Methods for Estimating Seasonal Populations with Water and Energy Data

Prepared by:
Jon Martindill
UC Davis Center for Water-Energy Efficiency
215 Sage Street Suite 200, Davis, CA 95616

Prepared for:
California Department of Water Resources
Report Number: WUES-DWR-2021-08.T1

Submitted on:
June 22, 2022

EXECUTIVE SUMMARY 1
1 INTRODUCTION 3
2 BACKGROUND 4
2.1 Seasonally Occupied Homes 4
2.2 Literature Review 5
3 METHODS 6
3.1 Data 7
3.1.1 Water Consumption Data 8
3.1.2 Energy Consumption Data 8
3.1.3 Additional Data 8
3.2 Detect Daily Occupancy Step 9
3.2.1 Detecting Daily Occupancy with Hourly Water Data 9
3.2.2 Detecting Daily Occupancy with Hourly Electricity Data 10
3.3 Classify Homes Step 12
3.3.1 Classifying homes with daily occupancy patterns 12
3.3.2 Classifying homes with monthly water data 14
3.4 Calculate Efficient Water Use Step 18
3.4.1 Calculate efficient water use with daily occupancy data 19
3.4.2 Calculate efficient water use with monthly water data 19
3.4.3 Calculate efficient water use with census data only 20
3.4.4 Adjust efficient water use calculation with vacation rental data 21
4 RESULTS & DISCUSSION 22
4.1 Occupancy Detection Results 22
4.1.1 Detecting occupancy with hourly water data 22
4.1.2 Detecting occupancy with hourly energy data 24
4.2 Home Classification Results 26
4.3 Seasonal Population Efficient Water Use Volume Calculation Results 29
5 CONCLUSION 31
REFERENCES 32
APPENDIX A – Detecting Seasonal Occupancy with Hourly Water Data 33
Removing Irrigation Water Use 33
Removing Leaks 34
APPENDIX B – Detecting Seasonal Occupancy with Hourly Electricity Data 37

EXECUTIVE SUMMARY

The Indoor residential water use efficiency standard in California sets a budget of the aggregated indoor residential water use for urban retail water suppliers based on the census population of their service areas. This water budget does not account for water use that occurs in residential homes by non-permanent residents. Homes that are primarily occupied by non-permanent residents, also known as seasonally occupied homes, effectively have an indoor water use budget of zero under the current indoor residential water use efficiency standard. Urban retail water suppliers with large numbers of seasonally occupied homes are therefore faced with inappropriate assessments on their water use efficiency. The purpose of this study, which was funded by the California Department of Water Resources, is for the Water-Energy Efficiency (CWEE) of the University of California, Davis to develop methods for detecting seasonally occupied homes and calculating adequate efficient indoor water use for those homes.

Three separate methods were developed by this report to calculate efficient indoor water use by seasonal population. Each method relies on different types of data and data processing tools, representing the range of what different urban retail water suppliers are capable of. The first method uses either hourly water consumption data or hourly electricity data to detect daily occupancy patterns for all homes, then classifies homes as 'seasonally occupied' or 'permanently occupied' based on those occupancy patterns and other property and account characteristics. The second method uses only monthly data to classify homes as seasonally occupied, and then uses those classifications to calculate the efficient indoor water use of the classified homes. The third method uses census data only to estimate the efficient water use of seasonally occupied homes.

The result from this study suggests that identifying seasonally occupied homes with hourly water data is the most accurate approach. However, this approach requires hourly water data, which is not available to many urban retail water suppliers in the state. Performing the same calculation with hourly electricity data shows similar results, but slightly less accurate; however, urban retail water suppliers are generally unable to obtain such data. Identifying seasonally occupied homes with monthly water data is more accessible for most water suppliers but has significantly lower accuracy than the methods using hourly water or electricity data.

Once seasonally occupied homes are identified, the next step is to calculate the volume of water that indicates the efficient indoor water use by seasonal populations. This calculation requires estimating the number of people-days that seasonal populations spend in residential homes in the urban retail water supplier's service area, and multiplying that value with the indoor efficient water use standard. Performing this calculation with seasonal occupancy estimates using hourly water and electricity data tends to result in the lowest number of estimated people-days, and therefore results in the smallest volume calculation of all methods. Estimating seasonal populations with monthly water data can effectively identify which homes are seasonally occupied, but cannot measure the number of days those home are occupied, which results in a less precise estimate of people-days. The approach outlined in this report tends to overestimate people-days of seasonal populations when using monthly water data alone, leading to a larger calculated volume of indoor efficient water use. Finally, the calculation using census data yields similar results to the method that uses monthly water data but is less prone to overestimating people-days in the sample of water suppliers tested in this study. These results suggest that, if accuracy is the highest priority, the method that uses hourly water or electricity data is the best. However, since that method relies on data unavailable to most water suppliers as well as complex methodologies, the methods using monthly water data and census data alone are more realistic for most urban retail water suppliers.

1 INTRODUCTION

In 2018, Assembly Bill 1668 and Senate Bill 606 (2018 Legislation) laid out a new framework for water conservation and drought planning in California. The California Department of Water Resources (DWR) and the State Water Resources Control Board have been collaborating with stakeholders across a range of use cases in the residential, commercial, industrial, and institutional sectors to establish new efficient water use standards, variances, and performance measures within this framework. These standards will be used to calculate annual urban water use objectives (UWUOs) for urban retail water suppliers. In addition to standards, DWR will recommend appropriate variances for certain unique uses that may materially impact a supplier's ability to meet its UWUO. One such unique use is from a significant presence of seasonally occupied homes in a supplier's service area. Seasonally occupied homes are residential units that have occupants who do not identify as permanent occupants in the urban retail water supplier's service area, and therefore do not count toward the population under the U.S. Census. Subject to DWR's recommendations, urban retail water suppliers with high numbers of seasonal populations could receive a variance to their efficient indoor residential indoor water use standard (IRWUS).

The IRWUS is defined on per-capita water use and was set to 55 gallons per capita per day (gpcd) in 2020, stepping down to 52 gpcd in 2025, and 50 gpcd in 2030, in the 2018 Legislation. DWR may recommend different standards for the Legislative's consideration. The amount of water use is to be estimated using the population as reported by the U.S. Census Bureau or California Department of Finance. This method presents an issue for many urban retail water suppliers, as the census data only reports "usual residents" in total population numbers and defines "usual residents" as people who live and sleep at a given home most of the time (U.S. Census Bureau, 2021). Therefore, water use in homes classified as residential by non-usual residents is captured in the measurement of residential water use, but the associated water users are not. The appropriate variance for remedying this discrepancy will estimate the number of people-days in those seasonally occupied homes such that the IRWUS can be applied to that seasonal population.

The goal of this study, which was funded by DWR, is for the Water-Energy Efficiency (CWEE) of the University of California, Davis to develop methods for identifying seasonally occupied homes, to estimate the water usage of seasonally occupied homes, and to inform the development of an appropriate variance for significant water use of seasonal populations under a variety of data-availability scenarios. This report explains the developed methodologies in detail, demonstrates their use and efficacy using examples of three urban retail water suppliers, and illustrates the corresponding potential aggregated water use volume that could be claimed by these urban retail water suppliers under a variance.

2 BACKGROUND

This section provides background on seasonally occupied homes, their prevalence in California, and existing literature on identifying occupancy patterns.

2.1 Seasonally Occupied Homes

The census data categorizes homes that are occupied for part of the year as “seasonal, recreational, or occasional housing” (U.S. Census Bureau, 2021). This category includes second homes, vacation homes, and vacation rentals, provided that the home is still categorized as a residence. Seasonally occupied homes do not need to have any particular seasonal pattern of occupancy – only that they are not the residences for any usual residents. For the purposes of this report, all residential homes with seasonal, recreational, or occasional occupants will count as seasonally occupied.

According to census data, most urban retail water suppliers in California have less than 20% of homes defined as seasonally occupied (Figure 1). This is especially true for the most densely populated areas of the state. However, in some vacation communities the majority of homes are seasonally occupied. In such cases, the allowable water use volume based on the census population data and IRWUS could only represents a fraction of residential water users and thus, is not appropriate for determining indoor residential water use efficiency of the associated urban retail water supplier.

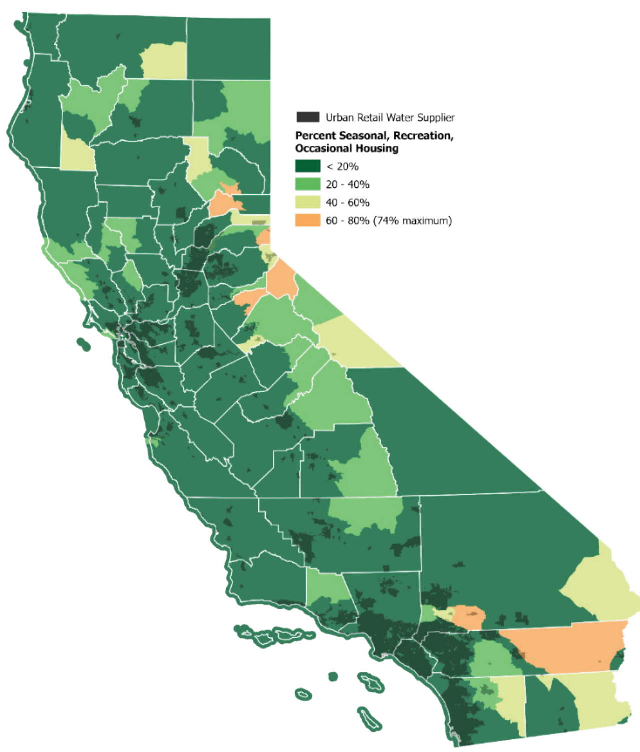

Figure 1: Percentage of homes seasonally occupied in each county in California (DWR 2021)

Seasonally occupied homes vary widely in their occupancy characteristics. The two main categories of seasonal homes that we consider in this report are second homes and vacation rentals. We define second homes as residences that are occasionally occupied by a person or group who have a usual residence elsewhere. Second homes may be owned or rented but are only ever occupied by one particular person or group. If we know the occupancy pattern of a home, second homes are easier to identify through their characteristics of water and energy uses because, by definition, they can only be occupied up to 50% of the time before they become a usual residence. The other type of home is a vacation rental. Vacation rentals are homes that can be rented to any group for short periods of time and may be listed through services such as Airbnb and VRBO. Since they are occupied by different groups at different times, the overall occupancy rate of these homes can far exceed 50%. According to Airdna.co (AirDNA 2021), vacation rentals listed on these websites are often occupied 70% of the time or more, depending on location. Such homes are more challenging to identify as seasonally occupied, as the water- and energy-use characteristics of these homes may be very similar to those of permanently occupied homes.

Seasonally occupied homes also vary by occupancy pattern, often related to the specific climate conditions of a community. Communities that offer popular summer recreation, such as boating and swimming, may see higher rates of seasonal occupancy in the summer. Skiing destinations, on the other hand, may experience high seasonal occupancy in winter months. Patterns of seasonal occupancy have a large influence on the number of days that a second home or vacation rental is occupied, and will play a role in calculating an appropriate variance.

2.2 Literature Review

Through a thorough literature review, no previous research was found to identify seasonally occupied homes using water or electricity data; however, there is an existing body of research on non-intrusive methods for detecting household occupancy. Occupancy detection is one of the primary goals of this report, but it must be done with data that could reasonably be available to urban retail water suppliers or to an entity that water suppliers could hire to perform the calculation. The most applicable prior research would have relied on monthly or hourly water data, or electricity data as the basis for occupancy detection. However, very few of these papers use water data to detect occupancy, and those that do use it along with other household sensors (e.g., Wang *et al.* 2021). There are many papers that detect household occupancy status using electricity data, but also rely on information or methods that are generally not applicable to this current research effort. Such papers typically use energy data at a higher resolution than that available to CWEE and employ machine learning techniques that rely on training datasets (Causone *et al.* 2019; Huchuk, Sanner, and O'Brien 2019, Kleminger and Staake 2013, Razavi *et al.*2019). The methods presented in these papers rely on data sources and computer models that are unavailable to urban retail water suppliers. Therefore, these approaches are not applicable in the context of this research because the intent is to develop methods that urban retail water suppliers can undertake on their own.

3 METHODS

The purpose of this report is to develop methods to calculate the volume of efficient indoor water use for seasonal populations. The IRWUS is based on water per person per day, so the methods created by this report must generate an estimate for the number of people-days of seasonal occupants within the service territory of an urban retail water supplier. Therefore, the method to calculate the volume of efficient water use by seasonal population should provide as accurate estimates as possible of the following metrics:

1. Number of seasonally occupied homes
2. Days per year each seasonal home is occupied
3. Number of people in each seasonal home when occupied

After calculating the number of people-days in seasonally occupied homes, it is reasonable to assume that the IRWUS can be applied to the identified seasonal population to calculate the efficient water use volume. The methods outlined in this report generally strive to find average metrics for the entire service area of an urban retail water supplier. That is – the number of seasonally occupied homes, the average number of days each seasonally occupied home is occupied over the course of a year, and the average number of people who use a seasonally occupied home when it is occupied.

The first metric is the number of seasonally occupied homes in a urban retail water supplier's service area. These are homes that are classified as residential accounts for the urban retail water supplier, but as seasonally occupied residences according to the U.S. Census. The U.S. Census already publishes this information, however; it is not necessary for any method in this report to accurately calculate the number of seasonally occupied homes. Instead, this report proposes two methods to identify which homes are seasonally occupied – an information is useful for estimating the occupancy patterns of seasonally occupied homes.

The second metric, the number of days that seasonally occupied homes are occupied, can only be accurately estimated with a method that detects daily occupancy patterns for each home. Section 4.2 proposes two approaches for calculating an estimate of occupancy patterns for each home with good accuracy. Both approaches are methodologically complex and require data that some urban retail water suppliers are be unable to access. Therefore, alternative estimates for the number of days that seasonally occupied homes are occupied are necessary.

