

ARTICLE OPEN



Long-term water conservation is fostered by smart meter-based feedback and digital user engagement

Andrea Cominola^{1,2}✉, Matteo Giuliani³, Andrea Castelletti³✉, Piero Fraternali³, Sergio Luis Herrera Gonzalez³, Joan Carles Guardiola Herrero⁴, Jasminko Novak^{5,6} and Andrea Emilio Rizzoli⁷

Consumption-based feedback has been demonstrated to encourage water conservation behaviors. Smart meters and digital solutions can support customized feedback and reinforce behavioral change. Yet, most of the studies documenting water conservation effects induced by feedback and smart meter data visualization evaluate them in short-term experimental trials only. Here we show that water conservation behaviors promoted by smart meter-based consumption feedback and digital user engagement interventions might persist in the long term. We developed an analysis of 334 households in Valencia, Spain. We find that approximately 47% of the households engaged in our water conservation program achieved a long-term 8% reduction of volumetric water consumption, compared with pre-treatment observations. Water conservation behaviors persisted more than two years after the beginning of the program, especially for the households receiving sub-daily smart meter information. Our results provide empirical evidence that smart meter-based water consumption feedback and digital user engagement can effectively promote durable conservation behaviors.

npj Clean Water (2021)4:29; <https://doi.org/10.1038/s41545-021-00119-0>

INTRODUCTION

Changing individual and community water consumption behaviors is essential to achieve community-wide and state-wide water conservation targets and address water security in the near future^{1–3}. Consumption-based feedback and user engagement are powerful enablers of behavioral change as they overcome the drawbacks of mandatory interventions^{4,5}. The implementation and extent of conservation strategies such as mandatory water usage restrictions or price-based strategies are limited by ethics, equity, and acceptance issues, along with the generally low price elasticity of water demand^{6–8}.

The effectiveness of feedback interventions to promote conservation behaviors has been extensively documented in the energy sector^{9–13}. Behavioral programs with randomized controlled trials led to savings ranging from 1% to over 20% also in the absence of volunteer selection bias and monetary incentives^{14,15}. While there is less evidence on the effectiveness of similar programs in the water sector, the increased deployment of smart water meters and the digital transformation of the utility sector is revealing the potential for the customization of consumption-based feedback^{16,17}. A growing body of experimental and observational studies shows that short-term water savings between 2.5% and 28% can be achieved by consumption-based feedback in near-real time⁴. However, the variability of these empirical results does not support a conclusive assessment of the effectiveness of consumption-based feedback due to their differences in research design, context, type of feedback, and sample size. In addition, understanding how consumption-based feedback impacts long-term behavior change, before rebound effects emerge, remains an open question, hindered so far by the limited time frame of most behavioral studies and smart water metering trials^{4,17}.

Here, we quantify for the first time the long-term effects of smart meter-based water consumption feedback and digital user engagement on residential water consumption. We formulate this overarching research question: *Can smart meter-based feedback and digital user engagement foster long-term water conservation efforts in the residential sector?*

To address this question, we conducted an observational study with case-control design and longitudinal measurements on 334 households in Valencia, Spain, by monitoring changes in water consumption over a three-year observation period. Differently from previous studies that investigated long-term changes in residential water and energy consumption in relation to the evolving socio-hydrologic and policy context^{18,19}, here we investigate behavior change in relation to water consumption feedback based on smart meter data that are communicated to customers via web/mobile media. At the beginning of the observational study (see “Methods” section), we provided a subset of the monitored households (treatment group) with access to the SmartH2O digital user awareness platform^{20,21} (see “Methods” section, Supplementary Figs. 1–7, and Supplementary Notes 1). Anonymized smart meter data of the treatment group gathered at different time frames during the observational study are compared with the pre-treatment water consumption data and with the behavior of a self-selected control group of other households. The control group had no access to either smart meter-based consumption feedback or the SmartH2O application for the whole duration of the longitudinal study.

This investigation on long-term residential water consumption behavior changes addresses our main research question from a three-fold angle. First, we inquire whether any water consumption change emerges from the study population considered as a whole, at different times in the observational study. To tackle this

¹Chair of Smart Water Networks, Technische Universität Berlin, Berlin, Germany. ²Einstein Center Digital Future, Berlin, Germany. ³Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milano, Italy. ⁴Global Omnium Idrica, Valencia, Spain. ⁵IACS-Institute for Applied Computer Science, University of Applied Sciences Stralsund, Stralsund, Germany. ⁶European Institute of Participatory Media, Berlin, Germany. ⁷IDISIA, USI-SUPSI, Lugano, Switzerland. ✉email: andrea.cominola@tu-berlin.de; andrea.castelletti@polimi.it

question, we assess average water consumption changes across all households in the two groups during the entire period of the observational study. Second, we investigate whether heterogeneous water consumption and behavior change patterns could be identified for different subgroups of the study population. Behavior change patterns of interest include durable/incremental conservation patterns and rebound effects. We thus run a segmentation analysis (see “Methods” section) and group the long-term behavior change patterns of different households in the treatment group in separate clusters. Finally, we discover which features of the behavior change program have most likely influenced the observed behavioral change. We correlate the identified household segments with the temporal smart meter sampling frequency and the level of usage of the digital application (see “Methods” section), to derive conclusions and draw recommendations for future digitally enabled water conservation and behavior change interventions.

RESULTS

Long-term behavior change by smart meter-based feedback and digital user engagement

At the beginning of 2016, we recruited volunteer households in Valencia to take part in our observational study (see details on the quantitative variables used in this study, population, study timeline, and exclusion criteria in the “Methods” section). All houses were previously equipped with smart water meters by the local water utility Global Omnium—EMIVASA, providing a pre-treatment baseline period. From June 2016 to February 2017 we provided the treatment group with access to the SmartH2O digital platform^{20,21}, a web and mobile application where they could visualize smart meter-based information and feedback about their water consumption and compare it with the water consumption of peer households (Supplementary Figs. 1–2), set individual water conservation targets (Supplementary Fig. 3), interactively learn and share water-saving tips (Supplementary Fig. 4), and engage with gamified tasks to earn points, badges, and rewards (Supplementary Figs. 5–6). Throughout the treatment period, we collected anonymized user activities on our platform (see “Methods” section). At the end of the treatment period, we continued to monitor water consumption till February 2019.

The overall data we collected allows quantifying the short-term change during the treatment period (June 1st, 2016–February 2nd, 2017), the medium-term change one year after the treatment (June 1st, 2017–February 2nd, 2018), and the long-term change nearly two years after the end of the treatment intervention (June 1st, 2018–February 2nd, 2019). In addition, a control group of self-selected households was used to comparatively quantify water consumption levels in households not provided with the SmartH2O digital platform or consumption-based feedback for the same observation and baseline periods. We quantify average water consumption changes for the treatment and control groups by first computing inter-annual differences of water consumption for each household and then calculating average changes across households in the two groups.

