

Article

Gamified Crowdsensing for Non-Intrusive Energy Disaggregation in Office Buildings

Thanasis G. Papaioannou ^{1,*}, , Dimosthenis Kotsopoulos ^{2,†},  and Cleopatra Bardaki ³

¹ Department of Digital Industry Technologies, National and Kapodistrian University of Athens, 34400 Psachna, Greece

² Department of Management Science and Technology, Athens University of Economics and Business, 10434 Athens, Greece; dkotsopoulos@aueb.gr

³ Department of Informatics and Telematics, Harokopio University, 17676 Athens, Greece; cleobar@hua.gr

* Correspondence: atpapaioannou@uoa.gr; Tel.: +30-222-802-1870

† These authors contributed equally to this work.

Abstract

User activity significantly contributes to energy wastage in buildings, yet traditional load analysis relies on costly, intrusive, and privacy-invasive metering infrastructure. This paper proposes gamified crowdsensing as an alternative approach to energy disaggregation, where building occupants act as “human sensors” to report energy-consumption activities. We develop a challenge-based mobile application utilizing a risk-based scoring mechanism (Impact $\times$ Likelihood) to prioritize the reporting of high-intensity loads and overcome user resistance. To ensure data integrity, user claims are cross-checked via a sparse, low-cost verification infrastructure including NFC stickers, BLE beacons, and minimal smart metering. Pilot experiments involving 80 employees across office buildings in Greece and Spain demonstrate that user claims for PC and lighting usage alone can explain 65% of the variation in aggregate device energy consumption and 62% of the variance in total building energy consumption. Our findings reveal that while human-centric models are highly effective, they are sensitive to local office culture and require site-specific calibration. This approach provides a scalable, low-intrusive, and cost-effective alter-native to traditional energy disaggregation for understanding occupant-driven loads.

Keywords: energy disaggregation; smart grid; crowdsensing; gamification

1. Introduction

The EU building stock remains largely old and energy-inefficient; up to 85% of EU buildings were constructed before 2001, and among those, approximately 75% exhibit poor energy performance [1,2]. Across the EU, the building sector is responsible for 40% to 42% of aggregate energy usage and contributes to over a third of all greenhouse gas emissions linked to energy [2,3]. Despite the urgency, the annual renovation rate for buildings is only around 1%, with deep energy renovations occurring at a mere 0.2–0.3% per year [4]. Recent policies, such as the revised Energy Performance of Buildings Directive (EPBD) [5] further increase this urgency, as they set stricter targets for the decarbonization of the building stock and the acceleration of deep renovation rates across the EU. Acting on building energy efficiency is therefore critical for reducing energy bills, enhancing energy independence, and achieving a fully decarbonised building stock by 2050. Public buildings consume approximately 40% more energy than residential buildings [6], with public offices representing 23% of the non-residential sector and being responsible for 26%



Academic Editor: Xiaohu Yang

Received: 22 June 2026

Revised: 23 July 2026

Accepted: 24 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of its energy consumption. User activity is a major driver of this consumption [7], is often wasteful, and is difficult to mitigate. This is consistently reported in existing research, with recent evidence identifying occupant behaviour as one of the most significant sources of uncertainty in building energy performance [8]. Occupants in public office buildings lack direct financial incentives to save energy; they neither pay utility bills nor can they easily spare time from their primary tasks for energy conservation. Even when intrinsic motivations for sustainable behaviour exist (e.g., altruism, energy awareness, personal norms), people frequently lack the knowledge and ability to act effectively [9].

Efforts to enhance energy efficiency within office environments frequently employ gamification or demand-response (DR) strategies to encourage energy-conserving behaviours in public facilities [10–14]. Broadly speaking, gamification aims to leverage users' intrinsic motivations, boosting their engagement to fulfil system goals via interactive, game-like experiences. Nevertheless, tracking occupants' energy usage for these gamified or DR methods typically necessitates an extensive network of meters and sensors [10,11,13]. This heavy reliance on instrumentation presents two major issues: first, ethnographic research indicates that building occupants often resist dense sensor deployments due to privacy apprehensions [15]. Second, deploying such invasive hardware is often prohibitively expensive and economically unviable, particularly in aging infrastructure (noting that 85% of EU buildings date back to before 2000, with 75% of those demonstrating subpar energy efficiency [2]).

In parallel, the concept of mobile crowdsensing (MCS) has matured as a data collection model where individuals equipped with sensing and computing devices collectively gather and share information about phenomena of common interest [16]. Originally applied to environmental monitoring, traffic sensing, and urban management, MCS has since expanded into a broader paradigm termed heterogeneous crowdsensing, or Crowdsensing 2.0, that integrates human participants, IoT devices, environmental sensors, and AI-driven analytics [17]. Building energy monitoring represents a natural application domain for this paradigm: occupants are already present, they directly influence energy consumption through their daily activities, and their behavioural data carries semantic richness that automated sensors alone cannot capture.

This paper proposes an alternative, human-centred energy disaggregation mechanism in buildings. The energy consumption per appliance by the buildings' occupants is estimated with the support of crowdsensing where we can see how the actual users perceive their energy usage and savings. Specifically, we demonstrate how a mobile gamified application is used as an incentive mechanism for motivating users to enhance their energy consumption and, simultaneously, participate to mobile crowdsensing of their energy usage in their buildings. It is a team-played gamified application [15] that proposes energy saving challenges to users daily. Every challenge is formulated as a specific rule, with a proprietary rule engine [18] evaluating user actions to ensure compliance. Although players are organized into teams that compete against one another for prestige, point allocation is based strictly on individual participation. Points are distributed to any player who actively accepts a challenge and successfully adheres to its guidelines without violation. Apart from individual challenges, there exist team ones, the accomplishment of which necessitates coordinated action by a team of users regarding the use of a shared device. The specific number of points per challenge is selected in accordance with the user ability to perform the respective energy-saving behaviour and the associated energy consumption to the challenge. Successfully completing a challenge earns the user points, a process validated through a specialized instrumentation strategy. This verification relies on a sparse network of energy meters and sensors within the building, combined with the participants' smartphone sensors, to explicitly track the specific user behaviours required

for each challenge. We have employed our gamified crowdsensing application to office buildings in two different workplaces in two European countries, namely at DAEM, the IT Company of the Municipality of Athens in Greece and at ICAEN Catalan Energy Agency in Spain. The game challenges chosen concern the energy consumption of lights and PCs in offices. The results from the office buildings show that user claims can explain 65% of the variation in the aggregate energy consumption of PCs and lights, and 62% of the variation in the total energy consumption in the building.

Non-intrusive load monitoring (NILM) approaches [7] provide a more cost-efficient software alternative to dense instrumentation, yet their average estimation accuracy deteriorates with higher numbers and highly homogeneous devices [19,20]. In commercial and office environments, where multiple identical devices operate simultaneously, NILM performance drops significantly because load signatures become highly homogeneous and difficult to distinguish [19,20]. Furthermore, traditional dense sensor networks face prohibitive retrofitting costs and severe privacy concerns from occupants [15]. This gap between laboratory accuracy, scalability, and real-world field performance reveals a fundamental limitation in current smart grid research.

To address this fundamental limitation, this study proposes a *paradigm shift*: moving the task of energy disaggregation from the electrical signal layer to the behavioural layer. Instead of attempting to computationally untangle identical electrical signatures from a single smart meter, we bypass the algorithmic bottleneck entirely by utilizing building occupants as “human sensors”. By deploying a gamified crowdsensing application, we “disaggregate” energy at the source, i.e., the user, providing a highly scalable, privacy-friendly, and cost-effective alternative to hardware-intensive sub-metering.

Specifically, the novel contributions of this paper are fourfold:

- Human-in-the-Loop Disaggregation: We introduce and validate a crowdsensing methodology that uses self-reported, gamified user claims as a reliable proxy for high-granularity appliance load estimation.
- Risk-Based Incentive Design: We formulate a unique scoring mechanism based on risk management principles ($\text{Impact} \times \text{Likelihood}$) to strategically incentivize the reporting of high-yield energy actions, ensuring data quality without overburdening users.
- Sparse Verification Architecture: We demonstrate that a highly economical IoT setup (NFC tags and BLE beacons) is sufficient to verify user claims and maintain system integrity, drastically reducing the capital expenditure required for traditional building retrofits.
- Contextual Validation via Mixed-Effects Modelling: We rigorously evaluate our framework across different organizational sites, utilizing Linear Mixed-Effects Modelling (LMM) to mathematically prove that human-centric energy models are inherently bound to local office cultures, highlighting the necessity for localized baseline calibration in behavioural energy models.