Estimating the number of people inside seasonally occupied homes is the third necessary metric for calculating an appropriate seasonal population efficient water use. Unfortunately, there is no reliable and non-intrusive way to measure this value when calculating the efficient water use of seasonal populations. This report proposes several approaches that can be used to estimate this number.

CWEE developed three distinct methods to estimate efficient water use of seasonally occupied homes, which are determined by data availability. The first method – "Daily" – relies on hourly water or energy data to detect daily occupancy for all homes (see Section 4.2), then use those daily occupancy estimates to classify homes (see Section 4.3.1) and calculate efficient water use for seasonally occupied homes (see Section 4.4.1). The second method – “Monthly” – does not attempt to detect occupancy patterns of homes, but instead simply classifies homes as seasonally occupied or permanent based on monthly water data (see Section 4.3.2). Then, it calculates efficient water use based on those classifications (see Section 4.4.2). The third method – “Census” – uses only publicly available data from the U.S. Census to calculate the efficient water use of seasonally occupied homes (see Section 4.4.3). Figure 2 shows how each of these methods breaks down into steps and shows where they are detailed in this report.

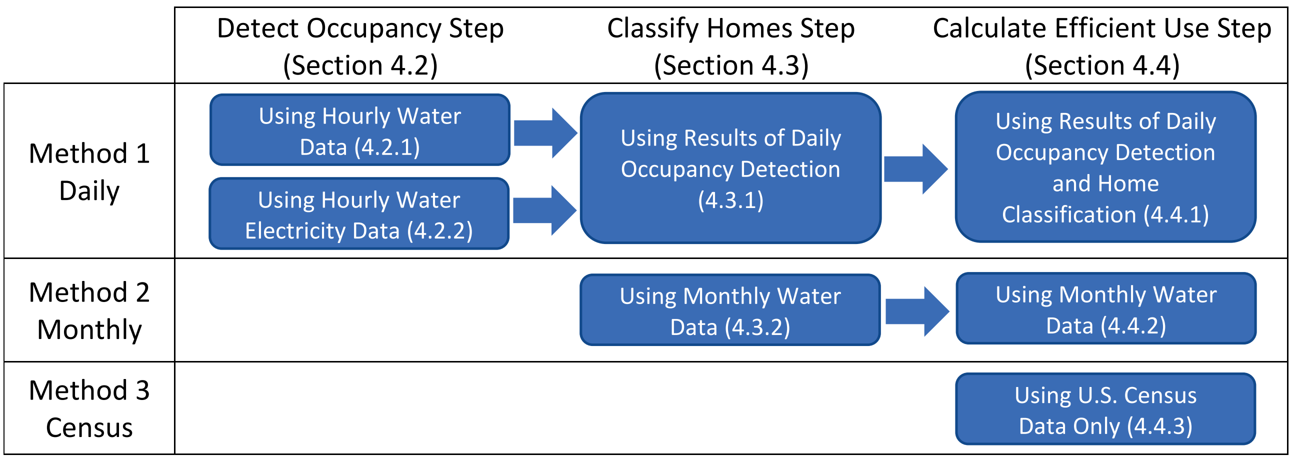

Figure 2: Relationship between methods in this report, calculation steps, and the section numbers in which each method is described

This section describes the datasets used to perform each method, broadly presents the techniques for detecting occupancy and classifying homes, and finally details the approaches that can be used for calculating the seasonal population efficient water use, and which are most appropriate given the input data. For the technical details of these methods, see the Appendices to this report.

3.1 Data

This study aims to develop methods that use data that is available to urban retail water suppliers. To do so, CWEE collaborated with three different urban retail water suppliers with high numbers of seasonal populations as well as the electricity suppliers whose service areas overlap with those of the three urban retail water suppliers. This report used private data and discussed water use patterns of the participating water suppliers. Therefore, all participating urban retail water suppliers will remain anonymous, and we will discuss them using generic names as Communities A, B, and C. Community A, served by Water Supplier A, offers both winter and summer activities, and 50% of all homes are seasonally occupied. Water Supplier B serves Community B, which is about 70% seasonally occupied and primarily offers summer activities. Community C, served by Water Supplier C, is about 25%

seasonally occupied and is primarily a spring and fall destination due to extreme heat in the summer months and a lack of winter activities. To further protect anonymity, some figures in this report are reported in approximate numbers.

Each community has access to different types of data. CWEE has identified three scenarios of data availability:

1. Hourly water consumption data is available (only available to some water suppliers with Advanced Metering Infrastructure (AMI))
2. Hourly electricity data is available (not available to most water suppliers)
3. Monthly water consumption data is available (available to majority of water suppliers)

The three participating water suppliers all have different data availabilities and provide an ideal case study for each of the above scenarios. In addition to water and energy data, some supplementary publicly available datasets (described in 4.2.3) are necessary and used for some elements of the analysis.

3.1.1 Water Consumption Data

The three participating urban retail water suppliers provided account-level residential water consumption data to CWEE for use in this study. Water Supplier A provided hourly data from 2018 to 2020, and Water Supplier B provided hourly data from 2013 to 2020. Water Supplier C did not have access to hourly water consumption data but provided monthly water consumption data from 2015 to 2020. Water Supplier A provided over 11,000 accounts, Water Supplier B provided about 8,000, and Water Supplier C provided over 100,000 accounts. All water consumption datasets include both the service address and the billing address for each account. The datasets went through pre-processing before use, including geocoding and removing extreme data points that are at or beyond the 99th percentile of the data range.

3.1.2 Energy Consumption Data

Electricity utilities provided account-level, hourly energy consumption data from 01/2017 to 09/2020 for all single-family residences in regions intersecting Communities B and C. The electricity utility supplying energy in the service area of Water Supplier A was not willing to share data for the purposes of this study. The available energy use datasets went through pre-processing before use, including filling missing record and eliminating extreme data points. Billing information, including billing address, were not provided for electricity accounts.

3.1.3 Additional Data

CWEE acquired data from the U.S. Census to calibrate seasonal occupancy identification and calculate the water use of seasonally occupied homes. All data used are from the American Community Survey 5-year estimates (U.S. Census Bureau, 2019).

Hourly weather data was used to detect occupancy with electricity data. CWEE acquired this data from California Irrigation Management Information System (CIMIS) weather stations closest to the addresses in each water supplier's service area. These data included dry-bulb air temperature, precipitation, and relative humidity. Missing data were imputed by fitting a LOWESS (Locally Weighted Scatterplot Smoothing) smoothing function of time between the previous and following non-missing values at the same location.

City-specific data on vacation rentals were collected from the AIRDNA Marketminder tool (AirDNA 2021). This tool lists the number of active rentals on Airbnb and VRBO and breaks down average occupancy characteristics of these listing. These data are publicly available without any paid account.

3.2 Detect Daily Occupancy Step

CWEE developed two methods to detect daily occupancy in residences. The purpose of these two methods is to be considered in developing a variance for seasonally occupied homes to calculate the allowable volume under the variance.

The two occupancy detection methods align with two of the data availability scenarios – hourly water data and hourly electricity data. These methods have been optimized for practicality and simplicity, where possible, so that they may be replicated by urban retail water suppliers hoping to classify occupancy in their own service territories.

3.2.1 Detecting Daily Occupancy with Hourly Water Data

This method uses hourly account-level water data to determine the daily occupancy for each home. This section briefly describes this method with multiple steps, which are explained in more detail in Appendix A.

Water use in any given hour may be an indicator of occupancy, but it is first necessary to rule out water use that is not associated with occupancy. Two common water uses that are not associated with occupancy are water use for irrigation, and water use associated with leaks. Scheduled irrigation controlled by an automatic irrigation system is not influenced by household occupancy, but household water meters record this usage unless it is metered separately. Slow leaks can cause meters to register continuous or periodic usage, even when no person is using water at a residence. Therefore, to accurately detect daily occupancy with hourly water data, it is critical to identify and remove water use associated with automated irrigation and leaks.

To detect water use associated with scheduled irrigation, CWEE leverages the fact that irrigation events typically have higher water use intensity than most normal household uses. Also, scheduled irrigation will generally occur at the same time of day and same day of week. For this analysis, all water use during a given hour is flagged as "irrigation" if the total usage of that hour exceeds 6 cubic feet and the usage is no more than 10% different from the usage during the same hour of the day one week prior or one week following. Hours classified as irrigation are removed from the subsequent analysis for detecting daily occupancy. This method is unable to detect automated irrigation that uses less than 6 cubic feet per hour. Lowering the threshold of detection below 6 cubic feet could result in over-identification of irrigation events, since some households could routinely use four to six cubic feet in an hour for non-irrigation purposes. In locations where low-water-use irrigation is prevalent, further refinements of methods to detect such irrigation may be necessary.

Leaks may lead to a misidentification of occupancy if the only water use at a home during a day is due to leaks. There are two possible procedures for removing leaks from hourly water data, depending on the type and precision of hourly water reading. One of the participating water suppliers provided readings from positive displacement water meters that report at a precision of one cubic foot. Positive displacement meters emit a signal when one cubic foot of water has been released since the previous signal. Removing leaks from positive displacement meter readings must account for the fact that low, steady usage less than one cubic foot per hour will appear as alternating ones and zeros. To identify a day as "leak-only", it must be shown that all readings are either 0 or 1 (or 1 and 2 to allow for leaks >1 cubic feet per hour) and that the readings occur at regular intervals throughout the day through metrics defined as "evenness" and "consistency". Any day that meets those criteria is flagged as "leak-only" and is counted as unoccupied.

One of the water suppliers provided readings from a velocity flow meter, which provides precise volumetric readings for each hour to the 0.01 cubic-foot precision. Removing leaks from velocity flow meters is much simpler because each hour has a precise recording of water consumed. If every hour in a given day has usage within a range of 1 cubic foot, and there are no hours with usage below 0.03 cubic feet, all of the usage in that day is considered "leak-only". If all hourly usage is within one cubic foot, it implies that either there is a consistent leak or that a person is in the home but is only using very small amounts of water. Using a toilet or sink only can easily lead to no hours with usage greater than one cubic foot. Therefore, to identify occupancy on low-usage days, the crucial identifying feature is that many hours will see zero water usage. If a home has some hours with greater than 0.1 cubic feet of water usage, but others with zero, it is likely that it is occupied. However, many houses rarely meet the threshold of zero water use, likely due to very slow leaks. In testing, approximately 5% of homes had more hours with 0.01 cubic feet of water use than hours with zero, and 4% had more hours of 0.02 cubic feet than zero. Only 1% of homes had more hours of usage at 0.03 cubic feet, however. Therefore, a threshold of 0.03 cubic feet per hour was selected to be the practical measurement of zero water usage. In the context of identifying leaks, hours with usage below 0.03 indicate that usage greater than 0.1 in other hours is not likely to be caused entirely by leaks.

Once hours flagged as irrigation have been removed from analysis, and days flagged as "leak-only" have been classified as unoccupied, the final designation of occupancy is simple – add up all water use for each day, and if there is at least one cubic foot of water use during the day, it is flagged as occupied. The resulting daily occupancy patterns detected for each home are then used to classify homes in Section 4.3.1.

3.2.2 Detecting Daily Occupancy with Hourly Electricity Data

This method uses account-level hourly electricity data to determine daily occupancy for each home. It is explained briefly in this section and in more detail in Appendix B.

While water consumption is usually limited to discrete uses around the home, modern homes have a wide variety of energy draws, often at all times of the day, and there are no clear signals of electricity use associated with occupancy. Refrigerators, for example, can use significant amounts of energy at irregular intervals, regardless of occupancy. Many appliances use small amounts of energy at all times. Some homes have lights that turn on automatically at certain times of the day. Finally, homeowners will often leave HVAC systems operational, even when away for extended periods of time, although they may adjust settings to minimize heating and cooling loads. While some studies have attempted to address these challenges by disaggregating individual uses with high-resolution energy consumption data (such as Klemionger and Staake 2013), such an approach is not possible with hourly energy data. Instead, the approach we propose relies on the relationship between outdoor temperature and daily energy use and the fact that this relationship often changes depending on whether homes are occupied. Due to the complexity of this analysis, some advanced statistical modeling is required.

CWEE developed a method of detecting daily occupancy by combining daily total electricity consumption for each home with daily average temperature data measured at the nearest weather station. In particular, this study used heating degree-days (HDD) and cooling degree-days (CDD) to capture differing responses to temperature depending on whether the daily average temperature is above or below 65 degrees Fahrenheit. The threshold of 65 degrees Fahrenheit is used by the National Weather Service (NWS CPC 2022) to indicate the cutoff at which daily temperatures require heating or cooling loads in buildings. Then, CWEE solved a latent class mixed model (LCLMM) for each home. The model solves two linear regression equations simultaneously for each home's daily energy use, where each equation represents the home being in one of two states – occupied or unoccupied.

The first equation estimates α, the energy use when the daily average temperature is at 65 degrees Fahrenheit and the home is occupied, β_1, the change in energy use for each CDD when the home is occupied, and β_2, the change in energy use for each HDD when the home is occupied. The second equation estimates $\tilde{\alpha}$, the energy use when the daily average temperature is at 65 degrees Fahrenheit and the home is unoccupied, β_1, the change in energy use for each CDD when the home is unoccupied, and β_2, the change in energy use for each HDD when the home is unoccupied. The model then estimates two different models in the way that best explains the data, and assigns each day a probability that it belongs to either the first or second class – and is therefore either occupied or unoccupied. Equation 1 below shows the equation represented symbolically.

$$E_{i,d} = \begin{cases} a + \beta_1 CDD_{i,d} + \beta_2 HDD_{i,d} \\ \tilde{a} + \widetilde{\beta_1} CDD_{i,d} + \widetilde{\beta_2} HDD_{i,d} \end{cases}$$

Equation 1

The LCLMM analysis was performed in R 4.0.3 (R Core Team, 2020) with the function *hlme* in the package *lcmm* (v1.9.2; Proust-Lima, Philipps, and Liquet, 2017). To run the analysis, specify the basic form of the model (i.e., $E_{i,d} = CDD_{i,d} + HDD_{i,d}$) and specify that each day belongs to one of two classes. The function then runs the model and generates two models – one for each of the two classes. It simultaneously calculates the probability that each day belongs to the two classes. If there is at least a 5% probability that a day belongs to the class with higher energy use, we identify that day as occupied.