Our analysis of average water consumption changes reveals that water conservation behaviors were already visible for the treatment group during the first period of adoption of the SmartH2O platform (Fig. 1). In the short term, these households reduced their water consumption, each with respect to its pre-treatment baseline period by approximately 3.9% on average (Fig. 1a); 4.1% median value reported in Fig. 1b). The prominence of this behavioral change emerges when looking at the water consumption of the households in the control group: during the same period, they registered an average 19% increase (15.7% median value) relative to their baseline reference. The relative consumption changes in the treatment group correspond to an average water consumption

decrease of 14.8 L/day for each household. Saving 14.8 L/day would lead, in a month, to an avoided water consumption equivalent to the volume of water normally used for about 9 showers with a duration of 5 min (a flow of 9.5 L/min is considered in this calculation²²). Our longitudinal study demonstrates the potential for behavior change also in the medium and long term. In the medium term, households in the treatment group reduced their water consumption by approximately 5.9% on average (9.1% median value), each with respect to its baseline consumption level, with an overall 10.6% reduction of volumetric water in 2017–2018 (Fig. 1c). Average water conservation weakens in the long term, with households in the treatment group reducing their water consumption on average by 1.25% with respect to their baseline values (Fig. 1a). Yet, the comparison of these average values with the 9.5% long-term individual median water consumption reduction (Fig. 1b) and with the 8% overall long-term reduction of volumetric water used by the treatment group (Fig. 1c) suggests that only a subset of the households in the treatment group rebounded the water consumption behavior to pre-treatment levels. Substantial behavioral differences between the treatment group and the control group emerge consistently during the entire duration of the longitudinal study. The degree of difference and similarity between these two groups changes in different periods of our longitudinal study and the behavior changes observed for each group in the short term, medium term, and long term in comparison to baseline values (Fig. 1a–c) are also statistically significant in most of the cases (significance level of 5%, see Supplementary Notes 2 and Supplementary Tables 1–2 for more details). The conservation efforts of households in the treatment group resulted in overall volumetric water use reductions between 5.6% and 10.6% per year (Fig. 1c), while the control group increased its total volumetric water usage with respect to the baseline level by approximately 15.5% in 2016–2017 to 19.5% in 2017–2018.

Our results suggest that this trend for the control group can be explained by the observed increase in water consumption during the dry summer period, likely related to outdoor water usage. Notably, while all four years considered here registered average summer temperatures above the 1950–2018 average in July and August (Supplementary Fig. 8a–b), abnormally dry conditions were registered especially in the month of July in 2016–2018, with precipitation values well below the 1950–2018 average. In particular, no precipitation events were recorded in July 2016 and dry records with less than 5 mm cumulative monthly precipitation were registered also in July 2017–2018 (Supplementary Fig. 8c). In these conditions, the total volumetric water usage of the control group increased by 9.8% in July and August 2016, with respect to the baseline water consumption level in July–August 2015 (Fig. 1d). This is equivalent to an average additional consumption of 17.5 L/day per household with respect to baseline levels (corresponding to the water needed for 11 additional showers in a month, each with a duration of 5 min). Total volumetric water usage of the control group increased by over 14% in summer 2018, corresponding to 29.6 additional L/day per household (over 18 additional 5-min showers in a month). Conversely, households in the treatment group show a positive water conservation outcome across the whole longitudinal study. This effect might be primarily attributed to the initial household motivations and high engagement in our behavior change trial started at the beginning of summer 2016. However, water meter observations reveal that the treatment group exhibited effective water conservation efforts also in the summer of 2017 and 2018, with total volumetric water consumption change ranging between –5.4% and –5.6% with respect to baseline summer values.

The consistent behavioral differences between the treatment group and the uninformed control group illustrate that the short-term effects of smart meter-based feedback and digital user engagement persisted in the medium and long term and can thus foster durable water conservation efforts. Yet, the discrepancy

Water consumption change

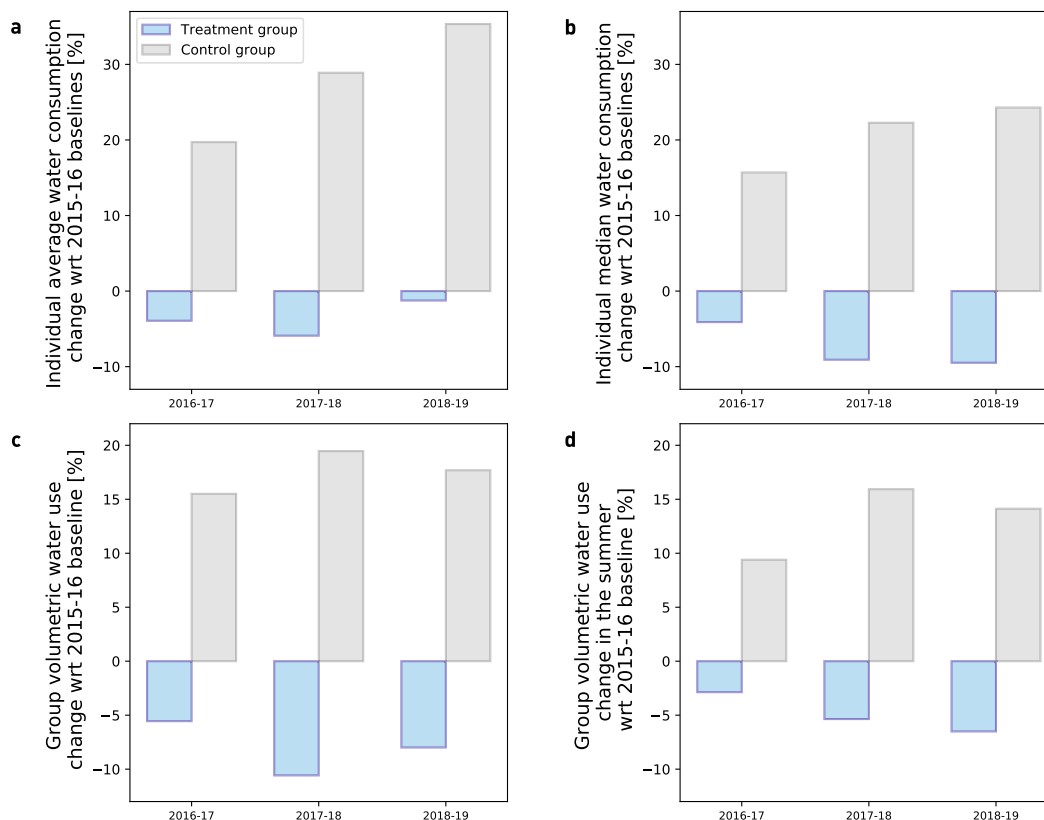


Fig. 1 Water consumption change. Water consumption changes are evaluated for the households in the treatment group (blue bars) and for the households in the control group (gray bars). Short-term, medium-term, and long-term water consumption changes are evaluated as averages (a) and medians (b) of percentage household water consumption changes with respect to individual baseline consumption levels, and total volumetric water use changes for the two groups over group-wide baseline consumption levels (c). Baseline consumption levels refer to pre-treatment average consumption levels observed in the period June 1st, 2015–April 30th, 2016. Short term refers to the treatment period (June 1st, 2016–February 2nd, 2017), medium term to a period one year after the treatment (June 1st, 2017–February 2nd, 2018), and long term to a period two years after the treatment (June 1st, 2018–February 2nd, 2019). Summer water consumption changes (d) are evaluated for the months of July and August in 2016, 2017, and 2018 as total volumetric water use changes for the two groups over group-wide baseline consumption levels (water consumption levels in July–August 2015).

between average and median values of water consumption changes in the treatment group, along with the scattered distribution of individual water consumption change values (Fig. 2), demonstrates highly heterogeneous individual responses to our stimuli. The key question is whether these heterogeneous behaviors reveal defined characteristics that can be used to design more effective behavioral programs^{23,24}.