Figure 1 provides a comparison between our proposed human-centric disaggregation method and traditional energy disaggregation approaches in order to better illustrate its conceptual difference.

The remainder of this paper is organized as follows: In Section 2, we overview the related work in different research areas. In Section 3, we describe the proposed incentive mechanism and the risk-based scoring system. In Section 4, we present the hardware and software infrastructure used for verification, followed by the evaluation of the pilot studies in Greece and Spain. Section 5 discusses the findings and, finally, Section 6 concludes the paper.

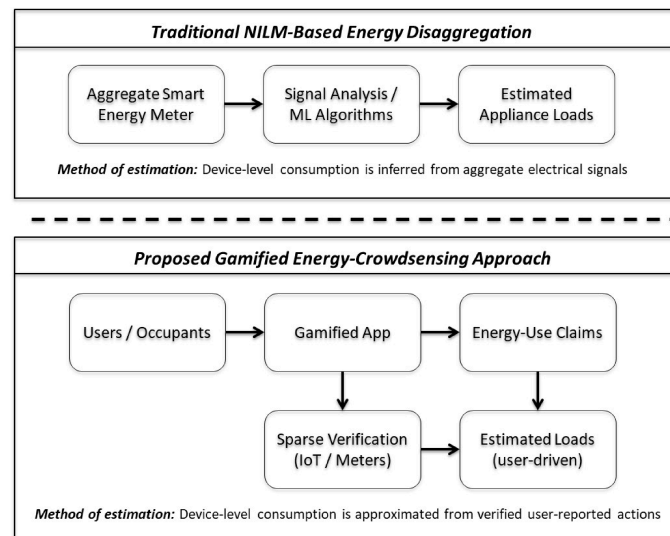


Figure 1. Conceptual comparison between traditional NILM-based disaggregation and the proposed gamified crowdsensing approach.

2. Related Work

Often referred to as “the Holy Grail of energy efficiency”, energy disaggregation provides the ability to gather data on the discrete power usage of individual appliances [21,22]. Non-Intrusive Load Monitoring (NILM) enables deducing this appliance-level energy consumption by analysing the overall aggregate power load of a building [7,23,24]. Users’ feedback has been utilized in the past in order to increase the precision of disaggregation algorithms in Non-Intrusive Appliance Load Monitoring (NIALM) [25–27]. Self-reported pro-environmental behaviour is a convenient and cost-effective means of assessing objective pro-environmental behaviour, with a strong association between them recorded in a variety of studies [28]. This holds true specifically for energy use; an in-depth comparison between tracked residential electricity loads and energy-focused time-use diaries revealed that self-reported behaviours serve as fairly accurate predictors of actual power demand [29]. In addition, occupancy-aided energy disaggregation, which focuses on extrapolating energy-use, based on occupancy in buildings, has been described and attempted repeatedly in the past with various results, mainly in household environments [30–33]. Moreover, it has emerged as a key paradigm for improving building energy consumption estimation and prediction, and this amplifies the need to explicitly incorporate human behaviour into energy models [33,34].

Apart from statistical approaches, crowdsensing energy consumption is another interesting direction. The idea of employing humans as sensors is not new and it has been already employed in various contexts. Apart from the wide utilization of automated sensors for the recording of natural phenomena, humans can themselves also directly act as sensors for data collection, utilizing their own natural means of perception [35]. The intelligence and mobility of humans can be leveraged in order to collect high quality and/or semantically complex data that usually requires the use of sophisticated (and costly) hardware and software [16]. Participatory sensing entails a conscious choice by the “human sensors” to opt to meet the application requests with regards to when, where, what, and how to sense [36]. Mobile crowdsensing (MCS) functions as an essential foundation for new Internet of Things (IoT) deployments [37] and facilitates a wide variety of large-scale systems [36]. This process takes place when users equipped with sensor-enabled computing devices, such as smartphones, collaboratively gather, share, and process data to evaluate and chart subjects of mutual interest [16]. Consequently, MCS has attracted considerable interest recently, emerging as an attractive data-gathering model that depends on the active information

contributions of mobile device users [38]. Sparse mobile crowdsensing represents a specific variation of Mobile Crowdsensing (MCS) that involves monitoring only a select, limited fraction of a target area. Following this initial data collection, extrapolation techniques are applied to accurately deduce the information for the remaining, unmeasured regions [39]. Moreover, extant research has also explored hybrid approaches that combine sensing infrastructures with occupants' behavioural inputs. The results from such studies have included the improved performance of smart building energy management systems [40]. Unlike most typical commodities, energy is an "abstract, invisible, and untouchable" resource that frequently goes unobserved due to its intangible nature [41]. Consequently, consumers routinely misjudge their electricity usage, typically overestimating the power drawn by low-energy appliances while underestimating the consumption of high-energy devices [42]. Hence, the direct assessment of energy consumption and savings by its users, based on their own perception, has been deemed "problematic and largely inaccurate" [43]. However, crowdsensing has been employed in the past with great success, in order to record building occupants' perceived levels of comfort with regards to ambient conditions [44,45], thus also leading to energy savings due to the optimal operation of HVAC systems [44]. Therefore, evidently crowdsensing may be successfully utilized to record users' preferences and actual actions with regards to energy consumption (instead of measurements of energy amount expended), thus providing a solid means for building an energy disaggregation mechanism. However, recent studies highlight persistent limitations of NILM approaches in real-world environments, especially in organizational or office building contexts, where homogeneous loads and complex usage patterns exist [19]. Moreover, emerging online NILM approaches which aim to improve real-time performance, seem to still face challenges with regards to issues of scalability and generalization [46]. Accordingly, to enhance NILM model performance in this context, human-in-the-loop learning techniques incorporating user feedback, reducing labelling effort and improving generalization across domains have been proposed [47].

Current crowdsensing platforms frequently lack effective incentive structures [48]. Motivation and rewards play a critical role in Mobile Crowdsensing (MCS) because they are essential for driving user participation and data contribution [37]. As a result, for MCS applications to be successful, they must implement proper incentive strategies to attract, involve, and sustain participants, thereby ensuring the completion of high-quality sensing tasks [18,49]. One major type of incentive is entertainment, which involves converting these sensing tasks into interactive, playable games [50]. Furthermore, social incentive models capitalize on participants' interpersonal connections to foster widespread cooperation; since a user's reward is heavily influenced by the actions of their friends, participants are driven to leverage their social relationships to sway their peers' behaviour and secure a greater payoff [49]. Lastly, in a crowdsourcer-centric setup, the crowdsourcer issues a unified reward that is subsequently divided among the active contributors [48]. Therefore, utilizing games where the incentives are provided by the crowdsourcer and shared among the group of participants, is an effective way of increasing the users' motivation and engagement for participating in a crowdsensing effort. Moreover, the effectiveness of gamification in promoting energy-saving behaviour in buildings, particularly when combined with feedback, social comparison, and goal-setting mechanisms, has been confirmed through recent systematic reviews [51]. However, the effectiveness of gamification on a long-term basis depends on sustained user engagement, as well as the appropriate design of the incorporated incentive mechanisms [52].

Based on all of the above, in this paper, we decided to design and utilize a team-played game in order to motivate the employees in office buildings to timely and accurately report on their energy consumption actions, with the aim of crowdsensing the energy

consumption in such environments, without the need for the utilization of costly energy disaggregation hardware. In addition, regarding incentivisation of users for crowdsensing, *scoring systems* can have considerable influence on the participation and satisfaction of players in crowdsensing games. They can act as a positive feedback system and provide a mechanism whereby game players are rewarded with points whenever they accomplish a task [53]. Hence, scores are often also a good proxy for the intention of players to achieve the objectives in a game. Furthermore, as highlighted by [54], the act of tracking scores is a crucial component that contributes to the overall enjoyment of the gaming experience. Therefore, improving the design of scoring systems can lead to increased satisfaction and participation in a game. If the points awarded satisfy the players, the latter are more likely to continue actively participating in the game and pursuing its goals, while they are guided towards performing specific actions through the scoring mechanism [53].

According to [55], a gamified system's primary driving force is its specific reward structure and mechanism, with point-based rewards being particularly effective at encouraging users to engage with or practice using a software platform. Furthermore, [53] identifies three fundamental components of effective scoring systems: Perceivability (the degree to which players are conscious of the scoring system's existence), Controllability (how much influence players have over their own scores), and Relation to Achievement (where higher potential achievements correspond to a greater level of challenge). Drawing from these principles, we conclude that for a gamified crowdsensing application to successfully engage its audience, its scoring framework must be highly perceivable, controllable, and directly correlated to the successful completion of the specific crowdsensing tasks at hand.