3.3 Classify Homes Step

CWEE developed two methods to classify homes as seasonally or permanently occupied. The first method relies on results from the daily occupancy detection method as described in the previous section, while the second method can be performed with monthly water data only. In other words, the second method can be used by urban retail water suppliers with no access of hourly water and energy use data.

3.3.1 Classifying homes with daily occupancy patterns

This method classifies homes as permanently or seasonally occupied using only the occupancy detection results (see Section 4.2) and other information available through an urban retail water supplier's billing system. By itself, daily occupancy patterns discussed above are not sufficient for determining permanent occupancy. A second home can only be occupied by its owner less than 50% of the time, and a permanent residence must, by definition, be occupied more than 50% of the time. However, shared homes and vacation rentals may be occupied more than 50% of the time and still fall under the definition of a seasonal occupied home. Therefore, CWEE developed a method to classify seasonally occupied homes using information that can be extracted from the occupancy data and the account's billing information.

The first step to classifying homes is to generate variables associated with permanent or seasonal occupancy. Variables are measurable and quantifiable metrics that are expected to have different values depending on whether a home is permanently or seasonally occupied. Table 1 summarizes five variables that have been generated as part of this analysis, and the corresponding justification of their usefulness as well as their limitations.

Table 1: Justification and limitations of variables used to classify homes as seasonal

Variable	Justification	Limitations
Overall occupancy rate	Seasonally occupied homes and vacation rentals are more likely to have lower occupancy rates.	Many vacation rentals have occupancy rates well over 50%.
Weekend/Holiday Occupancy	A vacation home is more likely to be occupied on days where most people do not have to work.	Permanent residents with non-traditional work schedules may also have higher occupancy on weekends.
Matching Billing Address	Second homes and vacation rentals are likely to have a different billing address from their service address.	Rental properties with permanent residents may have the same structure. Post Office (PO) boxes complicate matching.
Duplicated Billing Address	If one person or entity owns multiple residential properties, it is more likely to be a vacation rental.	A single owner or bill-payer for multiple properties could also indicate permanent rental use.

None of these variables shown in Table 1, alone, could identify seasonally occupied homes with much certainty, but they are effective when used together. Whether or not a property has a local billing address gives a strong distinction between full-time permanent residences and everything else, but it does nothing to identify vacation rentals and other atypical permanent residences from permanently occupied rentals. A low occupancy percentage, as determined from the analysis in 4.2.1 or 4.2.2, indicates a home that is not continuously occupied, but such a home could easily be the permanent residence for somebody who travels frequently. A high occupancy home is likely permanently occupied, but some vacation rentals have high occupancy too.

Using these variables, CWEE developed a set of rules to classify homes into three categories: Permanent Homeowners, Permanent Renters, and Seasonally Occupied. The criteria for these three categories are summarized in Table 2.

- A house meets the Permanent Homeowner criteria if its billing address matches the service address, and it is occupied more than 50% of the time with workday occupancy at least as high as weekend occupancy.
- A house meets the Permanent Renter criteria if its billing address does not match the service address and it is occupied more than 75% of the time. The occupancy criterion increases for this classification because a mismatch between the billing and service address also increases the likelihood that a home is seasonally occupied.
- Seasonally Occupied homes do not have specific classification criteria. Instead, homes that do not meet either of the criteria for permanently occupied homes by either homeowners or renters are classified as Seasonally Occupied if they are occupied at least 1% of the time.
- A home is classified as Vacant if it is occupied less than 1% of the time.

Table 2: Proposed classification criteria of permanent homes with daily occupancy data

Variable	Permanent Homeowner	Permanent Renter	Seasonally Occupied
Billing Address	Reasonable[1] and non-duplicated	Not reasonable[1] or duplicated	Not Applicable – Any home not classified as "Permanent" is either Seasonal or Vacant
Overall Occupancy	>50%*	>75%*	>=1%*
Workday Occupancy	Weekend Occupancy <= Workday Occupancy *	Weekend Occupancy <= Workday Occupancy *	Not Applicable (see above)

[1] Customers that use PO boxes as the billing address should be considered "reasonable" if the PO box is close to or in the same zip code as the service address.
** Proposed starting point to be adjusted to calibrate classification results to match census records*

The values proposed in Table 2 are used as starting points for calibration. Once all homes are classified, CWEE compares the total number of homes in each category to the values reported by the U.S. Census. Table B25003 from the American Community Survey summarizes the total number of owner-occupied and renter-occupied homes, and Table B25004 provides the number of homes for seasonal, recreational, or occasional uses. The criteria for overall occupancy and workday occupancy can be strengthened (by adjusting the threshold values up) or weakened (by adjusting the threshold values down) until the classifications match the most recent census data.

3.3.2 Classifying homes with monthly water data

Monthly water data is available for most urban retail water suppliers, if not all. Although it cannot be used to accurately detect occupancy patterns in homes as described in Section 4.2, it can still be used to classify homes into the three categories of Permanent Homeowner, Permanent Renter, and Seasonally Occupied.

To perform this calculation, use three to five years of monthly water use data for a group of homes. This group of homes may be all homes within the service area of an urban retail water supplier or a subset of homes with similar characteristics (as described later). Calculate the median monthly water use for all months and all homes in that sample. Then, for each home, calculate the number of months with usage below the median. For example, if the median monthly water use in a given year is 150 cubic feet, and a home has only three months with usage below 150 cubic feet, this home is calculated as having three low-water-use months. Finally, for all homes, calculate the percentile rank of each home's low water use month. The percentile rank converts the number of months with below-average water use to a number that represents the percentage of all homes in the analysis that have the same number or fewer of such months. Therefore, if a home has more low-water-use months than any other home it will score 100% - 100% of homes in the sample have below-average water use equally or less frequently than this example home. If a home has the lowest number of low-water-use months of all homes, it will score close to 0%. This value is the low-water-use percentile, and it is the primary variable we use to classify homes with monthly water data.

If most homes in a community have similar characteristics, the calculation of the low-water-use percentile can be performed for all homes together. However, this method can be improved by grouping together homes based on home size and/or landscape size. Home size is generally accessible through accessor data, and when matched to water accounts, provides a reasonable proxy for number of inhabitants. Some urban retail water suppliers maintain information on landscape size for each account, which would allow homes with small landscapes to be grouped together so that they are not classified along with homes with large, irrigated landscapes. Additional variables may be considered as well, such as age of home, number of bedrooms, neighborhood, or number of inhabitants. The intention is to create groupings of homes that would be assumed to have similar water use aside from occupancy patterns. The objective should be to create as uniform of groupings as possible while still having enough homes in each group to ensure robust analysis. This study used a minimum group size of 500 homes. For example, if an urban retail water supplier has data on the landscape size of each home, it could split homes into groups for <1,000 square feet, 1,000-5,000 square feet, and 5,000+ square feet, so long as each group has at least 500 homes. Then, within each group, the urban retail water supplier could perform the calculation as described in the previous paragraph. Larger urban retail water suppliers with diverse home types and more detailed data of home characteristics could split homes into groups based on multiple variables at a time, such as landscape size, home size, and home age.

The low-water-use-percentile is a reasonable but imperfect indicator that a home is permanent or seasonally occupied. Figure 3 shows the distribution of low-water-use percentile in homes in the Communities A and B, with home classification as calculated by the hourly water data method described in Section 4.2.1. The intersection point of the two distributions represents the threshold at which the low-water-use percentile should be set to classify homes as accurately as possible.

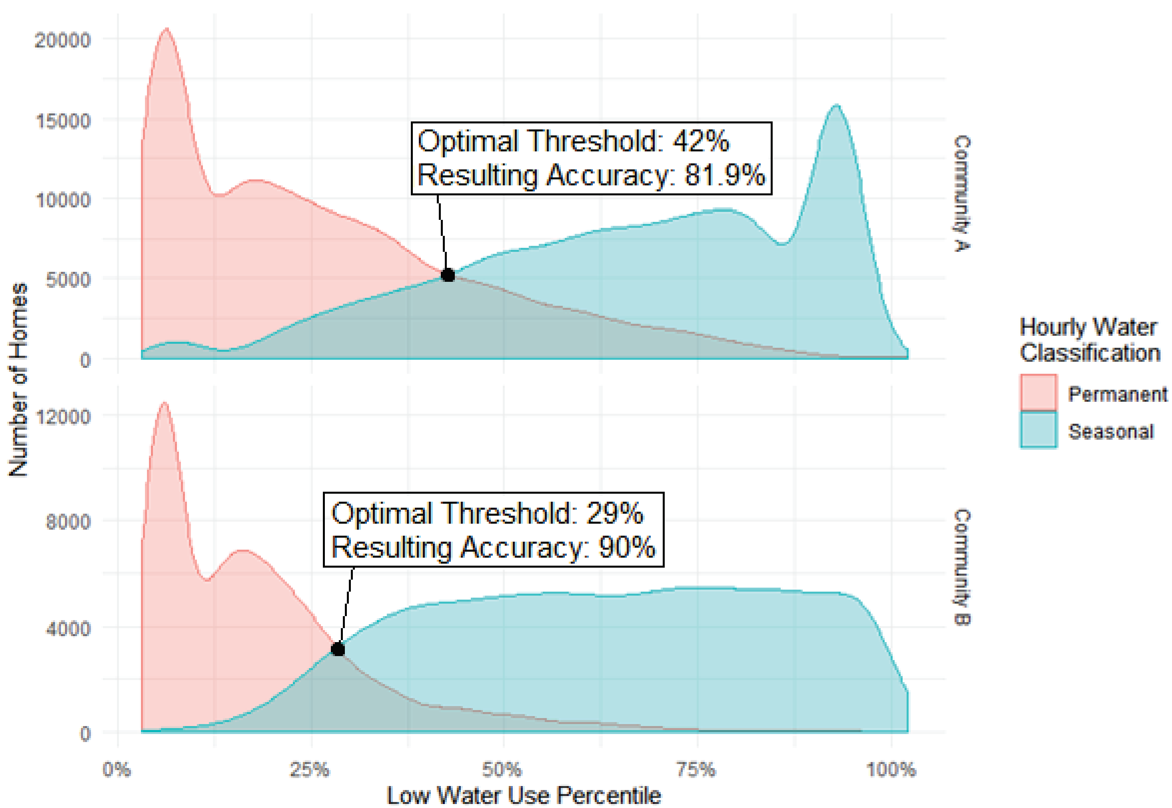

Figure 3: Distribution of low-water-use percentile in homes classified with the hourly water data method in the Community A (top) and Community B (bottom). The area under the red (blue) lines indicates the number of permanent (seasonal) homes of a given low-water-use percentile.

Community B has more pronounced patterns of seasonal occupancy than Community A. As a result, seasonally occupied homes in Community B are more likely to have very low water use rates during a large portion of the year. The error-minimizing classification threshold for Community A is 42%, as 80% of permanently occupied homes have low-water-use percentiles below 42%, and 84% of seasonally occupied homes have percentiles greater than that. Using a threshold of 42% results in an overall classification accuracy of 81.9%. For Community B, the optimal threshold is 29%, which results in an overall classification error of 90%. However, these thresholds can only be identified for a water utility if the hourly water classification is already completed. Therefore, urban retail water suppliers attempting to classify homes with monthly water data will need to use a threshold without the benefit of this analysis. Taking Communities A and B as a sample, the combined optimal classification threshold is 35%. If we apply the 35% threshold to each community, this results in an 80.1% classification accuracy for Community A, and an 88.1% classification accuracy for Community B.

As in Section 4.3.1, the billing address is also a useful indicator for whether a home is seasonally or permanently occupied. If the billing address matches the service address, the home is much more likely to be permanently occupied. Therefore, homes with reasonable billing have a less strict criteria to be categorized as permanent compared to homes that have either duplicated or not reasonable billing address. Figure 4 demonstrates that, within a given community, the optimal threshold will vary depending on whether the billing address matches the service address. In both communities, more than 70% of homes with billing addresses that match service address are categorized as permanent, while the inverse is true for homes that have billing addresses that are either duplicated or are located in a different zip code from the service address. Therefore, to minimize classification error, the identification threshold for homes with billing addresses that match service addresses should be higher than the identification threshold for homes with billing addresses that are duplicated or do not match the service address.

