400 26510 26511 26512 26513 26514 26520 26600 26700 26701 26702 26800 27110 27310 27320 27330 27400 27520 27900 28110 28120 28130 28131 28132 28140 28150 28210 28230 28240 28250 28290 28300 28301 28302 28410 28490 28910 28920 28922 28923 28930 28940 28950 28960 28990 29200 29202 29203 29320 30120 30200 30300 30400 30910 30990 32110 32300 32400 32401 32910 32990
Hard manufacturing	8	25940
Hard manufacturing	9	25930
Hospital	4	86100 86101 86102
Industrial chemistry	1	21100
Industrial chemistry	2	36000
Industrial chemistry	4	19100 19200 19201 19209 20110 20120 20130 20140 20150 20170 20200 20300 20301 20302 20410 20412 20420 20510 20520 20530 20590 20600 21200 24460 27200 35210

Industrial chemistry	7	20411
Industrial chemistry	9	20160
Irregular school	7	85520 85590
Irregular school	9	85510
Lab	6	72110
Lab	8	72190
Laundry	6	96010
Livestock	2	01629
Livestock	4	01410 01420 01430 01440 01460 01470 01490 01500 01620 01621 03220 46230 91040
Livestock	5	01450
Military	2	84240
Military	4	84220 84250
Mineral particulates	1	17120
Mineral particulates	3	23440
Mineral particulates	6	17230
Mineral particulates	7	25610
Mineral particulates	8	17110 17210 17211 17219 17220 17240 17290 23310 23320 23410 23420 23430 23490 23610 23620 23630 23640 23650 23690 23700 23910 23990
Mining	7	08120
Mining	8	05100 05101 05102 05200 06100 06200 07100 07210 07290 08110 08910 08920 08930 08990
Office	1	63120 64202 68201 79110 79909 94120
Office	2	46180 59131 61900 62020 62030 63990 65300 66220 66290 69109 70221 73110 73120 78109 82302 84120 90020 91012 94110 94910
Office	3	64203 74203
Office	4	59200 64205 64991 69101 84110
Office	5	58190 61300 64303
Office	6	58110 64910 64921 74901
Office	7	35140 35230 41100 46110 46120 46130 46140 46150 46160 46170 46190 58120 58130 58140 58141 58142 58210 58290 59120 59130 59133 60100 61100 62010 62011 62012 62090 63910 64110 64190 64191 64192 64200 64201 64209 64300 64301 64302 64304 64305 64306 64920 64922 64929 64990 64992 65110 65120 65200 65201 65202 66110 66120 66190 66210 66300 68100 68200 68202 68209 68310 68320 69100 69102 69200 69201 69202 69203 70100 70220 70229 71110 71111 71120 71121 71122 71129 72200 73200 74100 74200 74201 74202 74209 74300 74900 74902 74909 77400 78100 78101 78200 78300 79120 79900 79901 80100 80200 80300 82190 82200 82300 82301 82910 82911 82922 82990 84130 84210 84230 84300 88100 88990 91010 91011 91020 94200 94920 96090 99000
Office	8	64204 70210 71112 94990
Office	9	59132 61200 64999 82110

Residential	4	87200 87300 87900
Residential	9	87100
Retail	1	47820
Retail	2	47240 47250 47260 47510 47530 47540 47591 47599 47640
Retail	4	45320 47110 47190 47230 47290 47420 47421 47429 47430 47520 47590 47620 47630 47650 47710 47720 47721 47722 47730 47740 47750 47770 47780 47782 47789 47790 47799 47810 47890 77220 77290 77330