Durable conservation behaviors and rebound effects

To analyze behavior change patterns and identify different long-term responses to our behavior change program, we clustered the households in the treatment group in eleven distinct segments, each representing a specific behavior change pattern (see “Methods” section and Supplementary Fig. 9). The five largest clusters expose the main heterogeneous behaviors (Fig. 3). Three clusters (Fig. 3a–c) reveal durable water conservation behaviors and comprise households that cumulatively represent nearly half (47%) of the treatment group members. Households in these three segments show clear short-term, medium-term, and/or long-term decreases in their average water consumption and can be thus considered water savers. Among the water savers, we can distinguish different patterns of behavior change: 20.6% of the households in the treatment group exhibit a steep decrease of average water consumption in the medium term (Fig. 3b), while

9.4% present more visible conservation results in the long term (Fig. 3c). Conversely, 17% of the households in the treatment group displayed a substantial water consumption reduction right after the beginning of the treatment period (Fig. 3a). Their behavior change appears primarily driven by a strong early-stage engagement, but the effects of the treatment intervention flatten in the long term.

Only approximately 10% of the households in the treatment group manifest a rebound effect^{5,25} in the long term (Fig. 3d), i.e., after an initial water conservation effect, their water consumption does not maintain a decreasing trend but goes back to pre-treatment values. Notably, the short-term and medium-term water consumption saving trend of this cluster is comparable to that of short-term water-saving households (Fig. 3a). Yet, their average water consumption rebounded almost to pre-treatment levels in the long term.

Finally, 9% of the households in the treatment group show an almost steady average water consumption in the short and medium term, which in some cases increases by nearly 40% in the long term (Fig. 3e). While a detailed interpretation of this behavior change pattern is not straightforward with the available information for this study, the steady water consumption level in the baseline and short-term and medium-term periods might suggest that these households might have had limited water-

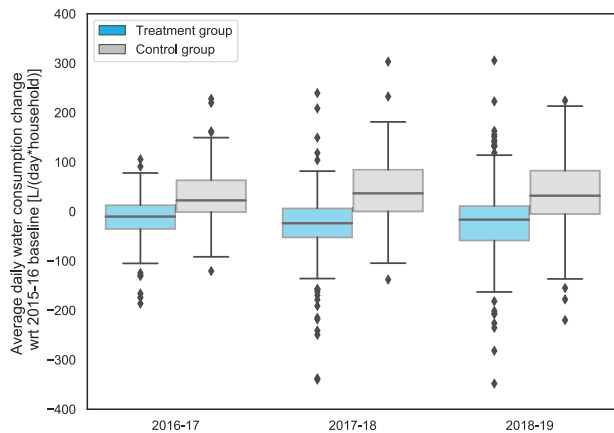


Fig. 2 Average daily household water consumption change distribution. Average daily water consumption changes in liters/day are evaluated for each household in the treatment group (blue boxes) and in the control group (gray boxes). Short-term, medium-term, and long-term water consumption changes are evaluated with respect to pre-treatment average consumption levels observed in the baseline period June 1st, 2015–April 30th, 2016. The line in the middle of each box represents the median value, boxes mark the first and third quartile. The length of box whiskers is equal to 1.5 times the interquartile range. Diamond markers are outliers.

saving opportunities. This might have discouraged their engagement, resulting in an undesired long-term increase in daily water consumption.

In addition to the behavior change patterns described above for the main household clusters, we identified six additional clusters, each of them including less than 7% of the households in the treatment group (Supplementary Fig. 9). The small size of these clusters and their generally oscillating water consumption patterns suggest that they include secondary behaviors justified by other drivers, i.e., households only marginally engaged with the SmartH2O platform and not aware or not willing to commit to water conservation.

Designing effective digitally enabled behavioral interventions

To further investigate the main factors that influenced the individual responses of different households, we correlate by logistic regression whether a household belongs to a specific behavior change pattern (or group of behavior change patterns) with descriptive data about the temporal sampling frequency of its smart meter and its level of engagement with the digital SmartH2O application, as described by the digital user engagement variables (i.e., the total number of logins, number of interactive actions in the application not associated with any reward, number of actions rewarded with gamification points, and cumulative reward importance, which accounts for the total amount of points, badges, and rewards).

The numerical results reveal meaningful associations (Fig. 4). The likelihood of households to be classified as water savers in the short, medium, and long term is positively correlated with their frequency of logins in the SmartH2O application and with the access to water consumption data at 1-h resolution (green line in Fig. 4). This empirical evidence suggests that durable water conservation behavior can be achieved by providing water consumers with detailed information on their water usage. Collecting high-resolution smart meter information and developing digital tools that enable users to visualize and interpret their water consumption in near real-time should thus be prioritized in the design of digitally enabled behavioral interventions. The cumulative reward importance is also positively weighted and, for a sub-set of water savers households, the number of actions

rewarded with gamification points, too. Advanced data collection and communication should be complemented with rewarded programs and non-monetary incentives to increase the retention of less environmentally minded or engaged households.

However, our results suggest that such incentives can be ineffective to achieve long-term conservation, when not coupled with sub-daily smart meter information. The access to hourly smart meter information emerges as an important factor that contributes to achieving long-term conservation effects (orange line in Fig. 4), along with intrinsic attitudes (non rewarded actions) and reward importance. Conversely, coarse information can produce undesired effects in the long term. Our results show that households are more likely to belong in the rebound segment when they receive smart meter information with a daily resolution, even if the coefficients for their non-rewarded actions (and, for some household, also rewarded actions) in the application suggest a good level of engagement (blue line in Fig. 4). This means that ICT-mediated engagement alone is not sufficient to motivate durable savings and must be backed by high-frequency consumption feedback.

Finally, the classification performance obtained for the logistic regression classifier (Supplementary Table 3) shows that we cannot classify the households in the different segments of the treatment group with high accuracy only by analyzing the temporal sampling frequency of their smart meters and their level of engagement with the digital SmartH2O platform. In many cases, the classification results are just slightly better than the expected performance of a random guess. Individual behaviors can be influenced by numerous physical and psychological determinants and the design of behavior change programs would benefit from detailed knowledge of such factors^{2,24,25}.

DISCUSSION

Long-term household water conservation behaviors are fostered by smart meter-based consumption feedback and digital user engagement. Most behavioral studies on water conservation document savings in household water consumption induced by feedback interventions in the short term, often followed by rebounding effects after the experimental trial as conservation awareness easily fades away^{4,5,17}. Here, our longitudinal study provides quantitative evidence that durable household water conservation behaviors can be observed in the presence of consumption feedback informed by smart meter data and user engagement mediated by a digital platform providing data visualization and interpretation, recommendations for water saving, and a gamification program. More than half (nearly 58%) of the households in the treatment group achieved substantial water savings in the short term. Almost half (47%) of all households in the treatment group also preserved durable water conservation behaviors two years after the start of the treatment.

We acknowledge that not all households engaged equally, some of them did not exhibit any predominant short-term or long-term behavior change, and a non-negligible fraction of them exhibited rebounding water consumption patterns. In addition, while average behavior changes are realistic, some extreme changes in water consumption might be due to variations in household and family composition or technological upgrades of indoor and outdoor water fixtures. Still, our observational study records a long-term 8% reduction in volumetric water consumption that is magnified by comparison with non-treated households, which conversely increased their total volumetric water usage by approximately 15.5% to 19.5% over the study period.