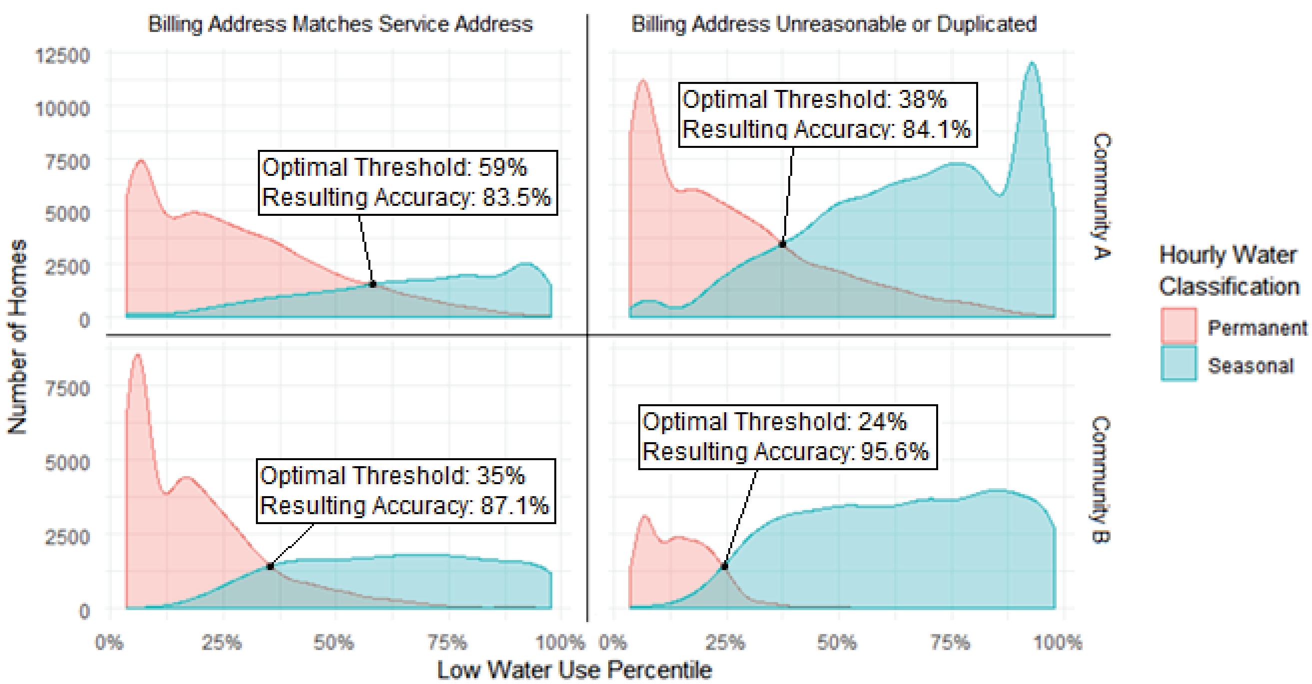

Figure 4: Distribution of low-water-use percentile in homes classified with the hourly water data method in the Community A (top row) and Community B (bottom row), and with billing addresses that

are either duplicated or do not match the service address (left column) and billing addresses that match the service address (right column). The area under the red (blue) lines indicates the number of permanent (seasonal) homes of a given low-water-use percentile.

In Community A, the error-minimizing threshold for homes with matching billing addresses is 59%, while it is 35% in Community B. For homes with billing addresses that do not match the service address, the threshold is 38% in Community A and 24% in Community B. The overall classification accuracy for using these optimal thresholds is 83.5% in Community A and 91.5% in Community B, and improvement of about 1.5% relative to the thresholds used in Figure 3. Combining the distribution of the two Communities allows CWEE to calculate an overall optimal threshold of 50% for homes with billing addresses that match their service address, and a threshold of 26% for homes with billing addresses that are either duplicated or are located in a different zip code. Applying these thresholds to both communities results in an 81.9% classification accuracy for Community A, and an 89.7% classification accuracy for Community B.

Table 3 summarizes the classification accuracy of the different approaches to calculating the threshold. Using the community-specific classifications to identify optimal thresholds always results in the best classification accuracy, but it is only possible to identify thresholds in this way when homes have already been classified by a different method. Using different identification thresholds depending on whether the billing address matches the service address also significantly improves identification accuracy in both communities.

Table 3: Summary of classification accuracies based on threshold calculation methodology with monthly water data

Threshold Calculation	Billing Address	Community A Classification Accuracy	Community B Classification Accuracy
Community-specific	Not Considered	81.9%	90.0%
Community-specific	Separated	83.5%	91.5%
Combined	Not Considered	80.1%	88.1%
Combined	Separated	81.9%	89.7%

Based on the results from the above analysis, CWEE proposes that urban retail water suppliers using monthly water data should classify homes as seasonal and permanent based on low-water-use percentiles grouped by whether the billing address is unique and whether it matches the service address on the account. Table 4 summarizes the proposed criteria to classify homes with monthly water data.

Table 4: Proposed classification criteria with monthly water data

Variable	**Permanent Homeowner**	**Permanent Renter**	**Seasonally Occupied**

Billing Address	Reasonable[1] and non-duplicated	Not reasonable[1] or duplicated	Not Applicable – Any home not classified as “Permanent” is Seasonal if annual water usage is greater than zero CCF.
Low-Water-Use Percentile	50^{th} Percentile and below*	26^{th} Percentile and below*	Not Applicable (See Above)

[1] Customers that use PO boxes as the billing address should be considered “reasonable” if the PO box is close to or in the same zip code as the service address.
** Proposed starting point to be adjusted to calibrate classification results to match census records*

The values in Table 4 are used as starting points for calibration. Note that all homes not classified as permanent are assumed to be Seasonally Occupied if their annual water usage is greater than zero. All homes with an annual water usage of zero are classified as vacant. Once all homes are classified, CWEE compares the total number of homes in each category to the values reported by the U.S. Census. Table B25003 from the American Community Survey summarizes the total number of owner-occupied and renter-occupied homes, and Table B25004 provides the number of homes for seasonal, recreational, or occasional use. The criteria for low-water-use percentile can be strengthened or weakened until the classifications match the most recent census data. Further, the homes that are classified as seasonally occupied with the lowest monthly water use can be further classified as vacant to match with other categories of non-permanent residence found in Table B25004.

3.4 Calculate Efficient Water Use Step

The final calculation step estimates the volume of efficient indoor water use for seasonal populations. This report assumes that this efficient water use is equal to the number of people-days of seasonal population multiplied by the IRWUS. Therefore, the primary aim of this calculation step is to estimate the people-days of the seasonal population within the service territory of an urban retail water supplier. In general, the number of people-days of the seasonal population is estimated by multiplying the number of seasonally occupied homes by the number of days per year each home is occupied and by the number of people in each seasonal home when occupied. However, the precise method of calculation will depend on the occupancy identification (as outlined in Section 4.2) and home classification steps taken (as outlined in Section 4.3).

If an urban retail water supplier has used hourly water or electricity data to detect occupancy and identify homes, then the method outlined in Section 4.4.1 is most appropriate. If monthly water data was used to identify seasonally occupied homes, the method in Section 4.4.2 is best. If neither of these approaches are feasible, the method outlined in Section 4.4.3 uses data from the U.S. Census to calculate the volume of efficient indoor water use of seasonal populations. Finally, Section 4.4.4 proposes a potential adjustment that can be applied using third-party information about vacation rentals.

3.4.1 Calculate efficient water use with daily occupancy data

The seasonal population efficient water use in gallons, V, can be calculated by multiplying the number of capita-days in which seasonally occupied homes are occupied by the IRWUS in gallons per capita-day, E. The number of people-days in seasonally occupied homes is calculated as the number of seasonally occupied homes, S, multiplied by the number of days per year those homes are occupied, O, and the average number of people inside each home, H. We can summarize this in an equation:

$$V = (S \times O \times H) \times E \qquad \textit{Equation 2}$$

Since S and O are given by the method in 4.3.1, the only remaining variables to account for are H and E. Since there is no reliable, non-intrusive method for measuring the number of people in each seasonally occupied home when it is occupied, estimates that rely on additional data sources are necessary. One possible and straightforward estimate is to use the average number of people per home among occupied homes in the urban retail water supplier's service area. This can be estimated with data from the U.S. Census. Dividing the total population (Table P1) by the total number of occupied housing units (Table S2501) will yield the average number of people per permanently occupied home. However, there is no way to verify whether the average number of people per permanently occupied home is equivalent to the average number of people in seasonally occupied homes. Seasonally occupied homes may have different property characteristics or use patterns that result in them having higher or lower average occupancies than permanently occupied homes in the same service area.

One way to account for differences in property characteristics of permanently and seasonally occupied homes is to use census data on home size. Dividing the average number of people per permanently occupied home by the average number of bedrooms per permanently occupied home will yield the average number of people per bedroom. Multiplying this number by the average number of bedrooms in each seasonally occupied home will yield an estimate for the average number of people in each seasonally occupied home when occupied. While this method may account for differences in property size, it will not account for any difference in occupancy patterns between permanently and seasonally occupied homes.

Once the number of person-days are determined, the final step is to multiply that number with the efficient water use. A baseline approach would be to use the current IRWUS. This approach assumes that seasonal occupants are, or should be, equally efficient in water use as permanent occupants. Among the participating water utilities, this appeared to be generally true – when occupied, seasonally occupied homes used similar amounts of water to permanently occupied homes.

3.4.2 Calculate efficient water use with monthly water data

If monthly water data is being used to classify homes (as outlined in Section 4.3.2), , but nothing is known about the occupancy patterns or the population of the homes. The fundamental calculation is similar to that in Section 4.4.1: calculate the number of capita-days in seasonally occupied homes, and multiply that number by the indoor residential water use efficiency E to calculate the seasonal population efficient water use volume, V. To calculate the seasonal population capita-days, we need to divide the total annual water use of seasonally occupied homes, A_S, by the average daily per-capita water use of permanently occupied homes. The daily per-capita water use of permanently occupied homes is calculated by dividing the average daily water use of permanently occupied homes, D_P, by the total population of permanently occupied homes, P_P. This calculation is summarized in Equation 3:

$$V = \frac{A_S}{D_P / P_P} \times E$$

Equation 3

Homes are classified as permanent or seasonal using the method in Section 4.3.2, so A_S can be calculated by adding up the annual water use of all homes that are classified as seasonally occupied and dividing that number by the number of homes classified as seasonally occupied. D_P can be calculated by first calculating the average daily water use for all homes classified as permanent, adding up those daily averages, then dividing that by the number of homes classified as permanently occupied. P_P can be obtained from Table P1 of the U.S. Census dataset.

3.4.3 Calculate efficient water use with census data only

Using the census data by itself can provide a simple calculation of seasonal population efficient water use. The overall approach is the same as Section 4.4.1 (see Equation 2) but would rely on average values and estimates for S, O, and H. The number of seasonally occupied homes, S, is available from the American Community Survey Table B25004. The population of seasonally occupied homes when occupied, O, could be calculated in the same way as proposed in Section 4.4.1. The simplest approach is to assume that seasonally occupied homes contain the same number of people as permanently occupied homes, on average. With this assumption, we calculate O by dividing the total population (Table P1) by the total number of occupied housing units (Table S2501).

For this method, the most challenging variable to estimate is the number of days in which seasonally occupied homes are occupied, since the census data does not provide any insight. In this study, using the occupancy detection method described in Section 4.2.1, seasonally occupied homes were on average occupied 40% of the time, or 146 days out of the year for Communities A and B. Different water suppliers may experience different occupancy rates of their seasonal populations, so 146 days may not be an appropriate estimate for all seasonal populations. However, the estimates for Communities A and B, which have different characteristics of their seasonal populations, are close to this 146-day average.

Therefore, calculating the seasonal population efficient water use using only census data requires two additional assumptions. The first assumption is that seasonally occupied homes, when occupied, contain the same number of people as permanently occupied homes. The second assumption is that the average number of days per year a seasonally occupied home is occupied is 146. With these assumptions, the seasonal population efficient use is calculated as the number of seasonally occupied homes, multiplied by the number of people per home in permanent houses, multiplied by 146 days, multiplied by the indoor residential water use standard in gallons per capita-day.

3.4.4 Adjust efficient water use calculation with vacation rental data

Stakeholders participated in the process of developing these methods have expressed concerns with the assumption that vacation rentals have proportional populations to permanently occupied homes. Vacation rentals, such as those available on Airbnb.com or VRBO.com, could be accounted for separately in the calculation of efficient water use to allow for different assumptions of their populations. Several external datasets exist that provide occupancy and size information for vacation rentals (Mashvisor 2021, AirDNA 2021). The data, available for free on AirDNA Marketminder, is the most useful data source CWEE identified, as it provides the number of active rentals, the occupancy rate of those rentals, and the average number of guests that vacation rental homes can accommodate, based on the listed maximum occupancy of all vacation rentals on Airbnb.com and VRBO.com. While there are limitations with these summary statistics, they provide valuable additional data. Multiplying the number of active rentals with the occupancy rate and the average maximum number of guests would generate a maximum value for seasonal population in vacation rentals.

This approach has advantages over previously identified methods for calculating efficient water use for vacation rentals. The other methods developed in this report have no external validation for identifying specific homes as seasonally occupied, or for the exact occupancy rate of those homes. AirDNA Marketminder gathers actual booking data from Airbnb and VRBO, and is therefore the only data source in this report that provides an independent measurement of occupancy capacities for seasonally occupied homes. The baseline assumption recommended by this report for its simplicity is that seasonally occupied homes house the same number of people as permanently occupied homes. However, if vacation rentals make up a large portion of seasonally occupied homes, and those homes have significantly higher capacities than the average permanently occupied home, it may be reasonable to account for this difference. AirDNA Marketminder provides maximum occupancy data for vacation rental properties, providing a second source of information to use beyond the population of permanently occupied homes. Both the baseline assumption and the assumption that vacation rentals are maximally occupied are potentially inappropriate, since neither is an accurate measurement of the actual number of people in seasonally occupied homes. Further, AirDNA is a private business, and it is not guaranteed that the data published on their website are accurate.

With these limitations in mind, the remainder of this section will detail a method that allows for the use of data from AirDNA Marketminder to adjust the water use calculation for vacation rentals. Since this method is an adjustment, it is first necessary to re-calculate the seasonal population efficient water use with $\acute{S}$ in place of S. $\acute{S}$ is calculated as the number of seasonally occupied homes S minus the number of vacation rentals S_R. The result is the seasonal population efficient water use for second homes $\acute{V}$. Then, we add the seasonal population efficient water use for vacation rentals, which is calculated by multiplying the capita-days in vacation rentals with the indoor residential water use efficiency standard E. The capita-days in vacation rentals is calculated by multiplying the number of vacation rentals, S_R, by the average occupancy rate of vacation rentals, O_R, and by the assumed occupancy rate of vacation rentals, H_R. H_R could be estimated using the average maximum occupancy of vacation rentals as provided by AirDNA Marketminder, though this could be an overestimate. Alternatively, H_R could be estimated using the baseline assumption that all vacation rentals have the same occupancy rate as permanently occupied homes, though this also could be a problematic assumption. Equation 4 summarizes this calculation.

$$V = \acute{V}(\acute{S} = S - S_R) + (S_R \times O_R \times H_R) \times E \qquad \text{Equation 4}$$

This adjustment relies on an external dataset that reliably catalogues vacation rentals and their occupancy rates, which may not be reliably available for all urban retail water suppliers. For now, it likely can increase the validity of results by helping to correct for the fact that high-occupancy vacation rentals may be more difficult to identify than other types of seasonally occupied homes with water and electricity data alone. Since such homes have high occupancy rates and may have higher population per home than other seasonally occupied homes, such an adjustment may have a large impact on the calculated seasonal population efficient water use.