Retail	7	47410 47610 47741 47749 47781 47910 47990 77291 77299
Retail	8	47220 47760 47791
Retail	9	47210 47300
Salon	2	96020
School	1	85320
School	7	85100 85200 85310 88910
Social	2	56301
Social	4	56100 56101 56102 56103 56290 56300 56302 59140 85420 85421 92000 93190 93191 93199
Social	5	90010
Social	6	85422 93290
Social	7	90040 93130
Social	8	93110 93120
Social	9	96040
Soft manufacturing	2	14132 15200
Soft manufacturing	4	13200 13920 13921 13930 13931 13939 14120 14130 14131 14140 14141 14142 14200 14310 15120 16210 16290 18110 18120 18121 31030
Soft manufacturing	5	13100 13923 18130
Soft manufacturing	7	13300 13922 14110 14190 18129
Soft manufacturing	9	14390
Terminal	2	49311
Terminal	4	49100 49310 49319 51100 51101 51102
Terminal	8	50100 50300
Textile manufacture	2	13910 13940 13950 13960 13990 15110
Tourism	4	91030
Tourism	6	93210
Waste processing	7	37000 38210 38220 38320 39000 46770
Wholesale	1	46320 46760
Wholesale	2	45112 46342 46360 46380 46620
Wholesale	3	46711
Wholesale	4	45111 46330 46491 46730
Wholesale	6	46499
Wholesale	7	45110 45190 45310 46210 46240 46310 46340 46350 46370 46390 46410 46420 46430 46431 46439 46440 46450 46460 46470 46480 46490 46510 46520 46610 46630 46640 46650 46660 46690 46710 46719 46720 46740 46900 52100 52101 52102 52103

Wholesale	8	46220 46341 46750
Workshop	2	95230
Workshop	6	98100
Workshop	7	16230 18140 31090 32120 32130 32200 32500 33110 33120 33190 45200 59110 59112 59113 60200 71200 82920 90030 95110 95120 95210 95220 95240 95250 95290 98200
Workshop	8	33130
Workshop	9	33140 45400 59111

8.2 COCOA Benchmarking

Through the use of consumption models, we produced the COCOA Benchmarking table, reported in Table A-3. This table shows how many litres per property per day per square meter a property within each COCOA category is expected to consume, and it also shows the level of uncertainty associated to the coefficients, in the same units. The COCOA Benchmarking table can be easily used to estimate the order of magnitude of consumption for a property by multiplying the coefficient (and the uncertainty bands) by the building area. In case the SIC code is not known, the functional use classification can be estimated (either by expert judgement, business name, or other forms of classification like VOA or ABP) and the standard data-driven classification, highlighted in bold font, can be chosen.

Table A-3: COCOA Benchmarking: estimated consumption in l/prop/day/m², by functional and data driven classification and month. Note that figures are expressed as a coefficient ± uncertainty, expressed in the same units.

Functional classification	Data driven classif.	January	February	March	April	May	June	July	August	September	October	November	December