We observe more frequent conservation behaviors for the households that received smart meter information with hourly sampling frequency, rather than daily, suggesting that the availability of high-frequency consumption data appears to be a prerequisite for an effective digital engagement of water users. This is a crucial result

water savers

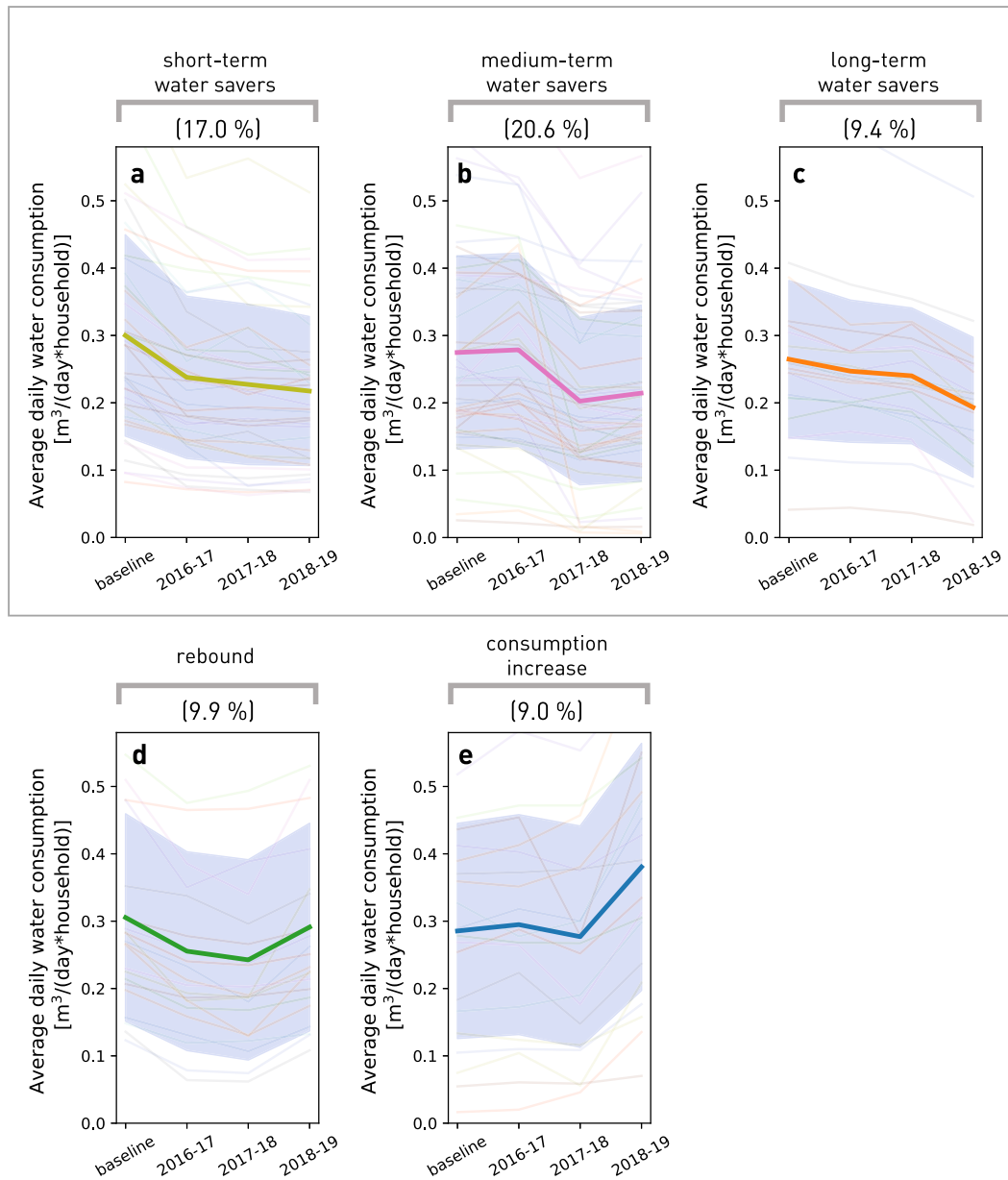


Fig. 3 Long-term water consumption change patterns. Five household segments (clusters) are represented as a selection of the eleven clusters obtained via household segmentation (Supplementary Fig. 9). Households in the treatment group are clustered in different segments based on the correlation distance of their water consumption pattern over the longitudinal study, where a water consumption pattern consists of the four values of average daily household water consumption during the 2015 baseline period and the three observation periods in 2016–2019. Each subplot represents a behavior change pattern, described by the labels on top. The percentage on top of each plot reports the percentage of households in the treatment group included in each household segment. Shaded blue areas are defined by the standard deviation around the mean (thick colored line in each subplot).

to substantiate the ongoing debate on the benefits and costs of different smart meter technologies and of other digital user engagement tools for water conservation, water utilities, and water consumers^{23,26,27}. We argue that consumption feedback based on fine-grained information can be more effective to help water consumers understand their consumption, identify opportunities for conservation, and monitor their conservation achievements.

From the findings of our study, we can derive some recommendations for the design of future water conservation and behavior change programs. Comprehensive data collection campaigns on water consumers' socio-demographic and

contextual information, and larger samples of participating households could help better characterize the most influential factors of a behavioral treatment program from other, non-controlled, effects at different temporal and spatial scales^{2,28}. It is suggested that possible biases are mitigated by performing future studies without volunteer selection bias¹⁴. Comparing the long-term effect of feedback-based interventions with other monetary and non-monetary demand management strategies in different social, economic, cultural, and geographical contexts, or in combination with other programs for sustainable water demand and supply and natural resources development^{29–32}, would help

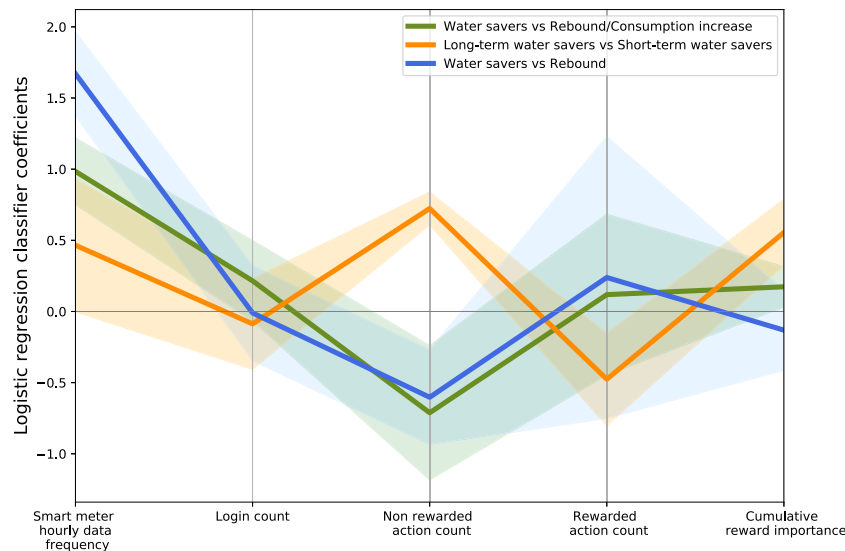


Fig. 4 Influence of smart meter data frequency and level of engagement with the digital SmartH2O application on household behavior change patterns. The coefficients of a logistic regression classifier weighting five normalized independent variables, i.e., smart meter hourly data frequency (binary variable for hourly vs. daily data), login count, non-rewarded action count, rewarded action count, and cumulative reward importance, are reported for three different tests. The first test (green line) classifies *water savers* against *rebound* or *consumption increase* households. The second test (orange line) classifies *long-term water savers* against *short-term water savers*. The third test (blue line) classifies *water savers* against *rebound* households. The different household clusters are characterized in Fig. 3. The shaded area around each line is defined by the standard deviation.

understand how scalable and general our results are³³. This study highlights the importance of individual engagement, smart meter technologies, and digital platforms as key elements to promote durable long-term water conservation behaviors.