4 RESULTS & DISCUSSION

In accordance with the three distinct methodologies (see Figure 2), these results are organized into three sections. The first section (5.1) demonstrates the results of occupancy detection methods, the second section (5.2) presents the results of home classification methods, and the third sections (5.3) presents results of seasonal population efficient water use calculation methods.

4.1 Occupancy Detection Results

CWEE developed two methods for detecting daily occupancy – one that uses hourly water data, and one that uses hourly electricity data. Of these, the method that uses hourly water data appears to be more practical for urban retail water suppliers and has a higher degree of accuracy. This section presents the results of both approaches and compares their relative performance and accuracy.

4.1.1 Detecting occupancy with hourly water data

CWEE used hourly water data to detect daily occupancy patterns in the Communities A and B. While no true ground-truth dataset was available, CWEE used a variety of methods to confirm that results were reasonable and accurate.

Figure 5 shows the daily occupancy rate of all homes in the Community B community for the years 2013 to 2017. The overall occupancy pattern matches closely with what representatives from Water Supplier B expected – higher occupancy during the peak summer months of June through September, as well as higher occupancy on weekends and much higher occupancy on major holiday weekends. Further, the “baseline” occupancy, represented by the dotted line which intersects weekday occupancy during winter months, is roughly 35%. This value matches closely to the estimated percentage of homes that are permanently occupied according to the U.S. Census.

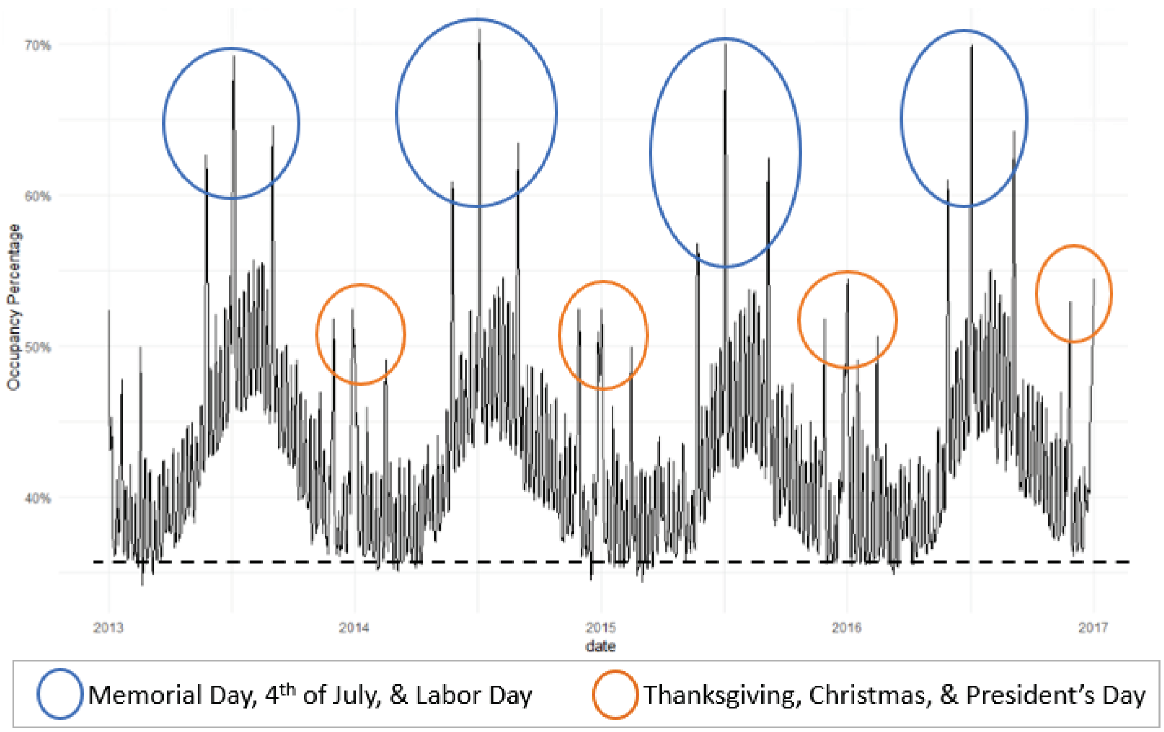

Figure 5: Occupancy percentage for Community B as calculated with hourly water data.

Figure 6 shows the daily occupancy rate of all homes in Community A for one year. The overall occupancy pattern matches closely with what representatives from Water Supplier A expected – higher occupancy during the peak summer months of June through September, as well as high occupancy during peak winter sports months of January to March. Again, a strong pattern of higher weekend occupancy and especially higher occupancy on holiday weekends persists. The baseline occupancy of about 50% again matches closely with the estimated percentage of homes that are permanently occupied according to the U.S. Census. The weekday-to-weekend occupancy pattern reveals that baseline occupancy during the summer is higher, with increases of less than 10% on weekends, while winter occupancy varies by over 10% from weekday to weekend. Representatives from the Water Supplier A confirm that many seasonal occupants will stay in the area for long periods of time in the summer, but often only visit for brief periods in the winter.

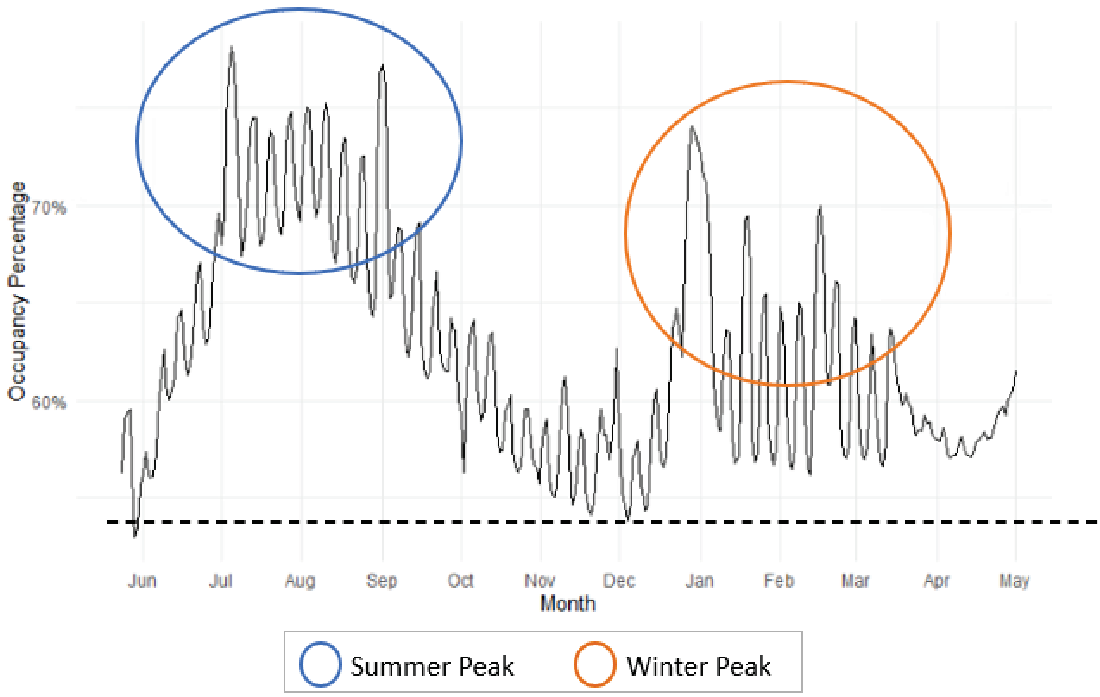

Figure 6: Occupancy percentage in Community A as calculated with hourly water data

Overall, it seems that this occupancy detection method estimates reasonable patterns of occupancy in the two settings it has been tested. On a household level, these readings are also remarkably consistent. In Community B, 80% of all homes had less than a 20% variation in annual occupancy percentage over the course of 5 years, and 50% of homes had less than 10% of variation in annual occupancy percentage using this hourly water use data to estimate occupancy.

4.1.2 Detecting occupancy with hourly energy data

Detecting occupancy with hourly energy data appears to lead to slightly less stable and reliable results than the method that uses hourly water data. Figure 7 shows that the overall occupancy chart for the Community C has a less-defined pattern than those observed in Figures 4 and 5. Weekends and holidays do not appear to cause as significant an increase in occupancy as they do in other communities. Part of the explanation for this is that the Community C is fundamentally different from Communities A and B. Over 70% of all homes are permanently occupied, so overall occupancy is likely dominated by the occupancy patterns of permanent residents. This population may be more likely to leave on weekends and holidays, which could offset the occupancy patterns of the seasonal population. Still, Community C demonstrates a clear pattern of increased occupancy during winter months, starting in November and decreasing in March and a second smaller peak in July and August.

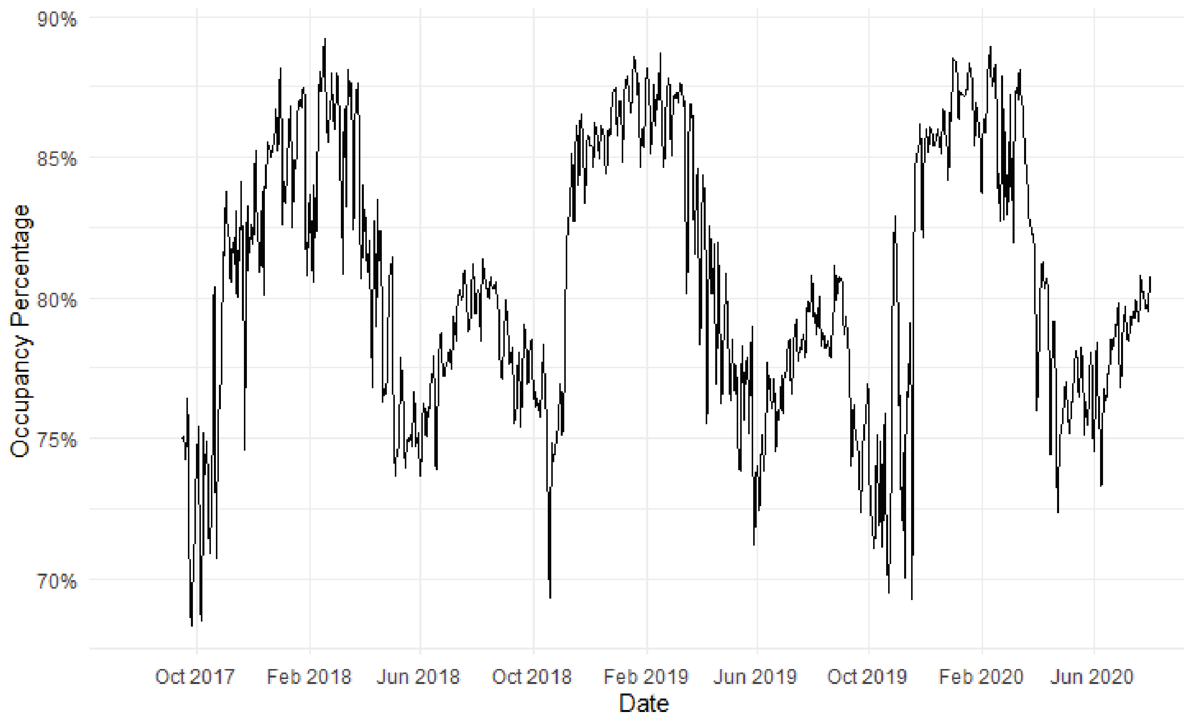

Figure 7: Occupancy percentage in the Community C as calculated with hourly electricity data

Daily occupancy results with both datasets are compared in Community B in Figure 8. In this case, the electricity occupancy detection seems to capture the weekly patterns of occupancy just as well as the water detection. However, the electricity results seem to suggest a slightly different overall trend of seasonality. In addition to the summer peak, it also captures a winter peak that is approximately 10% higher than what is predicted with hourly water data. One explanation for these results is that, in winter, there are some electricity uses that lead toward estimating more days as occupied compared to the approach that uses water data. Examples of such electricity uses include automatic lights, electricity use to heat homes, and any other automated or passive electricity use that can be associated with colder temperatures and longer nights in winter months. It is also possible that the water identification method underestimates winter occupancy rates because of lower overall water use due to reduced irrigation or other water use associated with summer months. However, the seasonal trend captured by the water data matches expectations for this community more closely, so it is more likely that electricity data overestimates winter occupancy than that water data underestimates it. While the electricity occupancy detection method seems to generally produce reasonable results, the water detection method appears to be more reliable, accurate, and stable.