Accommodation	**6**	**11.54±5.16**	**13.23±5.76**	**13.03±6.54**	**14.49±7.01**	**13.13±6.48**	**13.42±6.5**	**14.94±7.86**	**15.66±7.88**	**14.38±6.81**	**13.96±6.55**	**13.26±6.2**	**13.46±6.28**
Accommodation	8	21.68±24.7	21.91±24.19	19.36±21.32	20.61±23.27	21.71±25.02	20.73±24.16	21.37±24.42	20.2±22.4	23.31±27.31	24.99±29.45	24.55±28.93	25.67±30.05
Accommodation	9	13.13±5.47	14.59±5.92	14.12±6.49	15.52±6.91	14.52±6.64	14.61±6.47	15.98±7.43	16.39±7.33	15.69±6.81	15.58±6.76	14.9±6.43	15.24±6.54
Arable	**1**	**1.3±2.4**	**1.68±4.14**	**1.17±4.61**	**1.36±3.04**	**1.82±1.85**	**3.53±3.02**	**4.02±4.08**	**3.44±3.65**	**1.76±2.07**	**1.52±2.37**	**1.48±3.09**	**2.43±2.66**
Arable	2	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Arable	8	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Clinic	2	2.18±1.11	2.2±1.02	1.9±1.41	1.18±0.92	2.43±1.4	2.42±1.37	2.58±1.45	3.12±1.74	2.77±1.45	2.58±1.28	2.41±1.18	2.42±1.49
Clinic	**7**	**2.18±1.11**	**2.2±1.02**	**1.9±1.41**	**1.18±0.92**	**2.43±1.4**	**2.42±1.37**	**2.58±1.45**	**3.12±1.74**	**2.77±1.45**	**2.58±1.28**	**2.41±1.18**	**2.42±1.49**
Coolant	**1**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Depot	1	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1
Depot	2	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1
Depot	3	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1

Depot	4	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1
Depot	5	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1
Depot	6	0.47±0.12	0.45±0.16	0.47±0.19	0.47±0.19	0.52±0.15	0.48±0.13	0.53±0.16	0.54±0.15	0.58±0.15	0.51±0.14	0.53±0.15	0.34±0.1
Depot	**7**	**0.47±0.12**	**0.45±0.16**	**0.47±0.19**	**0.47±0.19**	**0.52±0.15**	**0.48±0.13**	**0.53±0.16**	**0.54±0.15**	**0.58±0.15**	**0.51±0.14**	**0.53±0.15**	**0.34±0.1**
Depot	8	1±0.74	1.16±0.77	1.29±1.36	0.84±0.9	1.43±1.15	0.83±0.65	1.18±0.89	0.98±0.76	1.23±1.02	0.72±0.66	1.14±0.87	0.85±0.64
Depot	9	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Drink processing	5	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Drink processing	**8**	**7.2±1.75**	**8.19±1.95**	**10.33±3.44**	**9.45±3.55**	**11.68±3.53**	**11.52±3.17**	**11.32±3.09**	**10.47±2.91**	**9.48±2.44**	**8.88±2.15**	**8.7±2.02**	**8.56±2.05**
Food processing	1	2.85±2.66	3.26±2.88	2.92±2.61	2.52±2.31	2.62±2.04	2.98±2.58	2.64±2.23	3.01±2.47	3.21±2.69	3.46±3.13	3.46±3.43	3.48±4.21
Food processing	2	10.72±10.09	9.86±8.9	7.64±6.97	8.03±7.67	8.19±7.27	8.13±7.36	9.91±9.7	9.77±9.17	9.33±8.25	10.42±9.8	11.05±10.63	11.2±10.68
Food processing	4	10.72±10.09	9.86±8.9	7.64±6.97	8.03±7.67	8.19±7.27	8.13±7.36	9.91±9.7	9.77±9.17	9.33±8.25	10.42±9.8	11.05±10.63	11.2±10.68