3. The Incentive Mechanism Design

To determine user requirements, we applied the Rapid User Centered Design (Rapid UCD) methodology, conducting ethnographic studies, online surveys, and on-site visits at two specific office buildings: the Development Company of Municipality of Athens (DAEM) in Greece and the Energy Directorate of Catalonia (ICAEN) in Spain. As detailed in [56,57], our research revealed several prevalent energy-wasting behaviours within these pilot locations. These actions include:

- Keeping windows open while the air conditioning is running;
- Leaving lights illuminated in unoccupied areas;
- Failing to power down computers and monitors at the end of the workday;
- Choosing to use elevators rather than the stairs;
- Leaving shared appliances, such as printers, photocopiers, and coffee machines, turned on after work hours.

While aiming overall to track such energy-consumption behaviours, we had to:

- associate energy-consumption with individuals or groups of users in a cost-effective and low-intrusive manner;
- track as much as possible energy consumption based on user activity.

In order to motivate employees to act as sensors of energy consumption and crowd-sense energy consumption at their workplace, we provided them with incentives. To this end, we took into account their preferences for an energy-saving competition at work, as well as conducted on-site visits to deduce the availability of specific actions for energy saving and their specific characteristics in the workplaces. These incentive mechanisms are described below.

3.1. Energy Consumption Claims—The Challenges

Note that within this workplace environment, many energy-consuming activities are integral to daily job duties and are therefore inelastic in nature. Consequently, our primary focus is on modifying employee habits that result in unnecessary energy waste. To achieve this, we implemented a gamified strategy in which employees participate in individual, challenge-based activities via a mobile application. For every distinct energy-wasting behaviour, one or more specific game challenge has been defined to encourage its avoidance. Saving electricity consumed by a device may necessitate individual or collaborative action by many users. For example, turning off the PC depends on individual action, while turning off the lights in a particular shared office space takes collective action to accomplish. To this end, we define individual and team challenges for individual or collective action, respectively. When an employee accepts and successfully completes a challenge, she/he gets awarded a number of points in an appropriate way, which is defined in the next subsection. Furthermore, employees are organized into teams that compete against one another to accumulate these points.

Challenges in the game may span different time periods, e.g., daily, weekly or even monthly, and, based on their individual or group nature, they can be categorized as:

- Personal challenges: These are initiated by single players, and their successful completion relies entirely on the individual's own actions. The outcome is reflected directly in that specific player's progress within the game.
- Team challenges: Although an individual participant can accept these, successful completion is dictated by the combined actions of every member within that participant's team. As a result, achieving these challenges requires a certain level of collaboration among teammates, and the outcome impacts the in-game progress of both the individual and the entire team.

Moreover, based on their objective, challenges can be categorized as:

- Volume-shifting challenges: They aim to alleviate energy-wasting actions altogether and thus reduce the total energy load within a workday.
- Time-shifting: They aim to move the electricity load associated to unavoidable tasks either to off-peak hours, thus smoothening the load, or to other convenient time periods, where there is abundance of energy, lower electricity price, etc.

Indicative volume-shifting personal challenges are the following:

- "Use the stairs, instead of the lift";
- "Turn PC off when away from desk for more than 30 min".

In addition, indicative volume-shifting team challenges are the following:

- "Turn-off lights upon leaving room": This challenge is accomplished when the last person leaving the room turns the lights off;
- "Keep windows closed, while A/C is on";
- "Turn off PC/printer/coffee machine upon leave".

In addition, if a (micro-) solar generation system is considered at a workplace (without a net-metering agreement with the utility company, as it is the case in the Greek pilot site), then it is desirable that the electricity load of unavoidable tasks is shifted towards time periods within the workday where the electricity generation is high. Such individual/team challenges can indicatively be among the following ones:

- "Shift meeting time";
- "Shift coffee/lunch break";
- "Shift phone charging";
- "Shift printing".

3.2. Maximize Impact—Scoring for Challenges

To design an effective incentive system, we treat the lack of user reporting as a ‘risk’ to the accuracy of our energy-load analysis. Each energy-wasting behaviour represents a risk event. We then employ a risk-prioritization methodology to assign point scores, where the ‘impact’ of a risk is the associated energy consumption, and the ‘probability’ is the likelihood of the user not performing the desired behaviour, derived from our baseline surveys. In this subsection, we present our score calculation approach in order to provide appropriate incentives to the employees towards reporting their energy-consumption-related actions at work. First of all, for fairness and energy-saving prioritization, individual (and team) scores are assigned according to the performance of employees to the various energy-reporting challenges. The employee is presented with several challenges in the game with each of them associated with a particular energy-reporting action. The employee may undertake any of these challenges, and upon their accomplishment receive a fixed amount of points. The maximum score that can be received by an employee is the sum of the points of all challenges with which they are presented. In these terms, the performance Δp_i of an employee i can be expressed as follows:

$$\Delta p_i = \text{Score for Accomplished Challenges by Employee } i / \text{Total Score of All Challenges Presented to Employee } i$$

The points associated with either individual or team challenges contribute to the team score of the team where each employee belongs. Therefore, the performance of team j with member set J is given by:

$$\Delta p_j = (\text{Score for Accomplished Challenges by Employee } i \in J + \text{Score of Accomplished Challenges for team } j) / (\text{Total Score of All Challenges Presented to Employee } i \in J + \text{Total Score of All Team Challenges for } j)$$

In order to calculate the scores for the accomplishment of each energy-reporting action, we consider employees not complying with the recommended reporting by the game as a potential *risk*. Moreover, each particular challenge pertaining to a particular energy-reporting behaviour has a different *impact* in terms of actual energy consumption to be reported. The specific scores assigned to each challenge have to be appropriately calculated, so as to motivate employees to accomplish the challenges that have greater impact on energy consumption. To this end, we decided to draw from the project management literature and employ risk management methodology in order to calculate the points associated with each particular challenge, so as to contain respective risks of not reporting a specific action on the accuracy of the total reported energy consumption.

Risk impact assessment involves evaluating the likelihood and the potential consequences of risk events should they actually materialize [58–60]. Having recorded our energy consumption sources, which depict our risks in this case (if they are not accounted for), we proceeded to assess their respective impacts at each site, based on the actual energy consumption figures recorded for each source of energy consumption therein. Accordingly, the impact of each risk (energy-consumption behaviour) was assessed, together with the probability (chance) that each risk event (not being accounted for) will occur. We initially defined the probability that the various energy-reporting behaviours associated with the game challenges are followed at the two pilot sites, by employing a self-reported questionnaire that assessed the adherence of each of the energy-consumption actions by employees (see Tables 1 and 2 for the specific items employed in the survey). All the items were rated on a Likert scale between 1 = Completely Disagree and 7 = Completely Agree.

Table 1. Self-reported energy consumption behaviour of employees in DAEM.

DAEM	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
When I am finished using my computer for the day, I turn it off.	42	6.3333	1.16172	0.17926	5.9713	6.6954	2.00	7.00
When I leave a room that is unoccupied, I turn off the lights.	42	6.4286	1.17167	0.18079	6.0635	6.7937	2.00	7.00
When I leave a bathroom that is unoccupied, I turn off the lights.	41	6.3171	1.40426	0.21931	5.8738	6.7603	1.00	7.00
When I am not using my computer, I turn off the monitor.	42	4.9762	1.88043	0.29016	4.3902	5.5622	1.00	7.00
When I leave my work area, I turn off the Air Conditioner(s).	42	6.6190	0.58236	0.08986	6.4376	6.8005	5.00	7.00
When I leave my work area, I turn off the printer(s).	42	4.8333	1.93702	0.29889	4.2297	5.4370	1.00	7.00
(Reversed) I often leave the windows open while the Air Conditioner is on.	40	4.8250	1.85206	0.29284	4.2327	5.4173	1.00	7.00
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	42	6.02	1.278	0.197	5.63	6.42	2	7

Table 2. Self-reported energy consumption behaviour of employees in ICAEN.