METHODS

Quantitative variables

The intervention described here relies on the IT platform “SmartH2O” for the collection and visualization of smart meter data, the provision of consumption feedback to the user, the delivery of water-saving recommendations, and the engagement of the consumer through a gamification program^{20,21,34,35}. We embedded a gamification mechanism in the digital platform to maximize user retention and stimulate the exploration and sharing of content and the setting and achievement of personal saving goals. Via the gamification mechanisms, users could collect reward points for different actions performed in the digital platform or the achievement of water-saving targets. Reward points consisted of virtual points that the users could redeem for physical rewards. The design of the SmartH2O digital platform and the behavioral change stimuli that have been introduced in the Valencia case-control study (e.g., web and mobile app, different reward schemes), along with their individual elements and the corresponding illustrative screenshots of the platform are provided in the Supplementary Information, consistently with the information published in a previous study²¹. Other platforms similar to SmartH2O or approaches for water conservation based on digital technologies are reported in the literature, including, e.g., real-time water consumption feedback on in-home displays, interactive dashboards, and games^{36,37}. Yet, to the author’s knowledge, SmartH2O is the first platform of its kind whose effect is rigorously assessed in the medium term and long term.

Household-scale water consumption data and smart meter sampling frequency. Water consumption readings measured at the household scale constitute the main quantitative variable of interest used in this observational study to identify behavior changes. The SmartH2O digital platform relies on water consumption information stored in a central database and enables data communication from the water utility to the water consumers (see Supplementary Fig. 1 for its software architecture). Water consumption data are collected by smart meters installed at the household premises, according to a schedule that considers the maximum available frequency of data sampling at each installation (hourly or daily). The consumption data are

anonymized by the utility company, filtered, and transferred to the central database of the SmartH2O platform. The content of the central database is published to the user via a web portal and a mobile application, which are the entry points of all users’ interactions with the platform.

Besides the time series of water consumption, we also stored the sampling frequency allowed by each household-scale smart meter. Two types of sampling frequencies were available in the considered population, depending on the installed smart meter hardware: hourly or daily.

Digital user engagement variables. The central database of the SmartH2O platform comprises content for improving user awareness, such as water-saving recommendations, and for implementing the gamification program, such as the description of virtual and physical rewards. The interaction of the users with the platform and the overall user experience features several functionalities, including user login, water consumption and smart meter-based feedback visualization, conservation goal settings, and different gamified water conservation awareness actions (see also Supplementary Notes 1). We monitored the activity of each user in the SmartH2O platform for the entire duration of the treatment period and gathered quantitative data on these four *digital user engagement variables*:

- (i) *Login count*, defined as the total number of logins executed by each user.
- (ii) *Non-rewarded action count*, defined as the total number of actions performed by each user, with no reward points associated.
- (iii) *Rewarded action count*, defined as the total number of actions performed by each user, with associated reward points upon their completion.
- (iv) *Cumulative reward importance*, defined as the total amount of points achieved by each user by completing the rewarded actions. It accounts for the total amount of points, badges, and rewards achieved by an individual user in the SmartH2O platform.

Each user profile in the SmartH2O platform was associated with a unique smart meter ID, which allowed linking the user activity in the platform with the household water consumption data. User confidentiality was maintained throughout the full study as data were anonymized by the water utility managing the water meters and the central database.

Population and study size

Our observational study was conducted in the city of Valencia, Spain. With a population of 794,288 inhabitants, as reported in 2019 by the Spanish

National Institute of Statistics (Instituto Nacional de Estadística)³⁸, Valencia is the third-largest city in Spain. The water utility of Valencia (Global Omnium–EMIVASA) has installed more than 425,000 smart water meters since the early developments in 2006 to monitor the water consumption of nearly all the population³⁹ (the last official census data, recorded in 2011, report 419,994 households in total in Valencia⁴⁰). The total population considered in this study after application of the exclusion criteria described in the next section included 334 individual households, each equipped with a water meter.

The architecture of the smart metering infrastructure deployed in Valencia has been designed in order to be vendor-independent, so it allows for different smart metering solutions to be integrated³⁹. While this is clearly an advantage for procurement, the diversity of hardware has an impact on data sampling and only one of the available technologies supports hourly data collection, which is a preferred requirement for water consumption data quality assessment and provision of sub-daily water consumption information to households in our case-control study. The number of hourly reading meters in Valencia amounts to 168,172 as of July 12th, 2020. EMIVASA also offered its customers access to a web platform where bills and invoices could be managed and also information about the current (daily and monthly) water consumption data was made available.

During our observational study, we integrated the digital SmartH2O platform^{20,21} in the EMIVASA portal. We invited users who already had an account in the platform and a compatible meter reading frequency to voluntarily join our observational study and sign up to the SmartH2O platform. The recruitment campaign was performed using different media channels, namely, newspaper articles on consumer magazines, radio programs, banners on the digital and printed invoices sent to EMIVASA customers, and also a Facebook campaign targeting the Valencia area. At the end of the recruitment campaign, we received 525 applications out of which we obtained a treatment group composed of 223 households after application of the inclusion/exclusion criteria. Out of the households who did not apply to join the case-control study during the recruitment phase, 111 households agreed to be monitored as part of the self-selected control group to be considered as a benchmark group not subject to treatment, after active recruitment via phone by the EMIVASA call center (client service management). Households in the control group had only access to their water consumption data through the already existing platform, which did not offer any type of smart meter-based consumption feedback, behavioral stimuli, and/or gamification elements.

Informed consent was obtained from the households monitored in this study. Moreover, the water utility (Global Omnium–EMIVASA) supervised and approved the collection, usage, and processing of the anonymized quantitative variables above described in compliance with the EU General Data Protection Regulation 2016/679 and the pre-existing Spanish law 15/1999 LOPD of 1999 (the SmartH2O study started before the adoption of the GDPR in 2016).

Baseline and observation periods

The treatment period of the case-control study lasted 8.5 months, from June 2016 to February 2017. We also continuously collected anonymized water consumption data for the study population from June 2016 to February 2019 both to conduct the longitudinal study presented in this paper and evaluate water consumption changes over time in comparison with a pre-treatment baseline (June 1st, 2015–April 30th, 2016), as well as to compare water consumption changes in the treatment and control groups. Consistently with the months included in the treatment period (short-term behavior change), we identify the observation period June 1st, 2017–February 2nd, 2018 for medium-term behavior change assessment, and the observation period June 1st, 2018–February 2nd, 2019 for long-term behavior change.

Exclusion criteria

The population considered for analysis of water consumption changes in this observational study was obtained by sequential application of the following exclusion criteria.