Food processing	6	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.2	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Food processing	**7**	**13.78±22.87**	**12.6±20.58**	**9.11±15.07**	**9.69±16.36**	**9.99±16.31**	**10.14±16.14**	**13±21.09**	**12.63±20.79**	**11.95±19.2**	**13.52±21.77**	**14.4±23.15**	**14.56±23.91**
Food processing	8	10.72±10.09	9.86±8.9	7.64±6.97	8.03±7.67	8.19±7.27	8.13±7.36	9.91±9.7	9.77±9.17	9.33±8.25	10.42±9.8	11.05±10.63	11.2±10.68
Food processing	9	10.72±10.09	9.86±8.9	7.64±6.97	8.03±7.67	8.19±7.27	8.13±7.36	9.91±9.7	9.77±9.17	9.33±8.25	10.42±9.8	11.05±10.63	11.2±10.68
Furnace	**4**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Furnace	7	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Furnace	8	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Furnace	9	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Further education	**7**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Hard manufacturing	1	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hard manufacturing	2	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hard manufacturing	3	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hard manufacturing	4	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28

Hard manufacturing	5	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hard manufacturing	6	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hard manufacturing	**7**	**0.87±0.32**	**0.96±0.37**	**0.93±0.47**	**0.81±0.42**	**0.91±0.47**	**0.84±0.39**	**0.94±0.43**	**0.98±0.46**	**0.88±0.38**	**0.87±0.35**	**0.9±0.35**	**0.8±0.29**
Hard manufacturing	8	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Hard manufacturing	9	0.82±0.26	0.9±0.29	0.89±0.41	0.75±0.36	0.85±0.39	0.8±0.33	0.9±0.37	0.93±0.38	0.81±0.3	0.8±0.28	0.84±0.28	0.82±0.28
Hospital	**4**	**8.35±8.62**	**8.63±8.82**	**9.39±13.59**	**8.31±12.96**	**8.76±8.1**	**8.3±8.45**	**7.72±7.16**	**8.55±8.21**	**7.85±7.15**	**7.99±7.87**	**8.43±8.21**	**8.18±8.54**
Industrial chemistry	1	1.06±0.69	0.99±0.66	0.77±1.06	0.87±0.78	1.08±1.02	1.09±0.99	0.59±0.62	0.6±0.64	1.15±0.8	1.56±1.02	1.14±0.8	0.97±0.65
Industrial chemistry	**2**	**1.42±1.24**	**1.29±1.21**	**0.23±1.68**	**0.91±1.33**	**1.19±1.75**	**1.28±1.67**	**0.53±0.97**	**0.51±0.95**	**1.19±1.17**	**1.79±1.53**	**1.57±1.43**	**1.37±1.26**
Industrial chemistry	4	1.06±0.69	0.99±0.66	0.77±1.06	0.87±0.78	1.08±1.02	1.09±0.99	0.59±0.62	0.6±0.64	1.15±0.8	1.56±1.02	1.14±0.8	0.97±0.65
Industrial chemistry	7	1.06±0.69	0.99±0.66	0.77±1.06	0.87±0.78	1.08±1.02	1.09±0.99	0.59±0.62	0.6±0.64	1.15±0.8	1.56±1.02	1.14±0.8	0.97±0.65
Industrial chemistry	9	1.06±0.69	0.99±0.66	0.77±1.06	0.87±0.78	1.08±1.02	1.09±0.99	0.59±0.62	0.6±0.64	1.15±0.8	1.56±1.02	1.14±0.8	0.97±0.65