ICAEN	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval		Minimum	Maximum
					Upper Bound	Lower Bound		
When I am finished using my computer for the day, I turn it off.	44	6.9545	0.21071	0.03177	6.8905	7.0186	6.00	7.00
When I leave a room that is unoccupied, I turn off the lights.	44	6.7727	0.56501	0.08518	6.6009	6.9445	4.00	7.00
When I leave a bathroom that is unoccupied, I turn off the lights.	44	5.9773	1.67733	0.25287	5.4673	6.4872	1.00	7.00
When I am not using my computer, I turn off the monitor.	44	4.7273	2.12804	0.32081	4.0803	5.3743	1.00	7.00
When I leave my work area, I turn off the Air Conditioner(s).	44	2.8182	2.04908	0.30891	2.1952	3.4412	1.00	7.00
When I leave my work area, I turn off the printer(s).	44	2.4091	1.71611	0.25871	1.8873	2.9308	1.00	7.00
(Reversed) I often leave the windows open while the Air Conditioner is on.	44	5.7727	1.72349	0.25983	5.2487	6.2967	1.00	7.00
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	44	3.98	2.205	0.332	3.31	4.65	1	7

Based on the identified risk events (energy consumption actions), their impact in terms of energy consumption (to be reported), and the probability of occurrence for each risk (the consumption behaviour remaining unaccounted for), we proceeded to derive scores for each of the energy consumption actions that need to be reported by the participants in the game. A very similar approach was followed by [61] in order to prioritize behavioural interventions. Their practical example application was regarding water conservation, albeit they propose the usage of their methodology in other behavioural interventions. Based on their research, behaviours are evaluated according to their potential impact, the probability of their adoption, and the current levels of participation among the target audience. Every one of these traits is assigned a score, making it possible to properly prioritize and provide incentives for these behaviours.

To calculate the scores for the energy-consumption behaviours that we monitored in the game, we followed the following steps:

- The actual energy consumption of the devices was employed, as higher energy-consuming devices can have a larger impact on total energy consumption (e.g., the A/C units where applicable). We took into account that, due to the large differences in energy consumption between different device types (e.g., a lamp consumes 60 Watts, while an Air Conditioner may consume 1000 Watts), the impact scores may need to be adjusted (e.g., similarly to a log-scale), so as to not de-motivate users from adopting the less impactful energy consumption behaviours at all. Given the orders-of-magnitude difference in energy consumption between devices (e.g., a light vs. an A/C unit), raw impact scores would disproportionately reward a few high-impact actions. To create a balanced and motivating points system that encourages a wide range of energy-saving behaviours, we applied a logarithmic transformation to the impact values, effectively compressing the scale. Therefore, the weighted impact of an action $a \in A$, where A are the available energy-saving actions, is calculated as:

$$\text{Weighted Impact}(a) = \frac{\log_{10}(\text{Energy Impact}(a))}{\max_{y \in A} \log_{10}(\text{Energy Impact}(y))},$$

where the energy impact, i.e., energy-consumption reduction, by an energy-saving action a related to a particular device is given by:

$$\text{Energy Impact}(a) = (\text{hours device will be turned off if action } a \text{ is taken}) \times (\text{Watts that are consumed / saved per hour by device})$$

- As per the calculation of the likelihood that an energy consumption behaviour occurs by the participants, we followed the following approach: First, we calculated the mean values for the self-reported adherence in energy-consumption behaviours, as they had already been recorded in the aforementioned questionnaire survey. We then used the mean values in each pilot site as indicators of the likelihood that the participants enact a behaviour.
- To calculate the score that should be attributed to each energy reporting action a , we calculated the score for the respective energy-saving challenge as:

$$\text{Score}(a) = \text{Weighted Impact}(a) \times \text{Likelihood}(a)$$

Therefore, the higher the weighted impact of an energy-saving action reported and the higher the probability that employees carry this behaviour, the higher the score (if they never perform the action, we do not need to record it, while if they do it all the time, e.g., turning off the lights when exiting a room, then the action needs to be recorded).

4. Metering Infrastructure

Our energy monitoring setup relies on a combination of smart plugs and N -channel smart meters, which can measure multiple electrical circuits simultaneously. We use the N -channel meters to track non-plug loads, such as elevators, air conditioning, and lighting, as well as the combined plug loads for clusters of up to four employee desks. Meanwhile, smart plugs are dedicated to monitoring shared office equipment, including coffee machines, photocopiers, and printers.

To link specific energy usage to individual building occupants, we deployed a supplementary sensor network. This includes Estimote Bluetooth Low Energy (BLE) beacons installed in elevators and office spaces to detect user presence. The system estimates a user's proximity based on the signal strength received by their smartphone, which then transmits this data to the Data/Core Backend System. Additionally, we attached NFC stickers to individual devices, light switches, and halfway up the stairwells. The game prompts users to swipe their smartphones against these tags to report energy-saving behaviours

(like taking the stairs instead of the elevator) and earn reward points for the energy they conserve. To monitor HVAC efficiency, Fibaro switch sensors are installed on all windows to detect if they are left open while the A/C is running. We also placed a Fibaro 4-in-1 sensor in each designated area to track motion/presence, luminosity, humidity, and temperature. This helps pinpoint energy-saving opportunities, map specific actions to individuals, and ensure ambient comfort. All structural mappings, linking circuits, smart plugs, and NFC IDs to specific devices, as well as associating window switches, BLE beacons, and 4-in-1 sensors to their respective rooms and the overall building layout, are securely maintained in the Data/Core Backend.

Driving the gamified application is a Data/Core Backend built upon the open-source SiteWhere IoT middleware (sitewhere.org). This architecture features five primary components: (a) A core system that manages and stores all incoming sensor and energy data. (b) A gateway responsible for collecting, filtering, and routing this data to the backend. (c) An analytics module that identifies energy-saving opportunities and actions, subsequently feeding them into the game. (d) The gamification system [15], which tracks game progression, issues user-specific challenges, and processes energy-saving events. (e) A proprietary rule engine [18] that evaluates whether a user has successfully completed a challenge and allocates the appropriate points. The overall system architecture is illustrated in Figure 2.

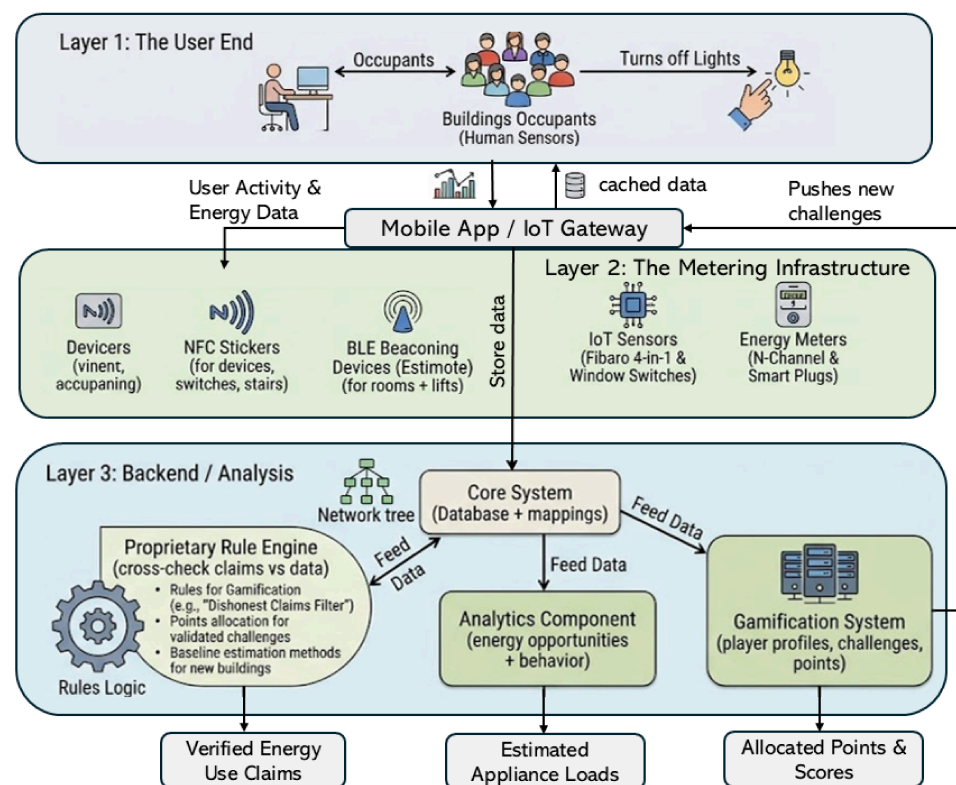


Figure 2. The overall system architecture.