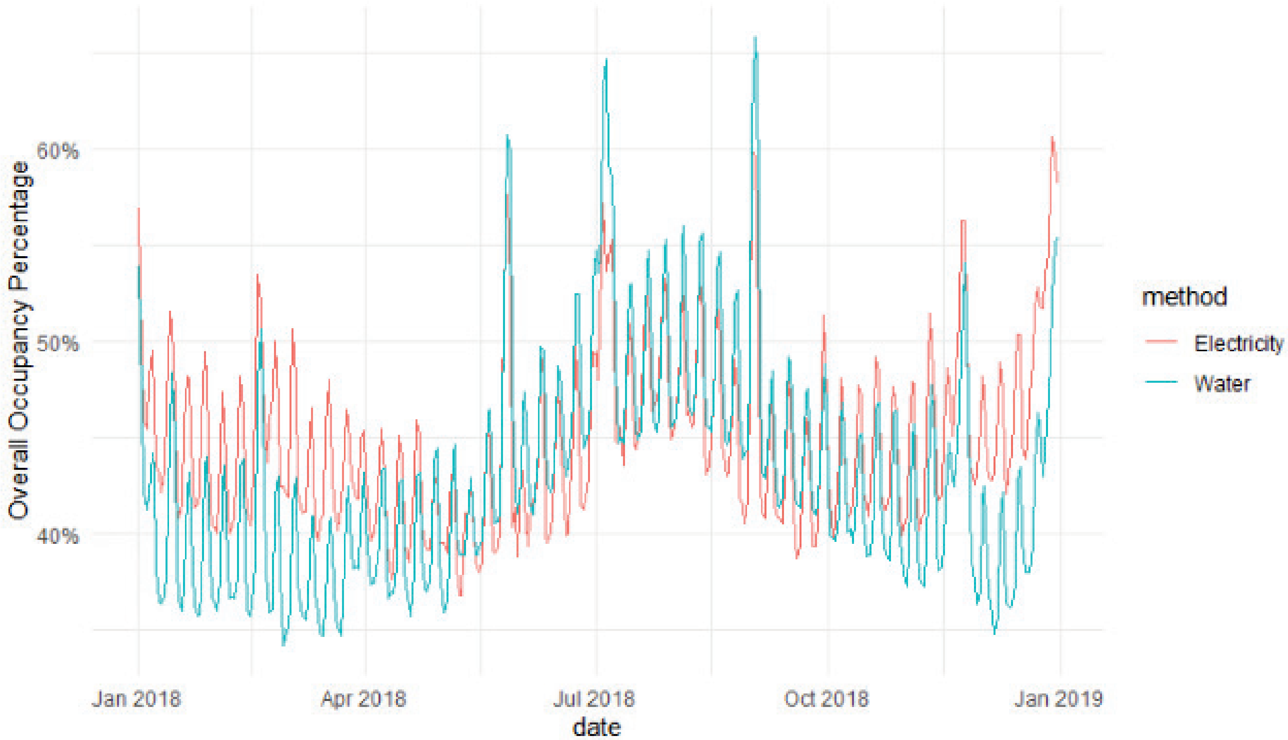

Figure 8: Occupancy percentage for all homes in the Community B calculated with electricity and water data. The water occupancy percentage matches more closely with expected seasonal patterns.

4.2 Home Classification Results

Home classification with daily occupancy data or with monthly water data provides critical input for two of the proposed methods to calculate the seasonal population efficient water use. In this section, we analyze the results of the home classification methods to confirm that the results are reasonable and to assess the level of agreement between different methods for classifying homes.

In the Community B, we observe that homes classified as Permanent Homeowner, Permanent Renter, and Seasonal all have nearly identical water use when unoccupied, but that seasonal homes generally use more water when occupied (Figure 9). The discrepancy is approximately 25 gallons per home per day in the peak travel season, and about 35 gallons per home per day in the peak season. This difference implies that either seasonal homes typically house more individuals than permanent homes when occupied, or that occupants in seasonal homes typically use more water per person than occupants in permanent homes.

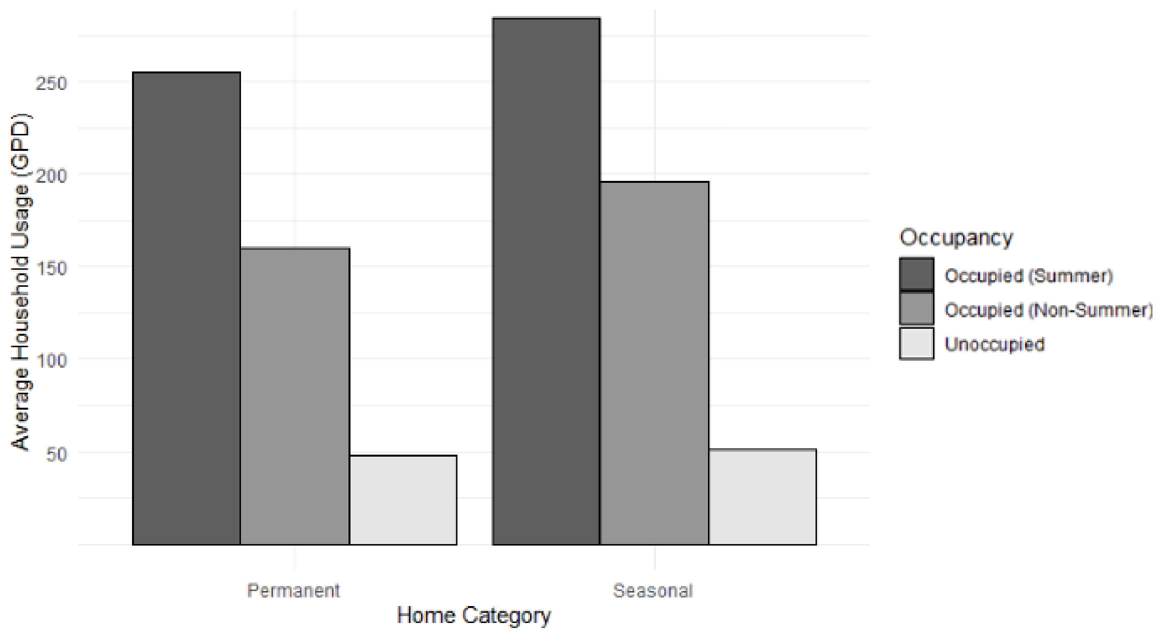

Figure 9: Average occupied and unoccupied household water usage in Community B homes

In Community B, seasonally occupied homes tend to use more water than permanently occupied homes in winter months, but less water in summer months (Figure 10). One explanation is that seasonally occupied homes in Community B have different characteristics from permanently occupied homes, such as smaller landscapes, that result in less water in the summer. It is also possible that the characteristics of seasonal populations are different in summer and winter, where winter visitors are more likely to crowd inside smaller homes because they are there for shorter periods of time.

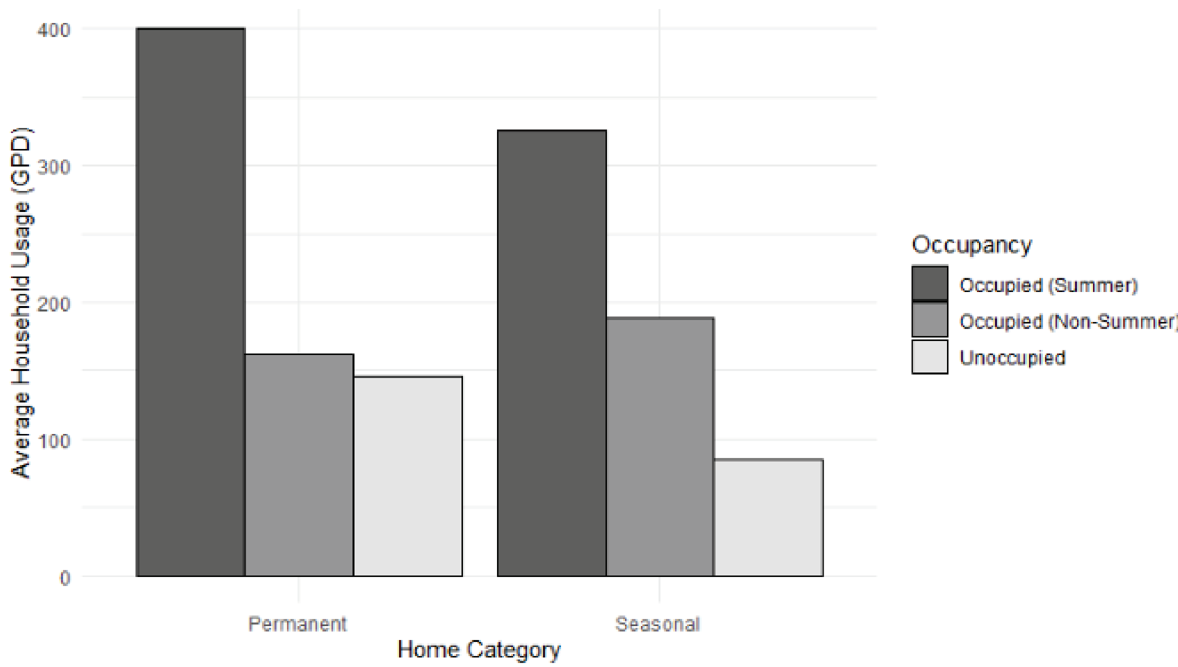

Figure 10: Average occupied and unoccupied household water usage in Community A homes

Water Supplier C does not have daily water usage data, and homes were classified using both hourly electricity data and monthly water data. Community C also covers a large heterogenous area, where some parts of the city have high proportions of seasonal populations, while others have much smaller proportions. Therefore, the accuracy of the two home classification methods can be assessed based on how accurately they capture the heterogeneity of seasonal population across zip codes within the service territory of Water Supplier C. Figure 11 demonstrates that classifying homes with hourly electricity data produces more accurate results in every single zip code. Both approaches were calibrated to match the census data on the scale of the entire service territory, and these results indicate that the calibration step could be further refined by performing it on the zip code level instead.

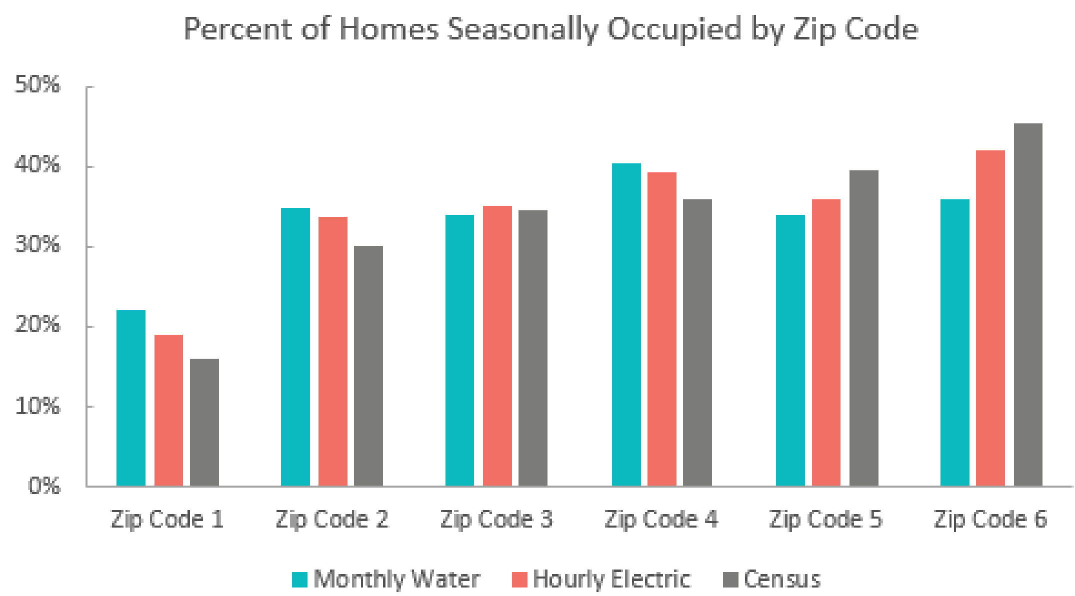

Figure 11: Comparison of classification results in Community C using the Monthly Water method, the Hourly Electric method, and actual Census data for seasonally occupied homes

Another way to assess home classification techniques it to evaluate the theoretical impact on the calculated volume of efficient indoor water use of seasonal populations. When applying the standard to all residential accounts, total water use is divided by only the permanent population, leading to an inflated and inappropriate measure of gpcd. Home classification allows for the water use of permanent accounts only to be divided by the permanent population, yielding a more appropriate measure of per-capita water use. Figure 12 shows that, in Community B, permanent homes use far below 55 gpcd when classified appropriately, regardless of classification technique. Similar results were observed in Community A and B, as the removal of homes classified as “seasonal” significantly improves the efficiency measurement of the remaining homes.

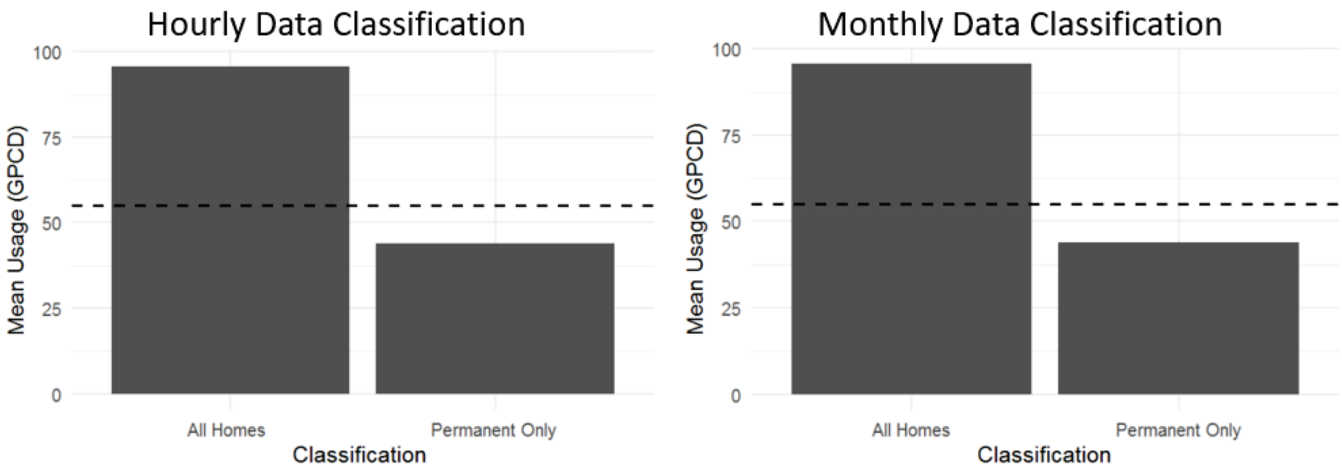

Figure 12: Estimated total GPCD of all homes and permanent-only homes using the hourly water classification and monthly water classification techniques in Community B

The three approaches to classifying homes generally had a high agreement. To analyze home classification accuracy, we assume that home classification using hourly water data is the most accurate. By comparing the classification results from other methods to the classification using hourly water data, we can calculate the relative accuracy of other methods.