1. *Exclusion of empty households.* First, we excluded the households with no data in the baseline and treatment period. We classified in this category also the households with a cumulative water consumption lower than 1.5 m³ over the whole baseline and treatment period (which together last nearly 20 months). This threshold value was identified as a conservative choice after consultation with the local water utility and comparison with the

average values of water consumption in the entire population (slightly above 0.21 m³/day) and the European average water consumption, which amounts to 128 liters per inhabitant per day (0.128 m³/day)⁴¹. A household in the considered population would use ~1.5 m³ in one week (0.21 m³/day × 7 days). While lower values than the average consumption are observed in those days in which the inhabitants spend little time at home, a cumulative consumption of 1.5 m³ over the course of more than 1 year can indicate that the house is generally empty (and possibly the observed water consumption is due to leaks).

2. *Exclusion of households with insufficient data length.* We removed the households with water consumption readings for less than 1000 h (approximately 6 weeks). This step guarantees a minimum representation of weekend/weekday water demand variation for more than 1 month (please note that the total duration of the treatment period is 8.5 months).
3. *Exclusion of partially empty households.* We excluded the households with more than 90% water consumption readings equal to zero in the baseline or observation period or completely lacking data for one of these two periods. We considered these households to be empty or equipped with faulty meters at least during one of the two short-term periods of interest. The above value threshold of 90% was identified with a trial-and-error procedure and expert-based data analysis that balance the rate of exclusion with the size of the remaining dataset.
4. *Exclusion of households lacking day-of-week representation.* We excluded the households with available observations for less than 7 unique day types, to guarantee a minimum representation of water consumption routines that depend on the day of the week. For those households with smart meters recording water consumption with hourly sampling frequency, we removed days with more than 4 h of gaps from the smart meter time series (anomalous meter data logging).
5. *Exclusion of households with anomalous high water consumption.* We considered hourly water consumption readings larger than 1 m³ as outliers (we thus removed these hourly readings) and we removed the households with a daily average water consumption larger than 1 m³ in at least one phase of the longitudinal study. High values of water consumption can be observed for specific days (e.g., when customers use water for outdoor irrigation or filling up a pool), yet average daily water consumption values over the selected threshold are more than three times higher than the European average (equivalent to approximately 0.3 m³/day per household). We did not apply more restrictive thresholds, in order not to bias our analysis and avoid unjustified exclusion of high water consumers.
6. *Exclusion of households with unrealistic short-term consumption change levels.* We excluded the households with extreme values of short-term consumption change during the treatment period, which were identified as outliers by Tukey's fences⁴². According to Tukey's fences, a data point x_i is considered an outlier if:

$$x_i \notin [Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)] \quad (1)$$

where Q_1 is the 25th empirical quartile (i.e., 25% of the data is lower than this point) and Q_3 is the 75th empirical quartile (i.e., 75% of the data is lower than this point), and $k = 1.5$. Tukey's fences with $k = 1.5$ approximate the 99.7% confidence interval defined for normal distributions by a distance of three standard deviations from the mean.

7. *Exclusion of households with anomalous conditions in medium-term and long-term.* We excluded 51 households that met the above exclusion criteria 1–6 during either the medium-term or long-term observation periods. Water consumption change patterns would be incomplete/anomalous for these households, with at least one missing/anomalous period out of the four periods of interest (i.e., baseline, treatment period, or following observation periods in 2018 and 2019).

With the above exclusion criteria, we obtained the 334 households considered for behavior change analysis in this observational study. More details on the population size after application of each exclusion criteria are reported with a flow diagram in Supplementary Fig. 10⁴³. It is worth noting that only less than 2% of high consumption households have been excluded, while most of the other excluded households had insufficient data or unrealistically low consumption levels. Also, the number of households in the sample considered here differ from those considered in

the evaluation of the SmartH2O project⁴⁴, due to the different temporal length of the two studies and the application of the exclusion criteria on data recorded in different periods (the SmartH2O project only included the baseline and treatment periods).

Adopting the same criteria to exclude households from the behavior change analysis only during the summer period (Fig. 1d) resulted in a reduced population of 179 households (101 households in the treatment group and 78 households in the control group), due to limited data availability for the summer period. Similarly, a subset of 198 households in the treatment group was considered for the correlation analysis by logistic regression (Fig. 4), as the excluded 25 households presented incomplete smart meter data or incomplete information on their usage of the digital SmartH2O application.

Data analysis and statistical methods

We performed customer segmentation to analyze heterogeneous long-term behavior change patterns (Fig. 3 and Supplementary Fig. 9). We applied agglomerative hierarchical clustering⁴⁵ to the patterns of average daily household water consumption during the entire duration of the longitudinal study. Here, a water consumption pattern of a household is a vector that contains four values of average daily water consumption, i.e., one for each period of the observational study, including the baseline (see “Methods” section—Baseline and observation periods). The only variable given as input to the hierarchical clustering algorithm consists of household-scale average water consumption per day for each phase of our observational study, which spans the baseline and the three observation periods in 2017, 2018, and 2019. Complete linkage and correlation distance were considered for hierarchical clustering. Complete linkage calculates the distance between two household clusters as the distance between the farthest pair of household water consumption patterns in the two household clusters, i.e., the maximum distance formulated as follows:⁴⁶

$$d(u, v) = \max(\text{dist}(u(x_i), v(z_i))) \quad (2)$$

where $d(u, v)$ is the distance between clusters u and v , x_i are the points belonging to cluster u and z_i those belonging to cluster v . Given two vectors of observations x_i and x_j , which in our study correspond to the water consumption patterns of two households (each with N elements, with $N = 4$, where each element is the household-scale average water consumption per day for the baseline and three observation periods) and their mean values (\bar{x}_i and \bar{x}_j) the correlation distance used by hierarchical clustering is calculated as follows:⁴⁷

$$\text{dist}(x_i, x_j) = 1 - \frac{(x_i - \bar{x}_i) \cdot (x_j - \bar{x}_j)}{\|x_i - \bar{x}_i\|_2 \|x_j - \bar{x}_j\|_2} \quad (3)$$

We considered hierarchical clustering as an appropriate choice because the analysis of the different hierarchical levels allowed the discovery of heterogeneous water consumption behaviors that would be potentially hidden if algorithms requiring a predefined number of clusters were used. We adopted complete linkage clustering to avoid that individual, mutually close households would force pairs of clusters representing different behaviors to merge. Also, we adopted correlation distance as we wanted to identify similarities in water consumption patterns over time, rather than in water consumption volumes.

After clustering the households in the treatment group with the above hierarchical clustering, similarly to a previous study¹⁸, we analyzed the coefficients of a logistic regression classifier cross-validated with binary tests to identify which candidate factors correlate with the main behavior change patterns that characterize the households in the treatment group (Fig. 4 and Supplementary Table 3). In this study, the input candidate factors consist of five independent variables that comprise the availability of smart meter hourly data frequency and the four digital user engagement variables, i.e., login count, non-rewarded action count, rewarded action count, and cumulative reward importance. First, we balanced the distribution of the households in the treatment group across the behavior change segments considered in the binary tests by Synthetic Minority Over-sampling Technique (SMOTE)⁴⁸. SMOTE oversamples the minority class to balance the sample distribution of a labeled dataset over the different classes. As we consider binary test where only two behavior change segments (or two groups of behavior change segments) are compared, the majority class represents the behavior change segment (or group of behavior change segments) with the highest number of samples and vice versa for the minority class. According to the SMOTE formulation⁴⁸, starting from a sample $c_{i,\text{initial}}$, which in this study is the

vector of input candidate factors for a household i in the minority class, a new sample $c_{i,\text{new}}$ is generated on the line between $c_{i,\text{initial}}$ and one of its k nearest-neighbors $c_{j,\text{initial}}$, with the following formula:

$$c_{i,\text{new}} = c_{i,\text{initial}} + \lambda(c_{j,\text{initial}} - c_{i,\text{initial}}) \quad (4)$$

where λ is a random number between 0 and 1, and $k = 5$ nearest neighbors computed based on Euclidean distance are considered by default⁴⁸. Among the possible options to perform class balancing, here we adopted a “not majority” strategy to over-sample the minority classes, i.e., we resample all classes but the majority class (which, in our binary problem, is equivalent to resampling the minority class).