Note that the aforementioned sparse IoT infrastructure, although it is employed in an office environment, is privacy-friendly. While BLE beaconing devices allow for the indoor positioning of employees, such user tracking has not been employed by our system. In fact, BLE beacons and/or 4-in-1 sensors are only employed to verify that the employee, identified by a user ID, has left a particular space within a timeout of 20 min after claiming a relevant challenge, e.g., lights OFF after leaving room. Therefore, no privacy or security threats emerge for the game participants. Moreover, our IoT infrastructure allows for fair point allocation through the verification of claims of employees regarding energy-saving

actions with 100% accuracy, i.e., TP is 100%. FN may be >0%, in general, if employees accidentally have their Bluetooth interfaces in their mobile phones turned off, the 4-in-1 sensor for detecting occupancy is blocked by an obstacle and the energy change associated to the challenge (measured by the N-channel smart meter) is insignificant (possibly also due to multiple concurrent loads in the same circuit). In that case, which did not occur in the pilot studies, the system does not assign any points to users for their respective challenges claimed and reminds them which sensors have to be ON for the verification of their claims. When there are connectivity failures, user claims are cached on the device and synchronized with the backend once connectivity returns, so that no data is lost. If a phone is fully discharged, the sensing infrastructure (smart meters, window and 4-in-1 sensors) keeps recording actual energy and presence, and only the users' claims are delayed until the next swipe and sync. Finally, while device tampering is a potential threat in any IoT infrastructure, the cross-verification of measurements regarding user position from multiple BLE beaconing devices and 4-in-1 sensors, when needed, minimizes this risk in our case.

5. Evaluation

5.1. Application Context: Two Workplaces in Europe

As already mentioned, we employed the crowdsensing mechanism at two office buildings in Greece and Spain, namely DAEM and ICAEN respectively. The employees utilized the gamified app daily and completed challenges concerning saving energy while using their PCs and lights in the workplace. Specifically, the gamified app was a team-played gamified application that proposes daily energy saving challenges, i.e., PC off after work and lights off after leaving office, to users. We had installed in the two workplaces IoT devices to monitor the actual energy consumption per device. Every game challenge functioned as a specific rule, with a proprietary rule engine [18] verifying user compliance based on their actions. Players were grouped into teams competing for prestige, but only individuals actively participating in a challenge were eligible to earn points. Ultimately, points were granted to players who accepted a challenge and successfully adhered to its rules, as confirmed by data collected from the metering infrastructure. Apart from individual challenges (i.e., PC off), there existed team ones (i.e., lights off), the accomplishment of which necessitated coordinated action by a team of users regarding the use of a shared device and the respective points are awarded equally to all team members.

On the one hand, DAEM employees were given challenges to switch off their PC when they leave their office, as well as switch off lights. A total of 51 employees of DAEM utilized the gamified app, split into seven teams, with each team being located at a separate work room. The pilot experiments lasted for 5 months. Each of the participants completed on average 31 “switch off PC” and 29 “turn off lights” challenges (i.e., 1567 and 1477 PC and Lights challenges, respectively, in total for all participants).

In terms of IoT infrastructure that monitored energy consumption, we installed smart metering infrastructure (three smart meters) in the panel of electrical circuits of DAEM to measure the energy consumption in an office area of 650 m². We utilized the meters along with other equipment: smart plugs, window-door sensors and NFCs to acquire an objective view of energy consumption.

On the other hand, 29 ICAEN participated in the pilot experiments. The employees were located in one open space, and they were given “switch off PC” and “turn off lights” challenges for a period of 6 months. The participants were split into six teams depending on their actual location at the workplace. Each of the participants completed on the average 20 “switch off PC” and four “turn off lights” challenges (i.e., 750 and 134 PC and Lights challenges, respectively, in total for all participants), during the pilot period.

5.2. Calculation of Challenge Scores

Prior to the pilot experiments, we surveyed the energy consumption behaviours of employees at DAEM and ICAEN using a 7-point Likert scale, i.e., Absolutely Disagree (1), Disagree (2), Somewhat Disagree (3), Neither Agree nor Disagree (4), Somewhat Agree (5), Agree (6), Absolutely Agree (7). The statistical analysis of the self-reported energy consumption behaviours of the employees at DAEM and ICAEN are illustrated in Table 1 and Table 2, respectively.

We first subtract 1 from the Likert scores, so that “Totally Disagree” corresponds to 0. Then, we calculate the likelihood of each self-reported behaviour that corresponds to the action dictated by a specific challenge as the mean value reported over the maximum possible value, i.e., 6.

The assumptions made in the calculation of the scores for each one of the energy-reporting challenges in these two pilot sites can be reviewed in Table 3 below:

Table 3. Estimation of impact for each of the energy consumption behaviours.

CHALLENGES	Hours	Load (W)	Impact
Challenge 1 (PC off when going home)	16	40	640
Challenge 2 (PC off when away from computer for more than 30 min)	8	40	320
Challenge 3 (Using the stairs instead of the Lift)	1	2 (2 Wh per floor for 1 s of use)	2
Challenge 4 (if A/C on then close window)	8		3200
Challenge 5 (lights and A/C in room turned off overnight)	12	1000	12,000
ADDITIONAL BEHAVIOURS			
Printer off after hours	16	60	960
Photocopier off after hours	16	170	2720
Fax machine off after hours	16	30	480
Coffee machine off after hours	16	40	640
BEHAVIOURS AS RATED IN QUESTIONNAIRE			
When I am finished using my computer for the day, I turn it off.	16	40	640
When I leave a room that is unoccupied, I turn off the lights.	8	200	1600
When I leave a bathroom that is unoccupied, I turn off the lights.	8	100	800
When I am not using my computer, I turn off the monitor.	8	30	240
When I leave my work area, I turn off the Air Conditioner(s).	16	800	12,800
When I leave my work area, I turn off the printer(s).	16	60	960
(Reversed) I often leave the windows open while the Air Conditioner is on.	8	400	3200
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	16	40	640

Moreover, we calculated the energy-consumption impact, i.e., a proxy of the energy savings, of each challenge. We then calculate the weighted energy impact per action, so that to decrease the difference among devices that inherently consume highly-different amounts of energy. The reward points for an accomplished action a are calculated by:

$$\text{Score}(a) = 100 \cdot \frac{\text{Weighed Impact}(a) \cdot \text{Likelihood}(a)}{\max_{y \in A} (\text{Weighed Impact}(\text{action } y) \cdot \text{Likelihood}(\text{action } y))}$$

The final score is rounded to the closest 5-point increment, so that users understand it better. The scores per energy-saving action are presented in Table 4 for DAEM and in Table 5 for ICAEN, while the in-game challenge scores employed in both pilot sites are depicted in Table 6.

Table 4. Impact–likelihood calculations per site for the targeted energy-consumption actions at DAEM.

Behaviour	Impact	Normalisation Factor	Weighed Impact	Likelihood	Weighed Impact × Likelihood	Score Per Completed Action	Rounded Score
When I am finished using my computer for the day, I turn it off.	640	2.81	0.68	88.89%	0.61	64.85	65
When I leave a room that is unoccupied, I turn off the lights.	1600	3.20	0.78	90.48%	0.71	75.37	75
When I leave a bathroom that is unoccupied, I turn off the lights.	800	2.90	0.71	88.62%	0.63	66.88	65
When I am not using my computer, I turn off the monitor.	240	2.38	0.58	66.27%	0.38	41.01	40
When I leave my work area, I turn off the Air Conditioner(s).	12,800	4.11	1.00	93.65%	0.94	100.00	100
When I leave my work area, I turn off the printer(s).	960	2.98	0.73	63.89%	0.46	49.54	50
(Reversed) I often leave the windows open while the Air Conditioner is on.	3200	3.51	0.85	63.75%	0.54	58.09	60
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	640	2.81	0.68	83.73%	0.57	61.09	60
I use the stairs (assumption that I use the lift 6 times per day for 1 floor)	12	1.08	0.26	80.00%	0.21	22.45	20

Table 5. Impact–likelihood calculations per site for the targeted energy-consumption actions at ICAEN.

Behaviour	Impact	Normalisation Factor	Weighed Impact	Likelihood	Impact × Likelihood	Multiplying Factor	Score Per Completed Action	Rounded Score
When I am finished using my computer for the day, I turn it off.	640	2.81	0.80	99.24%	0.79	125.71	99.88	100
When I am not using my computer, I turn off the monitor.	240	2.38	0.68	62.12%	0.42	125.71	53.03	55
When I leave my work area, I turn off the printer(s).	960	2.98	0.85	23.48%	0.20	125.71	25.12	25
(Reversed) I often leave the windows open while the Air Conditioner is on.	3200	3.51	1.00	79.55%	0.80	125.71	100.00	100
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	640	2.81	0.80	49.62%	0.40	125.71	49.94	50
I use the stairs (assumption that I use the lift 6 times per day for 2 floors)	24	1.38	0.39	80.00%	0.32	125.71	39.60	40

Table 6. Game score for different energy-saving behaviours at the pilot sites.