Irregular school	**7**	**0.24±0.66**	**0.19±0.22**	**0.2±0.27**	**0.13±0.2**	**0.21±0.25**	**0.24±0.28**	**0.18±0.24**	**0.15±0.21**	**0.32±0.6**	**0.35±0.66**	**0.34±0.49**	**0.2±0.24**
Irregular school	9	0.24±0.66	0.19±0.22	0.2±0.27	0.13±0.2	0.21±0.25	0.24±0.28	0.18±0.24	0.15±0.21	0.32±0.6	0.35±0.66	0.34±0.49	0.2±0.24
Lab	6	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Lab	**8**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Laundry	**6**	**9.92±6.71**	**9.33±6.07**	**9.69±8.15**	**7.83±7.77**	**10.84±9.02**	**12.93±10.53**	**11.19±8.32**	**12.15±7.9**	**11.91±7.93**	**11.37±7.37**	**11.51±7.15**	**9.12±5.56**
Livestock	**2**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Livestock	4	1.1±0.17	1.12±0.17	1.26±0.21	1.11±0.19	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Livestock	5	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Military	**2**	**1.1±0.17**	**1.12±0.17**	**1.26±0.21**	**1.11±0.19**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Military	4	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Mineral particulates	1	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Mineral particulates	3	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Mineral particulates	**6**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**

Mineral particulates	7	1.1±0.17	1.12±0.17	1.26±0.21	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Mineral particulates	8	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Mining	**7**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Mining	8	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Office	1	2.68±1.79	2.83±1.9	2.73±2.43	2.41±1.92	3.39±2.64	3.28±2.3	3.3±2.39	2.8±2.13	2.8±1.84	3.05±2.09	3.24±2.01	3.19±2.05
Office	2	1.08±0.15	1.08±0.16	1.2±0.21	1.15±0.21	1.18±0.2	1.15±0.17	1.15±0.17	1.13±0.17	0.99±0.15	0.99±0.15	1.04±0.15	1.01±0.16
Office	3	1.08±0.15	1.08±0.16	1.2±0.21	1.15±0.21	1.18±0.2	1.15±0.17	1.15±0.17	1.13±0.17	0.99±0.15	0.99±0.15	1.04±0.15	1.01±0.16
Office	4	1.08±0.15	1.08±0.16	1.2±0.21	1.15±0.21	1.18±0.2	1.15±0.17	1.15±0.17	1.13±0.17	0.99±0.15	0.99±0.15	1.04±0.15	1.01±0.16
Office	5	1.08±0.15	1.08±0.16	1.2±0.21	1.15±0.21	1.18±0.2	1.15±0.17	1.15±0.17	1.13±0.17	0.99±0.15	0.99±0.15	1.04±0.15	1.01±0.16
Office	6	5.45±6.61	10.83±18.06	2.34±4.96	1.09±2.24	1.26±2.44	1.22±2.28	6.83±8.65	10.48±15.28	10.57±14.78	10.06±14.12	10.06±14.18	9.95±14.38
Office	**7**	**1.11±0.16**	**1.11±0.17**	**1.3±0.24**	**1.21±0.24**	**1.22±0.21**	**1.2±0.18**	**1.21±0.19**	**1.16±0.18**	**1.05±0.16**	**1.05±0.16**	**1.1±0.16**	**1±0.16**
Office	8	3.59±3.91	3.58±3.94	1.53±2.11	1.08±0.83	1.26±0.94	0.92±1.31	1.66±2.17	1.09±1.33	0.94±0.64	1.05±0.72	1.07±0.69	3.79±4.15
Office	9	1.08±0.15	1.08±0.16	1.2±0.21	1.15±0.21	1.18±0.2	1.15±0.17	1.15±0.17	1.13±0.17	0.99±0.15	0.99±0.15	1.04±0.15	1.01±0.16