Second, we trained a logistic regression classifier⁴⁹ with k -fold cross-validation ($k = 5$) and evaluated its performance via weighted F1 score. In our binary problem, the logistic regression classifier models the class membership probability $P(y_{i,p} = 1)$ for household i , where $y_{i,p} = 1$ indicates that the household belongs to behavior change pattern p (else $y_{i,p} = 0$, according to the following logistic function:

$$P(y_{i,p} = 1) = \frac{1}{1 + \exp^{-f(c_i)}} \quad (5)$$

where $f(c_i)$ is a linear function where the input variables c_i are weighted by corresponding coefficients α :

$$f(c_i) = \alpha_0 + \alpha_1 c_{i,1} + \alpha_2 c_{i,2} + \dots + \alpha_M c_{i,M} + \varepsilon_i \quad (6)$$

In this study, $M = 5$, $c_{i,1}$ is a binary variable representing the availability of smart meter with hourly data frequency, $c_{i,(2,3,4,5)}$ are the four digital user engagement variables defined above, α_0 is the intercept of the logistic regression, and ε_i is random noise. We normalized the variables before logistic regression classification by subtracting the mean and dividing by the standard deviation to rescale them to comparable value ranges. The analysis of their corresponding logistic regression coefficients reveals how these variables discriminate among different clusters of water consumers and, thus, how they are potential determinants of defined water consumption behaviors. The F1 score (FS) is first calculated for each behavior change pattern (or group of patterns) p as the harmonic mean of the precision and recall achieved by the logistic regression classifier, formulated as follows:

$$FS_p = 2 \times \frac{(\text{precision}_p \times \text{recall}_p)}{(\text{precision}_p + \text{recall}_p)} \quad (7)$$

$$\text{Precision}_p = \frac{TP_p}{TP_p + FP_p} \quad (8)$$

$$\text{Recall}_p = \frac{TP_p}{TP_p + FN_p} \quad (9)$$

where, given positive and negative classes, TP_p , FP_p , and FN_p are the number of true positive elements (the classifier correctly predicts the positive class for them), false-positive elements (the classifier incorrectly predicts the positive class), and false negative elements (the classifier incorrectly predicts the negative class). A weighted average of the FS_p is then computed to account for class imbalance:

$$FS_{\text{average}} = \frac{1}{H} \sum_{p \in P} |p| \times FS_p \quad (10)$$

where P is the total number of classes p and H is the total number of elements aggregated across all classes.

Software implementation

We coded the exclusion criteria in Matlab and used the “prctile” function for the calculation of the quantiles in Tukey’s fences (last Matlab version tested: R2020b)⁵⁰. We implemented the customer segmentation analysis and logistic regression classifier in Python (version 3.7.1): the customer segmentation analysis relies on the hierarchical clustering included in the SciPy library⁵¹; the logistic regression classifier, along with its k -fold cross-validation and performance evaluation, were implemented using the machine learning library Scikit-learn⁵²; SMOTE oversampling was implemented using the Imbalanced-learn toolbox⁵³. A notebook with the Python code used to generate the results reported in this article is available in a public GitHub repository⁵⁴.

DATA AVAILABILITY

The anonymized smart meter data used in this study are from Global Omnium-EMIVASA and were collected during and after the SmartH2O project (<http://smarth2o.deib.polimi.it>). The full dataset that supports the findings of this study is available from Global Omnium-EMIVASA but restrictions apply to the availability of these data, which were used under license for this study, and so are not entirely publicly available, as they contain sensitive household-level information. The full dataset is however available from the authors upon reasonable request and with permission of Global Omnium-EMIVASA. Part of the data collected during the SmartH2O project is available on Zenodo (<https://doi.org/10.5281/zenodo.556725>)⁵⁵.

CODE AVAILABILITY

A notebook with the code used to generate the results reported in this article is available in this public GitHub repository: https://github.com/acominola/SmartH2O_longexperimentaltrial.git⁵⁴.

Received: 14 August 2020; Accepted: 7 April 2021;