DAEM	
Behaviour	Rounded Score
When I am finished using my computer for the day, I turn it off.	65.00
When I leave a room that is unoccupied, I turn off the lights.	75.00
When I leave a bathroom that is unoccupied, I turn off the lights.	65.00
When I am not using my computer, I turn off the monitor.	40.00
When I leave my work area, I turn off the Air Conditioner(s).	100.00
When I leave my work area, I turn off the printer(s).	50.00
(Reversed) I often leave the windows open while the Air Conditioner is on.	60.00
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	60.00
I use the stairs (assumption that I use the lift 6 times per day for 1 floor)	20.00
ICAEN	
Behaviour	Rounded Score
When I am finished using my computer for the day, I turn it off.	100.00
When I am not using my computer, I turn off the monitor.	55.00
When I leave my work area, I turn off the printer(s).	25.00
(Reversed) I often leave the windows open while the Air Conditioner is on.	100.00
When I am the last to take coffee in the afternoon at work, I turn the coffee machine off.	50.00
I use the stairs (assumption that I use the lift 6 times per day for 2 floors)	40.00

5.3. Data Analysis Methodology

We conducted all statistical evaluations of the gathered data using IBM SPSS Statistics v.23. To compare the measured variables and investigate the broader characteristics of our sample, we utilized descriptive statistics, specifically the mean and standard deviation (SD). Additionally, we executed linear regression analyses via the enter method to evaluate how effectively the data from one or more predictor variables could forecast the values of a separate outcome variable. All statistical models were estimated using standard linear regression, while the reported R^2 values reflect explanatory power rather than predictive generalization. Prior to modelling, Variance Inflation Factor (VIF) checks were conducted to ensure the absence of multicollinearity among predictors, and the normality of residuals was verified. We ran our serious game for 21 weeks between 10 December 2018 and 30 April 2019 and collected weekly game data, namely game points, number of accomplished challenges per different challenge, lights energy consumption, PC energy consumption, energy consumption per team and total energy consumption. While the data was collected during the 2018–2019 pilot phase, it remains highly relevant for evaluating crowdsensing feasibility today, as occupant PC and lighting behavioral patterns in office environments exhibit significant structural inertia over time. A total of six teams at DAEM and six at ICAEN were selected to participate in the game. We removed data concerning the Administration team at DAEM and the corresponding Management team at ICAEN due to their low participation in the game. In addition, we removed weeks 1, 5–8 and 21 at DAEM and the first 5 weeks at ICAEN due to system stability issues in these periods. After this data cleaning, we ended up with 15 weeks of data for DAEM for five teams and with 16 weeks of data for ICAEN for five teams, i.e., 155 data cases in total. Moreover, due to connectivity issues in the gateways in DAEM and ICAEN that were aggregating total energy consumption, the total energy consumption was found to be lower than the sum of the energy consumption of lights and PCs in 42 faulty cases. We resolved this problem in the following two ways: (a) we omitted all faulty cases keeping only 112 cases in the dataset, referring to that as the “clean dataset”, and (b) we corrected the faulty cases by estimating the total energy consumption based on its linear-regression relation with the sum of the PC and lights energy consumption in the rest of the cases, i.e., having in total 155 cases, referring to that as the “corrected dataset”.

To ensure the robustness of our findings and explicitly account for the hierarchical and temporal structures inherent in the dataset, we supplemented the standard linear regression with a Linear Mixed-Effects Model (LMM). Because our data consists of repeated weekly measures nested within distinct organizational teams, the assumptions of ordinary least squares (OLS) regression regarding the strict independence of observations may be violated. The LMM was configured to treat the total measured energy consumption as the dependent variable. The accomplished *PC* and *Lights challenges*, along with *Week*, were modelled as fixed effects to evaluate the human-centric signal and check for temporal autocorrelation. Concurrently, *Team* was modelled as a random intercept. This methodological addition allowed us to mathematically isolate the variance attributable to specific local office cultures (the team clusters) from the variance explained by the gamified user claims, establishing a more rigorous proof-of-concept for human-in-the-loop disaggregation. The LMM was formally defined as $Y_{it} = \beta_0 + \beta_1[Chal1_PC_{it}] + \beta_2[Chal5_Lights_{it}] + \beta_3[Week_t] + u_i + \epsilon_{it}$, where Y_{it} denotes the total measured energy consumption for team i at week t , β_{1-3} represent the fixed effects of the challenges and time, u_i represents the random intercept to account for unobserved, team-specific baseline characteristics, and ϵ_{it} represents the residual error.

5.4. Results of Pilot Studies

User engagement was high overall and sustained throughout the pilot period. As shown in Table 7 (Greek pilot) and Table 8 (Spanish pilot), the monthly challenge completion rate remained high, averaging over 70%, indicating that the gamified mechanism successfully maintained participant involvement. The in-game performances of different teams at DAEM and at ICAEN are depicted in Tables 9 and 10 respectively.

Table 7. Undertaken vs. fulfilled challenges per month in the Greek pilot.

Month	Completed	Failed	Overall (%)
December 2018	2026	1181	3207 (63.17%)
January 2019	3910	1568	5478 (71.37%)
February 2019	1128	331	1459 (77.31%)
March 2019	1217	437	1654 (73.57%)
April 2019	1268	584	1852 (68.4%)

Table 8. Undertaken vs. fulfilled challenges per month in the Spanish pilot.

Month	Completed	Failed	Overall (%)
December 2018	879	134	1013 (86.77%)
January 2019	535	174	709 (75.4%)
February 2019	633	231	864 (73.26%)
March 2019	719	226	945 (76.08%)
April 2019	876	268	1144 (76.57%)

Table 9. The in-game performance of different teams at DAEM. Important findings are highlighted in bold.

Team Name	Estimated Consumption (KWh)	Estimated Savings (%)	Total Score	Avg. 'Switch-off PC' Challenges Per User	Avg. 'Switch-off Lights' Challenges Per User	PC Energy Savings (%)	Lighting Energy Savings (%)
Administration	1225.61	102.93 KWh (8.39%)	1446.00	6.80	3.8	2.27 KWh (0.4%)	39.51 KWh (7.7%)
EU Projects	1881.82	256.25 KWh (13.6%)	8560.29	46.43	46.29	125.94 KWh (10.6%)	224.29 kWh (26.8%)
Helpdesk & IT Support	9788.90 kWh	2594.27 KWh (26.5%)	9056.64	23.0	25.36	1348.97 (18.7%)	118.22 KWh (25.64%)
Sales	2240.92	280.59 KWh (12.52%)	8483.38	52.13	45.5	204.76 KWh (14.87%)	77.82 KWh (8.4%)
Supplies	5334.55	786.83 kWh (14.74%)	6962.90	33.10	27.10	585.20 kWh (15.51%)	5.5 KWh (7.7%)
Technical	2773.51	982.83 KWh (39.5%)	7767.60	20.70	22.0	763.62 KWh (35.34%)	~0 KWh (0%) Model accuracy very low 12%

The results demonstrate significant variation in energy savings across teams, with some teams like 'Technical' at DAEM achieving remarkable PC energy savings (~35%), highlighting the potential of the approach when teams are highly engaged. Conversely, the negative savings for the 'Eficiencia' team at ICAEN are most probably associated with the low user engagement, i.e., <5 PC and <1 Lights challenges completed, similarly to the 'Direccio' team, i.e., with 0 challenges completed, that also has minimal energy savings. Potential issues with the baseline model or unobserved exceptional circumstances (e.g., special events, meetings, etc.) may provide further explanation for the higher energy consumption of the 'Eficiencia' team. In addition, the energy consumption of 'Direccio' and 'Eficiencia' teams that effectively did not participate in the game serve as a control group to show that the energy consumption without the game would either remain constant or even

increase. Evidently, the presence of the game alone is not enough to change behaviour; active participation (i.e., claiming challenges and accomplishing them) is the variable that drives the data accuracy and the savings.

Table 10. The in-game performance of different teams at ICAEN. Important findings are highlighted in bold.