In the two communities that had hourly water data, the classification using monthly water data led to relatively low error rates. Monthly water classification performed best in the Community B, where only 2.5% of homes were incorrectly categorized as seasonally occupied. The monthly water use method was less effective in Community A, which has a higher permanent population and a lower degree of occupancy seasonality, resulting in an error of 5.2%. These results suggest that the monthly water classification technique can result in a classification agreement of 90% to 95% with the hourly water method.

Home classification with hourly electricity and hourly water data also seem to have a high degree of agreement. In Community B – the only community with both hourly datasets, saw a 93% agreement in home classification. That is, approximately 3.5% of homes classified as seasonal with hourly water data were classified as permanent with hourly electricity data, and vice versa. In Community C, the agreement between classification using hourly electricity data and monthly water data was slightly lower at 87%.

These results indicate that home classification with monthly water data can produce similar results to classification from hourly water data. Classification with hourly electricity data also seems to lead to similar results, but may be slightly less accurate than the simpler approach that uses hourly water data.

4.3 Seasonal Population Efficient Water Use Volume Calculation Results

The seasonal population efficient water use calculations typically produce values that are close to the actual usage of seasonally occupied homes. Table 5 shows that the methods that use hourly water and energy data typically generate smaller calculations of efficient water use than those generated by monthly water data and census data alone. The seasonal population efficient water use generated with monthly water use typically produces the largest calculation of efficient water use, and often closely match the calculations based on census data alone. The methods that use hourly data produce calculations of efficient water use closest to actual water usage in seasonal homes in Communities A and B.

Table 5: Calculated seasonal population efficient water use volume for each method in million gallons (and million gallons per seasonally occupied home)

Community	Actual Use	Hourly Water	Hourly Electricity	Monthly Water	Census Data
A	161 (0.013)	250 (0.020)	No data	212 (0.017)	251 (0.020)
B	136 (0.018)	134 (0.018)	140 (0.018)	227 (0.029)	213 (0.027)
C	224 (0.014)	No data	296 (0.018)	336 (0.021)	322 (0.020)

Figure 13 replicates the per-household figures in Table 5. All calculated seasonal population efficient water uses are between 0.01 and 0.03 millions of gallons per home. All of the calculations allow for equal or greater water use than seasonally occupied homes already consume. The efficient water uses calculated with monthly water use and with census data generally exceed the actual water use of seasonally occupied homes by the greatest number.

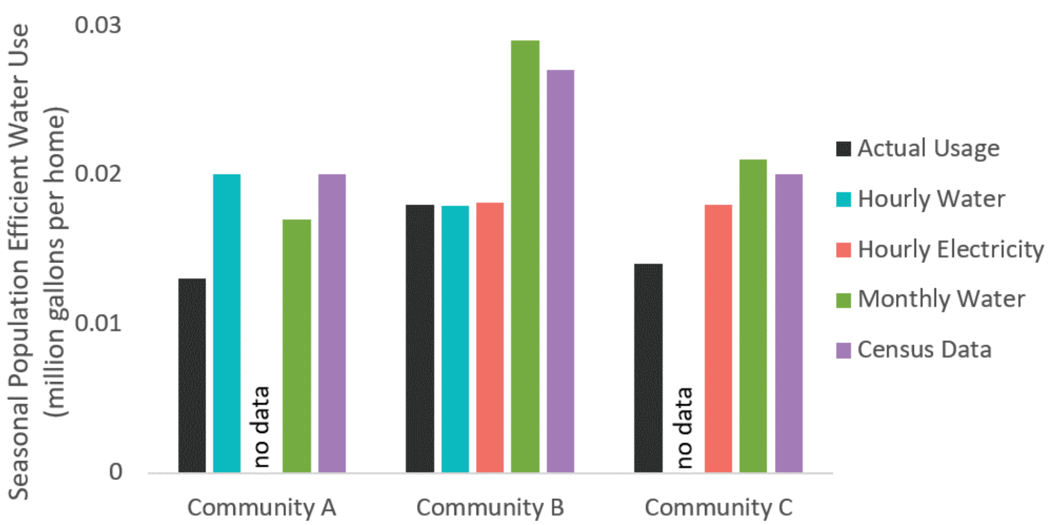

Figure 13: Actual usage and calculated seasonal population efficient water use for each identified seasonal occupied home in all three communities

5 CONCLUSION

The methods outlined in this report generally produce reasonable calculations of efficient water use for seasonally occupied homes in the three communities in which the methods were tested. The methods that generate daily occupancy patterns, using hourly water and electricity data, seem to generally produce the most accurate and useful results, but these methods also require a significant amount of high-quality data and large-scale data processing and analysis. The monthly water use method also seems to produce reasonable results and it only requires data that is already available to most urban retail water suppliers. The census method does not require any significant data processing and produces results that are generally in line with the other proposed methods.

REFERENCES

AirDNA. 2021. Marketminder: Cutting-edge vacation rental data. AirDNA. Retrieved September 15, 2021, from https://www.airdna.co/vacation-rental-data

California Department of Water Resources. 2021. Water Use Efficiency Water Use Studies Workgroup – Workshop: Variance – Seasonal Populations [PowerPoint Slides]. June 10.

Causone, F., Carlucci, S., Ferrando, M., Marchenko, A., and Erba, S. 2019. A data-driven procedure to model occupancy and occupant-related electric load profiles in residential buildings for energy simulation. Energy and Buildings, 202, 109342.

Cecile Proust-Lima, Viviane Philipps, Benoit Liquet. 2017. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. Journal of Statistical Software, 78(2), 1-56. doi:10.18637/jss.v078.i02

Huchuk, B., Sanner, S., and O'Brien, W. 2019. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Building and Environment, 160, 106177.

Kleminger W, Staake T, and Santini S. 2013. Occupancy Detection from Electricity Consumption Data. Build Sys ’13(14-15).

Mashvisor. 2021. raditional and airbnb investment property. Mashvisor Inc. Retrieved August 18, 2021, from https://www.mashvisor.com/

National Weather Service Climate Prediction Center (NWS CPC). 2022. Degree Day Outlook for Major United States Cities. National Oceanic and Atmospheric Administration. Retrieved June 13, 2022 from https://www.cpc.ncep.noaa.gov/pacdir/DDdir/INT1.html

Norris, A. 2021. Marketminder: Cutting-edge vacation rental data. AirDNA. Retrieved September 20, 2021, from https://www.airdna.co/vacation-rental-data.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Razavi, R., Gharipour, A., Fleury, M., and Akpan, I. J. 2019. Occupancy detection of residential buildings using smart meter data: A large-scale study. Energy and Buildings, 183, 195-208.

U.S. Census Bureau. 2021. Selected housing characteristics, 2015-2019 American Community Survey 5-year estimates. Retrieved September 10, 2021, from https://data.census.gov/cedsci/.

Wang, C., Jiang, J., Roth, T., Nguyen, C., Liu, Y., and Lee, H. 2021. Integrated sensor data processing for occupancy detection in residential buildings. Energy and Buildings, 237, 110810.

APPENDIX A – Detecting Seasonal Occupancy with Hourly Water Data

This appendix provides a more detailed and specific demonstration of detecting seasonal occupancy with hourly water data than is provided in the methods Section 3.2.1. Urban retail water suppliers with advanced metering infrastructure (AMI) may have access to hourly water use data for each residence in their service area. Household water use provides the most reliable signal of occupancy of all other data sources explored in this report.

The rule-based approach is algorithmic, requiring no statistical or machine-learning methodology. Under ideal conditions, the simplest rule can be used to detect occupancy; when the meter detects water use on a given day, categorize it as "occupied".

Figure 14 demonstrates the application of this simple rule. However, there are two primary obstacles that limit the reliability of this rule. First, the water meters usually measures both indoor water uses and outdoor water uses. This means that automated outdoor irrigation will register on the water meter and would cause a false signal of occupancy even when nobody is using water within the residence. The second obstacle to identifying occupancy is the presence of leaks. Slow leaks can cause meters to register continuous or periodic usage, even when no person is using water at a residence. Additional rules must be incorporated to handle water usage that is not associated with residential occupancy.

Date	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	Category
Sun, 2/17/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Mon, 2/18/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Tue, 2/19/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Wed, 2/20/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Thu, 2/21/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Fri, 2/22/2013	0	0	0	**8**	**7**	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	Occupied
Sat, 2/23/2013	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	Occupied
Sun, 2/24/2013	0	0	0	0	0	**1**	**1**	**2**	**1**	0	0	0	0	0	0	0	**1**	0	**4**	0	0	0	0	0	Occupied
Mon, 2/25/2013	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	**4**	**3**	0	0	0	0	0	0	0	0	Occupied
Tue, 2/26/2013	0	0	0	0	0	0	**1**	0	**5**	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	Occupied
Wed, 2/27/2013	0	0	0	0	0	0	0	**1**	0	0	0	0	0	0	0	0	**1**	**3**	0	0	0	0	0	0	Occupied
Thu, 2/28/2013	0	0	0	0	0	0	**3**	**3**	**1**	**1**	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	Occupied
Fri, 3/1/2013	0	0	0	0	0	0	**2**	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	Occupied
Sat, 3/2/2013	**1**	**8**	0	**1**	0	0	**1**	**2**	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	Occupied

Figure 14: Sample hourly water data. Each row is a day, and each column is an hour. Water use bold and outlined in thick black borders indicate water use associated with occupancy, and days categorized as occupied are highlighted in yellow.

Removing Irrigation Water Use

Water use associated with scheduled irrigation is relatively straightforward to identify. Irrigation water use is typically much higher than indoor water uses, and is usually above 10 cubic feet per hour, whereas the average individual's indoor water use is about 7 cubic feet per day. Also, automated irrigation usually follows strong patterns based on time of day and day of week. Therefore, rule-based algorithms can simply look for high water uses that are repeated at the same time of day and on the same day of the week over time, and simply flag those uses as irrigation. These assumptions may not hold for all communities, particularly ones that are especially dry or have heavy restrictions on irrigation practices, but they were effective for the communities tested in this study. The specific rules can be adjusted, but the following describes the current set of rules currently being tested.

Irrigation detection protocol:
Step 1: Define minimum hourly water usage to categorize as irrigation (default 10 cubic feet)
Step 2: Generate a *vector* containing all hourly water uses for a given day of the week and a given hour of the day

> e.g., Start with Mondays at 12:00 AM. The vector would contain hourly usage for 12:00AM on every Monday in the sample, creating a string of numbers that looks like: (…0, 0, 0, 1, 0, 0, 2, 0…)

Step 3: For each reading in the vector, calculate the *percentage difference* between it and the previous reading. Formula = $(CF_{x-1}-CF_x)/CF_x$
Step 4: For each reading in the vector, if the total usage is greater than 10 cubic feet AND the *percentage difference* is less than 10%, <u>flag the hour as "Irrigation"</u>
Step 5: Repeat steps 2-4 for each day of week & hour combination (7 * 24 = 168 vectors)
Step 6: For every hour flagged as "irrigation" with this method, also flag the proceeding and following hours as "irrigation".
Step 7: Remove all hours flagged as irrigation from the dataset to continue processing
Figure 15 demonstrates the result of applying the above protocol to a home's water use data. All of the water use at 4:00 AM on Wednesdays, Fridays, and Sundays are flagged as irrigation and removed, as are the water uses in hours immediately before and after that hour. This causes three days that would otherwise be categorized as occupied to instead be categorized as unoccupied but has no impact on three other days that also have water uses in other periods of the day.

Date	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	Category
Mon, 6/3/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Tue, 6/4/2013	0	0	0	8	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Wed, 6/5/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	0	0	0	0	0	Occupied
Thu, 6/6/2013	0	0	0	8	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Fri, 6/7/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Sat, 6/8/2013	0	0	0	9	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Sun, 6/9/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied
Mon, 6/10/2013	31	4	11	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Occupied
Tue, 6/11/2013	5	23	0	8	46	0	0	0	0	0	0	0	0	0	0	0	0	16	11	6	0	0	0	0	Occupied
Wed, 6/12/2013	0	9	2	0	0	0	0	0	0	0	0	0	0	0	0	14	17	10	0	0	0	0	0	21	Occupied
Thu, 6/13/2013	6	0	1	8	47	0	0	0	0	0	0	0	0	1	1	3	1	0	0	0	0	0	1	0	Occupied
Fri, 6/14/2013	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	1	Occupied
Sat, 6/15/2013	0	0	0	12	45	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	Occupied
Sun, 6/16/2013	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Unoccupied

Figure 15: Sample hourly water data from account with automatic irrigation. Hourly usage with a red border has been flagged as "irrigation" and is not considered water usage associated with occupancy. As a result, days with only irrigation water use are categorized as "Unoccupied".

Removing Leaks

Removing leak readings is particularly challenging with a rule-based algorithm because they are difficult to distinguish from sporadic small water usage by a person. Our assumption must be that usage patterns of humans are more varied, and some hours will have significantly higher uses than others. However, this may not be true every day. For example, a person may spend only a small amount of time at home on a given day and might not use the shower or cook in that time. As a result, we would see a handful of hours registering one or two cubic feet of usage. A leak may look similar, but often has the characteristics of lasting for several consecutive days and, alternating between readings of 0 and 1 or 1 and 2, and continuing at all times of day.