Published online: 07 May 2021

REFERENCES

- Beal, C. D., Gurung, T. R. & Stewart, R. A. Demand-side management for supply-side efficiency: modeling tailored strategies for reducing peak residential water demand. *Sustain. Prod. Consum.* **6**, 1–11 (2016).
- Quesnel, K. J. & Ajami, N. K. Large landscape urban irrigation: a data-driven approach to evaluate conservation behavior. *Water Resour. Res.* **55**, 771–786 (2019).
- Castilla-Rho, J. C., Rojas, R., Andersen, M. S., Holley, C. & Mariethoz, G. Social tipping points in global groundwater management. *Nat. Hum. Behav.* **1**, 640–649 (2017).
- Sønderlund, A. L., Smith, J. R., Hutton, C. J., Kapelan, Z. & Savic, D. Effectiveness of smart meter-based consumption feedback in curbing household water use: knowns and unknowns. *J. Water Resour. Plan. Manag.* **142**, 04016060 (2016).
- Fielding, K. S. et al. An experimental test of voluntary strategies to promote urban water demand management. *J. Environ. Manag.* **114**, 343–351 (2013).
- Marzano, R. et al. Determinants of the price response to residential water tariffs: meta-analysis and beyond. *Environ. Model. Softw.* **101**, 236–248 (2018).
- Rougé, C. et al. Assessment of smart-meter-enabled dynamic pricing at utility and river basin scale. *J. Water Resour. Plan. Manag.* **144**, 1–12 (2018).
- Barrett, G. Water conservation: the role of price and regulation in residential water consumption. *Econ. Papers* **23**, 271–285 (2004).
- Darby, S. *The Effectiveness of Feedback on Energy Consumption - A Review for DEFRA of the Literature on Metering Billing and Direct Displays* (Environmental Change Institute, University of Oxford, Oxford, 2006).
- Ehrhardt-Martinez, A. K. & Donnelly, K. A. *Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities*. (Washington, DC: American Council for an Energy-Efficient Economy, 2010).
- Kelly, J. & Knottenbelt, W. Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature. *Preprint at arXiv* 1605.00962 (2016).
- Tiefenbeck, V. Bring behaviour into the digital transformation. *Nat. Energy* **2**, 1–3 (2017).
- Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E. & Galinsky, A. D. The critical role of second-order normative beliefs in predicting energy conservation. *Nat. Hum. Behav.* **2**, 757–764 (2018).
- Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E. & Staake, T. Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nat. Energy* **4**, 35–41 (2018).
- Tiefenbeck, V. et al. Overcoming salience bias: how real-time feedback fosters resource conservation. *Manag. Sci.* **64**, 1458–1476 (2018).
- Stewart, R. A. et al. Integrated intelligent water-energy metering systems and informatics: envisioning a digital multi-utility service provider. *Environ. Model. Softw.* **105**, 94–117 (2018).
- Cominola, A., Giuliani, M., Piga, D., Castelletti, A. & Rizzoli, A. E. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: a review. *Environ. Model. Softw.* **72**, 198–214 (2015).
- Bolorinos, J., Ajami, N. K. & Rajagopal, R. Consumption change detection for urban planning: monitoring and segmenting water customers during drought. *Water Resour. Res.* **56**, e2019WR025812 (2020).
- Bolorinos, J., Rajagopal, R. & Ajami, N. K. Mining the gap in long-term residential water and electricity conservation. *Environ. Res. Lett.* **16**, 024007 (2020).
- Rizzoli, A. E., Castelletti, A., Fraternali, P. & Novak, J. Demo Abstract: SmartH2O, demonstrating the impact of gamification technologies for saving water. *Comput. Sci. Res. Dev.* **33**, 275–276 (2018).
- Novak, J. et al. Integrating behavioural change and gamified incentive modelling for stimulating water saving. *Environ. Model. Softw.* **102**, 120–137 (2018).
- Mayer, P. W. et al. Residential end uses of water. *Aquacraft, Inc. Water Eng. Manag.* **310**, 4309b (1999).
- Cominola, A. et al. Data mining to uncover heterogeneous water use behaviors from smart meter data. *Water Resour. Res.* **55**, 9315–9333 (2019).
- Cominola, A. et al. Segmentation analysis of residential water-electricity demand for customized demand-side management programs. *J. Clean. Prod.* **172**, 1607–1619 (2018).
- Gonzales, P. & Ajami, N. Social and structural patterns of drought-related water conservation and rebound. *Water Resour. Res.* **53**, 10619–10634 (2017).
- Monks, I., Stewart, R. A., Sahin, O. & Keller, R. Revealing unreported benefits of digital water metering: Literature review and expert opinions. *Water* **11**, 1–32 (2019).
- Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D. E. & Abdallah, A. M. Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. *Environ. Model. Softw.* **102**, 199–212 (2018).
- Maier, S. U., Raja Beharelle, A., Polanía, R., Ruff, C. C. & Hare, T. A. Dissociable mechanisms govern when and how strongly reward attributes affect decisions. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-020-0893-y>. (2020).
- Nakagawa, M. Trust in sustainable natural resource development. *Nat. Hum. Behav.* **3**, 542 (2019).
- Janssen, M. A. Modelling social norms of water conservation. *Nat. Hum. Behav.* **1**, 624–625 (2017).
- Daniel, D., Marks, S. J., Pande, S. & Rietveld, L. Socio-environmental drivers of sustainable adoption of household water treatment in developing countries. *npj Clean. Water* **1**, 1–6 (2018).
- Dorea, C. C. et al. Safely managed drinking water services in the Democratic People's Republic of Korea: findings from the 2017 Multiple Indicator Cluster Survey. *npj Clean Water* **3**, 1–7 (2020).
- Kennedy, D. S., Goemans, C., Klein, R., Lowrey, J. & Reidy, K. Residential water demand management: Lessons from Aurora, Colorado. *J. Am. Water Resour. Assoc.* **44**, 192–207 (2008).
- Castelletti, A. et al. Gamified approaches for water management systems: an overview. *Smart Water Grids* 169–201 <https://doi.org/10.1201/b21948-7>. (2019).
- Albertarelli, S., Fraternali, P., Novak, J., Rizzoli, A. E. & Rottondi, C. DROP and FUNERGY: Two gamified learning projects for water and energy conservation. In *Proceedings of the 11th European Conference on Games Based Learning, ECGBL 2017* 935–938 (2017).
- Clifford, E. et al. Interactive water services: the WATERNOMICS approach. *Procedia Eng.* **89**, 1058–1065 (2014).
- Terlet, J., Beach, T. H. & Rezzgui, Y. Smart meters and in-home displays to encourage water conservation through behavioural change. *Int. J. Inf. Commun. Eng.* **10**, 537–543 (2016).
- Instituto Nacional de Estadística (INE). Valencia/València: Población por municipios y sexo. <https://www.ine.es/jaxiT3/Datos.htm?t=2903#tabs-tabla>. (2019).
- Bertocchi, M. et al. The SmartH2O project. *Validation Report*. https://smarth2o.deib.polimi.it/wp-content/uploads/2017/03/sh2o_D7.2_SES_WP7_validation_report_v1.1.pdf (2016).
- Instituto Nacional de Estadística (INE). Censo 2011. Número de viviendas por tipo Viviendas por municipios (con más de 2.000 habitantes) y tipo de vivienda. <https://www.ine.es/jaxiT3/Datos.htm?t=3456#tabs-tabla>. (2011).
- The European Federation of National Water Services. *Europe's Water in Figures*. 22 (2017).
- Tukey, J. W. *Exploratory Data Analysis*. (Addison-Wesley Publishing Company, 1977).
- Moher, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, 336–341 (2009).
- Rizzoli, A. E. et al. The SmartH2O Project. *Final Overall Validation and Impact Report*. (2017).
- Brunton, S. L., & Kutz, J. N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. (Cambridge University Press, 2019).
- Müllner, D. Modern hierarchical, agglomerative clustering algorithms. *Preprint at arXiv* 1109.2378 (2011).
- Shirkhorshidi, A. S., Aghabozorgi, S. & Ying Wah, T. A Comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS ONE* **10**, 1–20 (2015).
- Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

49. Dreiseitl, S. & Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inform.* **35**, 352–359 (2002).
50. The MathWorks, I. MATLAB. <https://www.mathworks.com/products/matlab.html>. (2020).
51. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
52. Pedregosa, F. et al. Scikit-learn: machine learning in Python Fabian. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
53. Lemaitre, G., Nogueira, F., West, W. S., Mv, O. & Aridas, C. K. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **18**, 559–563 (2017).
54. Cominola, A. Long-term water conservation is fostered by smart meter-based feedback and digital user engagement [Code]. https://github.com/acominola/SmartH2O_longexperimentaltrial.git. (2020).
55. The Smarth2O project. The-Smarth2O-project/datasets: anonymized datasets of Smarth2O project (Version v1.0.1). <https://doi.org/10.5281/zenodo.556725>. (2017).

ACKNOWLEDGEMENTS

The research leading to these results has received partial funding from the European Union's Seventh Framework Programme (FP7/ 2007e2013) under grant agreement no. 619172 (Smarth2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption). The authors would like to acknowledge Global Omnium–EMIVASA for sharing the anonymized data used in this research and all partners of the Smarth2O consortium for their contributions to the design and implementation of the Smarth2O digital platform. The authors acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Technical University of Berlin for covering publication costs.

AUTHOR CONTRIBUTIONS

A.C. (Andrea Cominola), A.C. (Andrea Castelletti), M.G., P.F., and A.E.R. designed the research and developed the paper. A.C. (Andrea Cominola) and M.G. analyzed the data. J.C.G.H. provided the smart meter data. S.L.H.G. provided user data from the digital Smarth2O application. A.C. (Andrea Cominola) wrote the software code for data analysis, performed numerical analysis, and designed the visual elements of the paper. A.E.R., A.C. (Andrea Castelletti), P.F., J.N., J.C.G.H. secured funding during the

duration of the Smarth2O project (2014–2017) and led the implementation and validation of the Smarth2O digital platform. All authors reviewed the manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41545-021-00119-0>.

Correspondence and requests for materials should be addressed to A.C. or A.C.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021