Team Name	Estimated Consumption (KWh)	Estimated Savings (%)	Total Score	Avg. 'Switch-Off PC' Challenges Per User	Avg. 'Switch-off Lights' Challenges Per User	PC Savings (% from Total Est. PC Consumption)	Lighting Savings (% from Total Est. Lights Consumption)
ComGenercat	10,584.16	920.00 kWh (8.7%)	1600.63	12.13	0.88	42.07 KWh (−2%)	−292.70 kWh (−3.95%)
Direccio	14,668.07	374.48 kWh (2.6%)	0	0	0	5.51 KWh 0.00%	−941.49 KWh (−13.95%)
Eficiencia	15,465.47	−942.02 kWh (−25.03%)	716.43	4.43	0.14	−1055.37 KWh (−128.12%)	−2.42 kWh (~0%)
GEA	28,588.24	1938.65 kWh (6.78%)	7440.86	49.29	12.14	591.49 kWh (25.54%)	1363.86 KWh (−5.40%)
IntMobRen	20,249.38	1306.15 kWh (6.45%)	5183.40	33.33	5.17	28.42 KWh (2.07%)	987.35 KWh (5.35%)
Planificacio	17,030.09	3021.95 kWh (16.85%)	3475.00	11	1.43	626.07 KWh (29%)	2392.18 kWh (15.17%)

We also analysed the bivariate correlations of all variables of interest in the clean and the corrected datasets, as illustrated in Figures 3a and 3b, respectively. The total measured energy consumption is significantly correlated to the sum of the energy consumption for lights and PCs, both in the clean (i.e., 0.986) and in the corrected (i.e., 0.989) datasets. Moreover, taking the linear regression of the total energy consumption with respect to the sum of the energy consumption for PCs and lights, we derive $R^2 = 0.98$. Therefore, the sum of the energy consumptions for lights and PCs explains 98% of the variation of the total measured energy consumption. This nearly perfect correlation serves as a critical foundational discovery; it demonstrates that in public office environments, PC and lighting loads are the primary drivers of total building load variance. This validates the study's methodology, proving that monitoring these specific human-driven activities is sufficient to understand the building's overall energy profile without requiring the instrumentation of complex systems like the HVAC. The total measured energy consumption is significantly negatively correlated (i.e., −0.533 and −0.578 in the clean and the corrected datasets, respectively) to the number of achieved Lights challenges. By achieving this challenge, employees conserve energy, which is exactly the objective of this challenge, and thus the total energy consumption is reduced. This result becomes even clearer, observing that the more Lights challenges achieved, the lower the energy consumption for lighting and hence the negative correlation (i.e., −0.508 and −0.561 in the clean and corrected datasets respectively) between these two variables. The standalone effect of the achieved PC challenges (i.e., “turn-off PC after work”) to the total energy consumption is minimal, i.e., turning ON and OFF the PCs is not a good predictor of the variance in total energy consumption. A possible explanation for this could be that while PCs are ON, they consume energy proportionally to their workload, regardless of the rest of the office activities that consume electricity. However, looking at energy consumption for PCs alone, we find that it is moderately negatively correlated (i.e., −0.211 and −0.186 for the clean and corrected datasets, respectively) to the number of PC challenges, as was the case with lighting. This means that by achieving PC challenges, employees conserve energy consumed by their PCs. In addition, achieving more Lights challenges is highly correlated (i.e., 0.628 and 0.632 for the clean and corrected datasets, respectively) to achieving PC challenges.

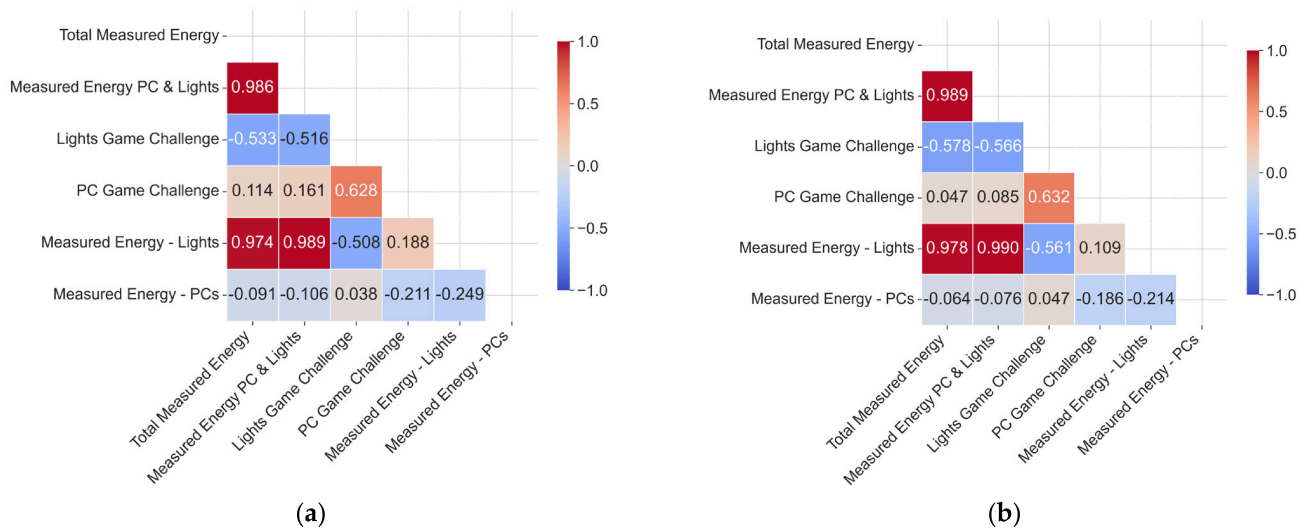


Figure 3. Bivariate correlations in the clean (a) and corrected (b) datasets.

The bivariate correlations show that “Lights Game Challenges” have a strong negative correlation with total consumption (−0.533 to −0.578), while “PC Game Challenges” show almost no correlation (0.047 to 0.114) with total building energy. This result suggests that lighting is a far more reliable indicator of building-wide energy trends than PC usage. This is likely because lighting is a “binary” state (on/off) and often room-wide, whereas PC consumption fluctuates based on CPU load even when “on.” This finding helps building managers prioritize which challenges to incentivize most.

Our linear-regression analysis revealed a two-step explanatory chain that underpins the value of crowdsensing for energy disaggregation. First, we established that the aggregated energy consumption of PCs and lights was an almost perfect predictor of the total building energy consumption ($R^2 = 0.98$), as depicted in Figure 4. Second, we found that the number of accomplished PC and Lights challenges was a strong predictor of the PC + lights energy consumption ($R^2 = 0.65$), as illustrated in Figure 5. By extension, this implies that user-reported data alone can explain approximately 62% of the variance in the total energy consumption of the office buildings, providing a robust foundation for a crowdsensed disaggregation model, as depicted in Figure 6.

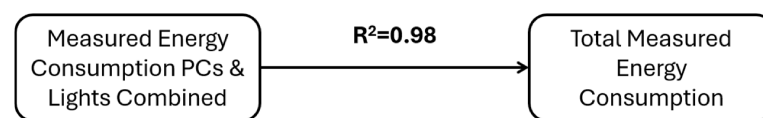


Figure 4. Aggregate energy consumption of PC and Lights as a predictor of the total energy consumption. This correlation is statistically significant ($p < 0.001$).

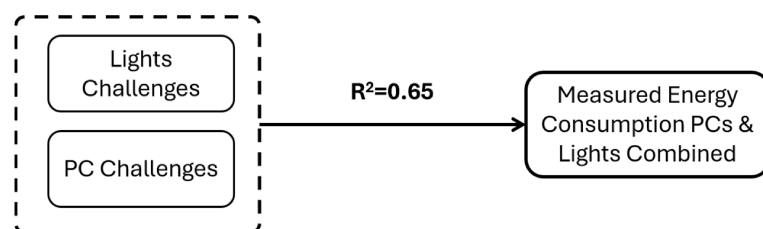


Figure 5. PC and Lights accomplished challenges as a predictor of the energy consumption of PC and lights. Linear Regression formula: [Predicted PC&Lights Energy Consumption] = $189.48 + \{19.01 \times [\text{Chal1_PC}]\} - \{25.43 \times [\text{Chal5_Lights}]\}$. This correlation is statistically significant ($p < 0.001$).