Residential	**4**	**7.43±3.89**	**8.08±4.11**	**10.44±6.24**	**6.89±4.68**	**7.39±4.4**	**8.53±4.82**	**8.53±4.73**	**8.72±4.68**	**8.04±4.58**	**7.67±4.41**	**8.74±4.78**	**8.28±4.51**
Residential	9	8.14±3.54	8.39±3.62	10.71±5.47	7.54±4.38	7.97±4.04	8.99±4.36	8.95±4.27	9.18±4.28	8.55±4.21	8.17±4.08	9.06±4.35	8.44±4.05
Retail	1	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.19	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Retail	2	0.53±0.13	0.57±0.12	0.55±0.14	0.38±0.1	0.44±0.11	0.48±0.14	0.42±0.1	0.44±0.1	0.43±0.09	0.5±0.13	0.52±0.12	0.5±0.1
Retail	**4**	**0.53±0.13**	**0.57±0.12**	**0.55±0.14**	**0.38±0.1**	**0.44±0.11**	**0.48±0.14**	**0.42±0.1**	**0.44±0.1**	**0.43±0.09**	**0.5±0.13**	**0.52±0.12**	**0.5±0.1**
Retail	7	0.53±0.13	0.57±0.12	0.55±0.14	0.38±0.1	0.44±0.11	0.48±0.14	0.42±0.1	0.44±0.1	0.43±0.09	0.5±0.13	0.52±0.12	0.5±0.1
Retail	8	2.31±1.4	2.69±1.72	3.37±3.09	3.19±4.28	3.49±5.12	3.52±5.7	3.62±6.63	3.35±5.32	3.1±3.65	2.48±2.03	2.52±1.8	1.97±1.39
Retail	9	0.53±0.13	0.57±0.12	0.55±0.14	0.38±0.1	0.44±0.11	0.48±0.14	0.42±0.1	0.44±0.1	0.43±0.09	0.5±0.13	0.52±0.12	0.5±0.1
Salon	**2**	**3.09±0.86**	**3.28±0.96**	**3.95±1.37**	**3.28±1.17**	**3.62±1.31**	**4.12±2.2**	**4.06±1.92**	**3.58±1.12**	**3.46±1.01**	**3.38±0.99**	**3.65±1**	**3.38±1.09**
School	1	3.63±0.98	3.29±0.94	3.81±1.18	3.09±1.11	3.62±1.12	3.44±1	2.82±0.79	1.91±0.72	3.66±1.13	3.21±1.46	3.86±1.56	3.26±1.04
School	**7**	**3.67±0.98**	**3.32±0.94**	**3.81±1.16**	**3.13±1.11**	**3.66±1.12**	**3.48±0.99**	**2.84±0.79**	**1.93±0.73**	**3.7±1.13**	**3.24±1.48**	**3.9±1.58**	**3.3±1.04**

Social	2	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Social	**4**	**1.17±0.31**	**1.29±0.33**	**1.22±0.38**	**1.31±0.38**	**1.5±0.38**	**1.6±0.39**	**1.5±0.38**	**1.44±0.36**	**1.41±0.36**	**1.27±0.35**	**1.21±0.34**	**1.29±0.34**
Social	5	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Social	6	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Social	7	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Social	8	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Social	9	1.17±0.31	1.29±0.33	1.22±0.38	1.31±0.38	1.5±0.38	1.6±0.39	1.5±0.38	1.44±0.36	1.41±0.36	1.27±0.35	1.21±0.34	1.29±0.34
Soft manufacturing	2	0.51±0.44	0.54±0.49	0.63±0.83	0.46±0.44	0.52±0.47	0.5±0.43	0.61±0.7	0.32±0.35	0.43±0.43	0.43±0.44	0.54±0.43	0.45±0.42
Soft manufacturing	4	0.51±0.44	0.54±0.49	0.63±0.83	0.46±0.44	0.52±0.47	0.5±0.43	0.61±0.7	0.32±0.35	0.43±0.43	0.43±0.44	0.54±0.43	0.45±0.42
Soft manufacturing	5	0.51±0.44	0.54±0.49	0.63±0.83	0.46±0.44	0.52±0.47	0.5±0.43	0.61±0.7	0.32±0.35	0.43±0.43	0.43±0.44	0.54±0.43	0.45±0.42
Soft manufacturing	**7**	**0.51±0.44**	**0.54±0.49**	**0.63±0.83**	**0.46±0.44**	**0.52±0.47**	**0.5±0.43**	**0.61±0.7**	**0.32±0.35**	**0.43±0.43**	**0.43±0.44**	**0.54±0.43**	**0.45±0.42**