Leak detection protocol (positive displacement meters):

Step 1: For each day in an account, determine whether all values are either 0 or 1, or 1 and 2 CF.
All 0s and 1s implies a leak with an average rate of less than 1 CF/hour.
All 1s and 2s implies a leak with an average rate of between 1 and 2 CF/hour
Step 2: For each day calculate the "evenness" of the leak rate as the difference between the number of hours with a reading of 0 and the number of hours with a reading of 1.
Step 3: For each day calculate the "consistency" of the potential leak as the number of hours where the less common value repeats itself
e.g., if "0" is the less-common reading, and a day has one instance of consecutive readings of "0", then the "consistency" score is 1.
Step 4: Flag a day as "leak-only" **IF** all values are 0 and 1 or 1 and 0 AND
- The "evenness" is less than **6** OR
- The consistency score is **0**
Step 5: Remove all days flagged as "leak only" from the sample
The reasoning behind these rules is as follows. First, we assume leaks are relatively steady outputs at or below 2 cubic feet per hour. We assume leaks greater than 2 cubic feet per hour are likely to be rare and noticeable enough that it is unlikely such leaks would persist long enough to cause a home to be miscategorized. If there is no other usage during a day aside from a leak, then readings will alternate between the two values closest to a leak. For example, a leak at about 0.3 cubic feet per hour will result in readings of either 0 or 1. A leak of about 1.3 CF per hour will result in readings of either 1 or 2. Leaks with more variance in flow rate will likely be missed by this method.

Second, we also need to confirm that the readings are either even or consistent. An "even" designation occurs when the flow rate of the leak is close to 0.5 CF/hr or 1.5 CF/hr. In this event, the number of 0's and 1's (or 1's and 2's) will be relatively even during a day, and the total usage in the first half of the day will be close to the total usage in the second half of the day. Uneven readings could imply that the leak rate is closer to 0, 1, or 2 than it is to a midpoint between any of those numbers. In that case, we check for consistency. If there is a constant flow rate, but there are more of one value than another (e.g., more 0s than 1s), we check to see if the less common value ever occurs twice in a row. For example, if a day has 20 0's and 4 1's, we would consider it "consistent" if none of the 1's occurred in consecutive hours.

None of the above rules alone to determine that a day has no occupancy, however. A day with sporadic and low water use could meet some or all of the above criteria. Therefore, the final test to determine whether a day's water use is entirely leaks is that either the previous or following day must also meet the same criteria. That way, the rule requires consecutive days of readings that appear to be leaks in order to rule out occupancy in those days. A home with a leak with many days matching the above criteria is demonstrated in Figure 16.

Date	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	Category
Sat, 9/17/2016	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	Unoccupied
Sun, 9/18/2016	1	0	0	4	2	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	5	1	0	1	Occupied
Mon, 9/19/2016	0	1	0	1	0	1	2	0	1	0	0	1	0	0	1	1	2	1	1	0	1	1	0	0	Occupied
Tue, 9/20/2016	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	Unoccupied
Wed, 9/21/2016	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	Unoccupied
Thu, 9/22/2016	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	Unoccupied
Fri, 9/23/2016	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	Unoccupied
Sat, 9/24/2016	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	Unoccupied
Sun, 9/25/2016	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	Unoccupied
Mon, 9/26/2016	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	Unoccupied
Tue, 9/27/2016	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	Unoccupied
Wed, 9/28/2016	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	Unoccupied
Thu, 9/29/2016	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	Occupied
Fri, 9/30/2016	1	0	0	0	3	1	0	1	0	0	1	0	0	0	0	1	1	3	0	0	0	1	0	1	Occupied

Figure 16: Sample hourly water data from account with a slow (<1 cubic foot per hour) leak. Hours bordered red are removed from the analysis. Days categorized as "occupied" have at least one hour with usage greater than 1 cubic foot, and all usage in those days is considered to be associated with occupancy, even if some of the usage is due to the leak.

After removing irrigation and designating days as "leak-only", the rule-based algorithm will categorize a day as "occupied" if it has at least one hour of recorded usage.

APPENDIX B – Detecting Seasonal Occupancy with Hourly Electricity Data

This method uses account-level hourly electricity data to determine daily occupancy for each home. CWEE developed a method of detecting daily occupancy by combining daily total electricity consumption for each home with daily temperature data measured at the nearest weather station. In particular, this study used heating degree-days (HDD) and cooling degree-days (CDD) to capture differing responses to temperature depending on whether the day is above or below 65 degrees Fahrenheit. Then, CWEE solved a latent class mixed model (LCLMM) for each home.

The LCLMM estimates α, the energy use when the daily average temperature is at 65 degrees Fahrenheit and the home is occupied, β_1, the change in energy use for each CDD when the home is occupied, and β_2, the change in energy use for each HDD when the home is occupied. The second equation estimates $\tilde{\alpha}$, the energy use when the daily average temperature is at 65 degrees Fahrenheit and the home is unoccupied, β_1, the change in energy use for each CDD when the home is unoccupied, and β_2, the change in energy use for each HDD when the home is unoccupied. The model then estimates two different models in the way that best explains the data, and assigns each day a probability that it belongs to either the first or second class – and is therefore either occupied or unoccupied. Equation 1 below shows the equation represented symbolically.

$$E_{i,d} = \begin{cases} a + \beta_1 CDD_{i,d} + \beta_2 HDD_{i,d} \\ \tilde{a} + \widetilde{\beta_1} CDD_{i,d} + \widetilde{\beta_2} HDD_{i,d} \end{cases}$$

Equation 1

This method is justified by reviewing patterns of daily water use and daily temperature in homes in Community B (Figure 17). For these six sample homes, homes seem to have different relationships between average temperature and energy use depending on whether they are occupied.

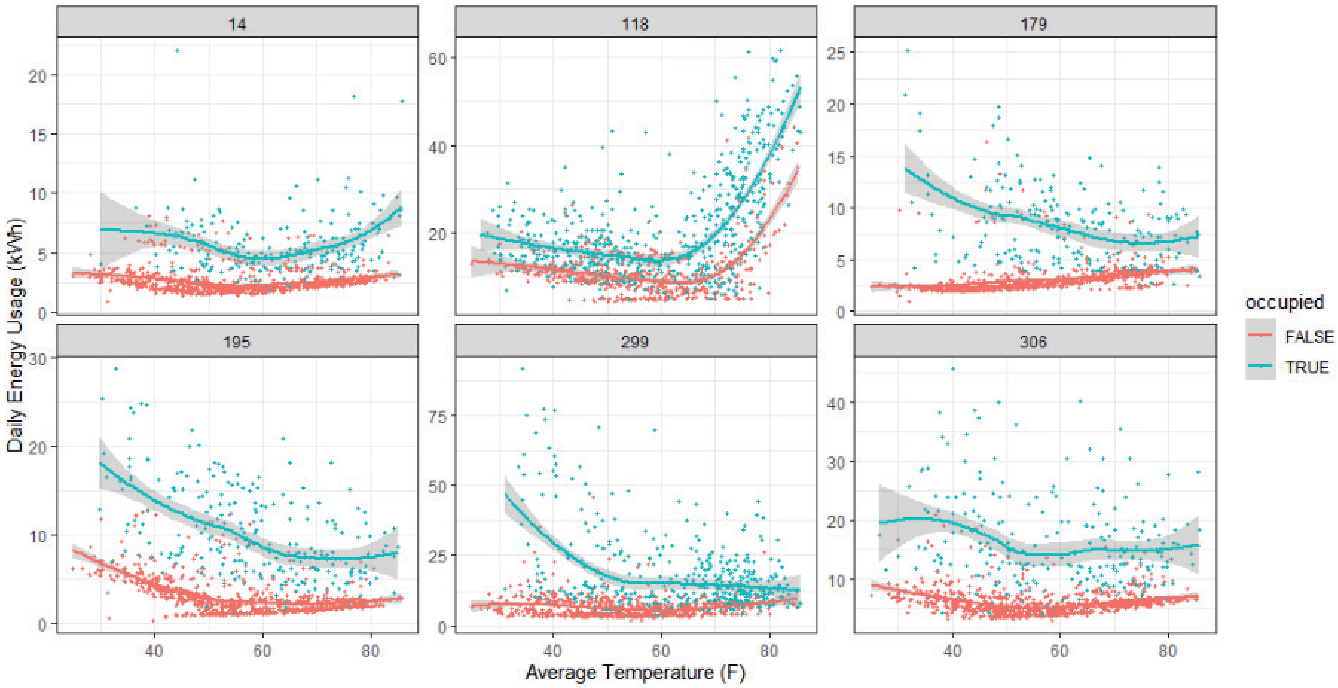

Figure 14: Patterns of daily energy use and daily average temperature for six sample homes in Community B with daily occupancy detected with the hourly water method

The application of the methodology yields promising results in many cases. Figure 18 demonstrates a scatter plot of one of the sample homes comparing daily average temperature and daily energy use. The electricity classification method identifies a very similar pattern of days as occupied when compared to the water categorization method, leading to a similar overall result in total occupancy. Many individual days with moderate energy use are categorized differently between the two models, however.

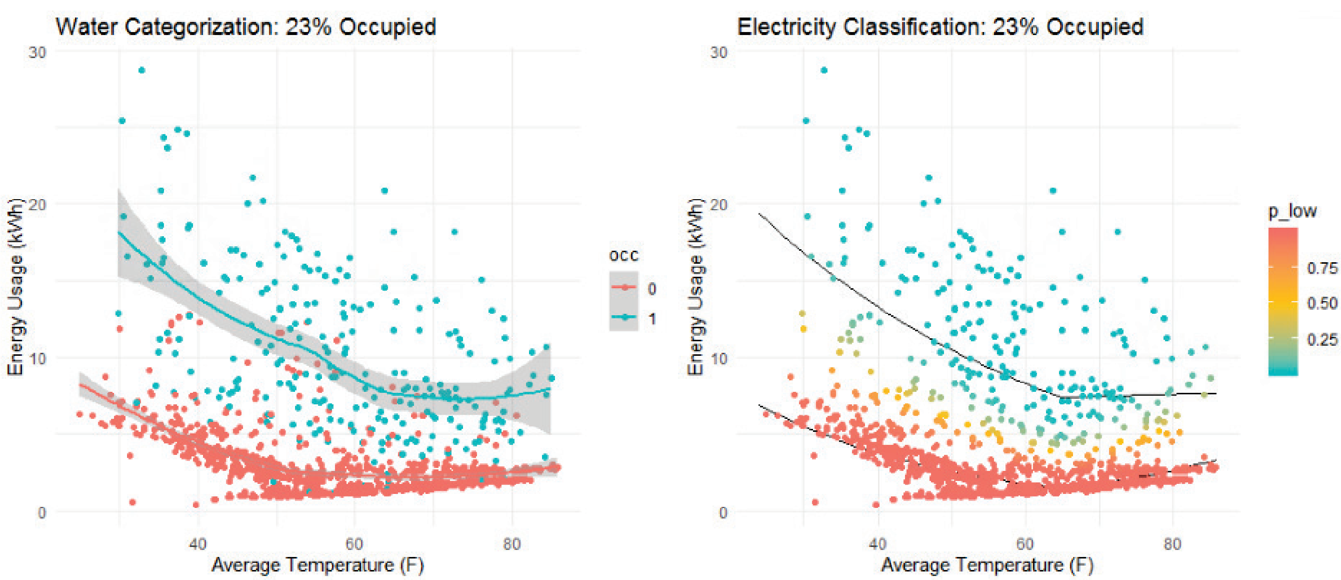

Figure 15: Example output of electricity occupancy detection compared to water detection results

Figure 19 shows the outcome of electricity categorization when the relationship of energy use to temperature in a home does not change significantly depending on when it is occupied. The large number of yellow points in the electricity classification plot indicates that the model is uncertain whether the day is occupied or unoccupied.

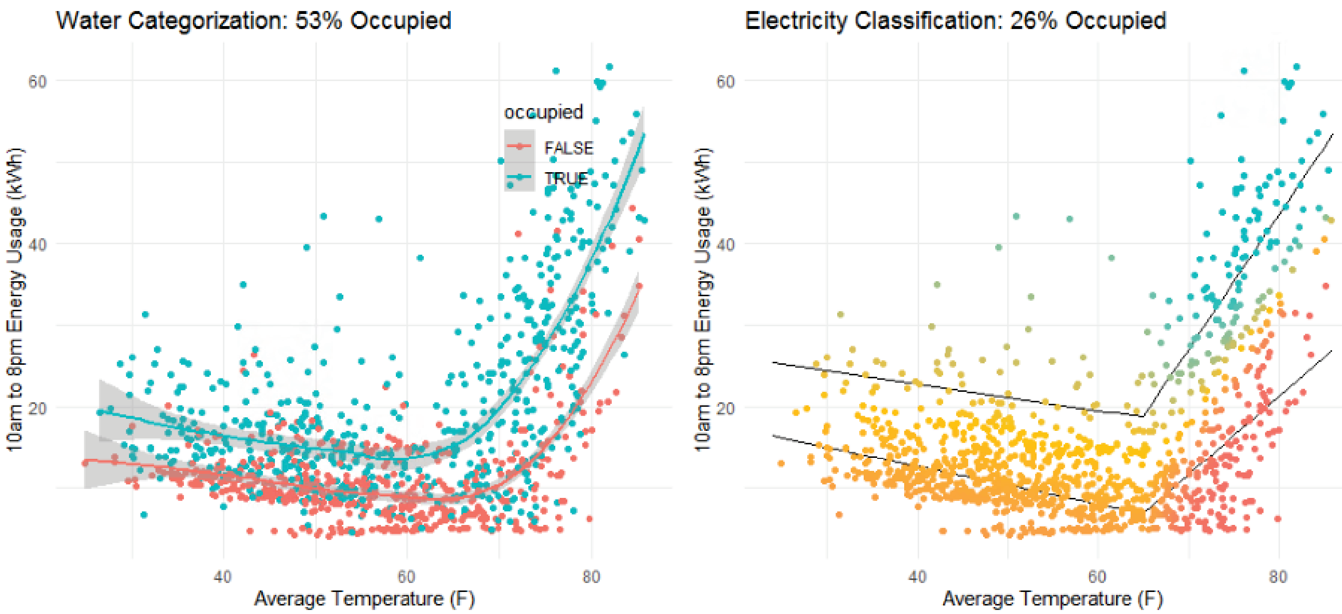

Figure 16: Example output of electricity occupancy detection compared to water detection results