Soft manufacturing	9	1.1±0.16	1.12±0.17	1.26±0.2	1.11±0.19	1.21±0.2	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.18
Terminal	2	0.66±0.47	0.47±0.39	0.4±0.32	0.4±0.36	0.36±0.28	0.41±0.3	0.46±0.33	0.35±0.27	0.42±0.3	0.43±0.3	0.43±0.3	0.63±0.42
Terminal	**4**	**1.3±1.01**	**0.81±0.75**	**0.7±0.52**	**0.63±0.47**	**0.73±0.52**	**0.85±0.59**	**0.94±0.63**	**0.69±0.48**	**0.84±0.56**	**0.88±0.61**	**0.85±0.59**	**2.06±1.14**
Terminal	8	0.66±0.47	0.47±0.39	0.4±0.32	0.4±0.36	0.36±0.28	0.41±0.3	0.46±0.33	0.35±0.27	0.42±0.3	0.43±0.3	0.43±0.3	0.63±0.42
Textile manufacture	**2**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Tourism	4	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Tourism	6	1.1±0.17	1.12±0.17	1.26±0.2	1.11±0.18	1.21±0.19	1.2±0.19	1.2±0.2	1.16±0.19	1.14±0.18	1.12±0.18	1.12±0.18	1.09±0.17
Waste processing	**7**	**1.1±0.17**	**1.12±0.17**	**1.26±0.2**	**1.11±0.18**	**1.21±0.19**	**1.2±0.19**	**1.2±0.2**	**1.16±0.19**	**1.14±0.18**	**1.12±0.18**	**1.12±0.18**	**1.09±0.17**
Wholesale	**1**	**0.48±0.39**	**0.48±0.38**	**0.44±0.39**	**0.42±0.39**	**0.5±0.39**	**0.48±0.42**	**0.45±0.41**	**0.42±0.37**	**0.4±0.34**	**0.43±0.36**	**0.81±0.72**	**1.1±1.24**
Wholesale	2	0.53±0.34	0.55±0.31	1.24±0.98	0.58±0.71	0.93±0.69	0.95±0.81	0.94±0.67	0.9±0.61	0.93±0.6	0.62±0.4	0.73±0.4	0.56±0.33
Wholesale	3	0.47±0.11	0.49±0.11	0.48±0.15	0.31±0.11	0.48±0.12	0.47±0.12	0.49±0.13	0.43±0.14	0.39±0.09	0.38±0.09	0.45±0.1	0.45±0.1
Wholesale	4	0.45±0.24	0.46±0.24	0.63±0.45	0.62±0.5	0.55±0.38	0.44±0.29	0.42±0.29	0.4±0.27	0.39±0.24	0.37±0.25	0.41±0.23	0.46±0.24
Wholesale	6	0.47±0.11	0.49±0.11	0.48±0.15	0.31±0.11	0.48±0.12	0.47±0.12	0.49±0.13	0.43±0.14	0.39±0.09	0.38±0.09	0.45±0.1	0.45±0.1

Wholesale	**7**	**0.46±0.13**	**0.49±0.14**	**0.39±0.17**	**0.2±0.09**	**0.44±0.13**	**0.45±0.15**	**0.49±0.15**	**0.41±0.12**	**0.36±0.1**	**0.35±0.1**	**0.39±0.1**	**0.37±0.11**
Wholesale	8	0.47±0.11	0.49±0.11	0.48±0.15	0.31±0.11	0.48±0.12	0.47±0.12	0.49±0.13	0.43±0.14	0.39±0.09	0.38±0.09	0.45±0.1	0.45±0.1
Workshop	2	0.45±0.13	0.47±0.12	0.39±0.13	0.32±0.11	0.46±0.13	0.44±0.21	0.38±0.14	0.41±0.13	0.39±0.12	0.43±0.12	0.5±0.13	0.33±0.1
Workshop	6	0.45±0.13	0.47±0.12	0.39±0.13	0.32±0.11	0.46±0.13	0.44±0.21	0.38±0.14	0.41±0.13	0.39±0.12	0.43±0.12	0.5±0.13	0.33±0.1
Workshop	**7**	**0.43±0.14**	**0.44±0.12**	**0.37±0.13**	**0.33±0.12**	**0.45±0.14**	**0.42±0.23**	**0.36±0.15**	**0.39±0.14**	**0.37±0.13**	**0.42±0.13**	**0.48±0.13**	**0.32±0.11**
Workshop	8	0.45±0.13	0.47±0.12	0.39±0.13	0.32±0.11	0.46±0.13	0.44±0.21	0.38±0.14	0.41±0.13	0.39±0.12	0.43±0.12	0.5±0.13	0.33±0.1
Workshop	9	0.73±0.37	0.78±0.41	1.75±1.32	0.23±0.36	0.72±0.39	0.73±0.42	0.73±0.43	0.66±0.4	0.67±0.39	0.62±0.34	0.72±0.34	0.58±0.3