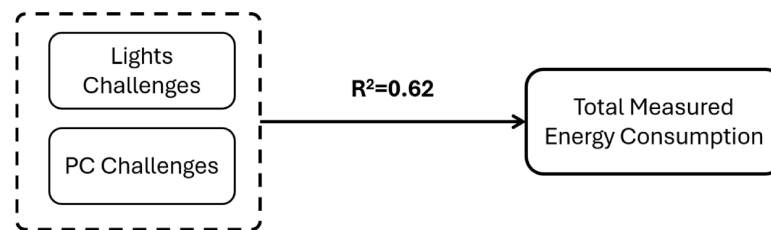


Figure 6. PC and Lights accomplished challenges as a predictor of the total energy consumption. Linear Regression formula: [Predicted Total Energy Consumption] = $227.33 + \{19.11 \times [\text{Chal1_PC}]\} - \{26.90 \times [\text{Chal5_Lights}]\}$. Pearson correlation is shown and is statistically significant ($p < 0.001$).

This result is important: Firstly, it means that the total energy consumption due to total user activities in office buildings is highly correlated to their PC and lights energy consumption. Secondly, since employing multi-channel smart meters for measuring energy consumption for PCs and lights would be very costly, being able to explain 62% of the total energy consumption based on mere user claims, as we have seen above, seems extraordinary.

Furthermore, to account for the hierarchical structure of the data (users nested within teams) and to test for temporal autocorrelation, we utilized a Linear Mixed-Effects Model (LMM). The LMM revealed a substantial baseline variance attributed to the random effect of the *Team* cluster (Group Var = 24,872.95), mathematically confirming that the vast majority of variance in total energy consumption is driven by baseline organizational differences between teams. Testing for temporal carry-over, the fixed effect for *Week* was highly insignificant ($p = 0.728$), indicating that the gamified challenges operated as independent events without confounding temporal habituation. Crucially, even after rigorously controlling for time and these massive group-level differences, the gamified user claims maintained their explanatory power: accomplished PC challenges remained a statistically significant predictor of total energy consumption ($p = 0.047$), and Lights challenges demonstrated marginal significance ($p = 0.073$).

The magnitude of this group-level variance yields a critical insight into the mechanics of office energy consumption: structural, baseline energy use is overwhelmingly determined by local organizational culture (e.g., team-specific operational hours, shared equipment usage, and collective norms), rather than temporal shifts. However, the fact that the gamified user claims retained statistical significance despite this massive organizational ‘noise’ proves the robustness of the human-in-the-loop signal. It demonstrates that while we cannot predict a team’s baseline energy consumption without local calibration, the behavioral changes induced by the gamified application consistently yield measurable, predictable reductions in aggregate load regardless of that starting baseline.

From a practical standpoint, these results offer a transformative insight for facility management. The ability to explain roughly 62% of the variance in total building energy consumption using only validated claims for PC and lighting usage proves that exhaustive device-level instrumentation is largely unnecessary for identifying macro-level energy trends. Instead of deploying expensive smart plugs on hundreds of individual office devices, building managers can achieve a highly actionable level of visibility by targeting just a few binary-state, human-driven loads. This significantly lowers the barrier to entry for energy auditing and continuous commissioning in legacy public buildings.

Subgroup Analysis

We have established that we are able to predict total energy consumption for the office building as a whole based on user claims up to a certain degree. In the sequel, we investigate whether this result can be true for subgroups of users and subareas in the

building as well. The correlations between the sum of the electricity consumptions for lights and PCs per team with respect to the total measured energy consumption per team based on the corrected dataset are depicted in Figure 7 below.

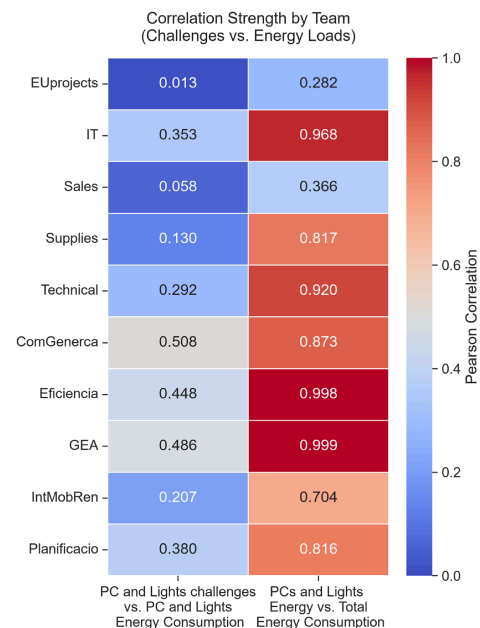


Figure 7. Per-team correlation of PC and Lights challenges with total energy consumption.

By providing incentives even to specific teams of users (i.e., subsets) to claim PCs and Lights challenges, we are able to explain significant fractions (i.e., over 50%) of the variation of the total measured energy consumption for 8/10 teams. This is a significant result, meaning that claims from selected teams can explain significant percentages of the variance of the total energy consumption.

Using eight teams as a training set for finding the linear regression coefficients of PC and Lights challenges for predicting the sum of the energy consumption of PCs and lights, and two teams as test set, we find that the predictive power of the model drops significantly depending on the sections of the teams that belong to the training and the test sets. This is expectable based on the table above that shows the correlations of PC and Lights challenges with the sum of the energy consumptions for PCs and lights per team to be significantly lower than their overall value of 0.646. This means that teams were motivated with different effectiveness to claim their PC and lighting activities.

Using only DAEM data, the aforementioned correlation drops to 0.172, while for ICAEN only, it is found to be 0.471. Using ICAEN as training set and DAEM as test set, we find a correlation of -0.23 (negative predictive power). Using DAEM as training set and ICAEN as test set, we find a correlation of -0.685 . Inside different organizations, using all other DAEM teams except for Technical as training set and Technical team as test set, we find -0.35 . Using all other ICAEN teams except for Planificacio as training set and Planificacio as test set, we find 0.098. Moreover, it is important to note that all leave-one-team-out analyses show positive correlations of PC and Lights challenges with the energy consumption of PCs and lights for all teams, except for Direccio and Eficiencia teams that effectively did not play the game. This result proves that human-centric energy models are highly sensitive to local office culture and the specific “effectiveness of motivation” within a group. Therefore, a “one-size-fits-all” regression model for crowdsensing cannot be applied across different companies without local calibration.

These empirical cross-prediction failures are explicitly mathematically explained by the Linear Mixed-Effects Model (LMM) results detailed in the previous section. The mas-

sive baseline variance attributed to the *Team* random intercept (Group Var = 24,872.95) quantifies exactly why models trained on one subgroup fail when applied to another. Because the overwhelming majority of the baseline variance in total energy consumption is dictated by the specific team being observed, a pooled ordinary regression model cannot successfully transfer across these distinct group boundaries. Ultimately, the LMM mathematically confirms what this subgroup analysis observes empirically: human-centric crowdsensing models are inextricably bound to their local organizational context and strictly require baseline calibration to account for the unique operational culture of each specific team environment.

6. Discussion

Our approach achieves an explanatory power ($R^2 = 0.62$ and $R^2 = 0.65$ for total and key loads, respectively) that is competitive with the lower end of NILM accuracy ranges reported in the literature (55–88% for electric vehicles, pool pumps, and refrigerators) [62], but through a fundamentally different, low-intrusive, and lower-cost paradigm. Instead of inferring device states from aggregate data using complex algorithms, we obtain them directly from occupants, turning a data analytics problem into a human engagement challenge. The infrastructure cost of our system (NFC tags, beacons, a few smart meters) is substantially lower than installing smart plugs or dedicated circuit meters on every device in an office, rendering it especially well-suited for upgrading existing building infrastructures. Therefore, our findings enhance those of recent research suggesting that behavioural interventions, including gamification, remain an underutilised yet critical lever for improving building energy efficiency [63]. However, we note that the reported R^2 scores should be interpreted as indicative of strong associations within the studied environments rather than as universally generalizable predictive models to be further corroborated through more extensive experimentation in various contexts and scales.

A significant finding from our bivariate analysis is the structural difference between load types. Lighting challenges proved to be far more reliable proxies for building-wide trends than PC usage. This asymmetry between lighting and PC challenges as predictors of total consumption aligns with the NILM literature on appliance characteristics. In [64], the authors classify appliances along several operational dimensions, including binary on/off versus multi-state and finite versus continuous operation. This is attributed to the ‘binary’ nature of lighting (on/off), whereas PC consumption fluctuates significantly based on CPU workload even while the device is powered on. A PC that is on may draw anywhere from 15 W (idle) to 60 W (under heavy CPU load), making the relationship between user claims (I turned off my PC) and actual energy reduction less deterministic. Consequently, future crowdsensing efforts should prioritize binary-state loads, where user claims serve as more reliable proxies for measurable energy changes to maximize predictive accuracy.

This asymmetrical reliability between load types highlights a crucial theoretical divergence between traditional NILM and our crowdsensed approach. NILM struggles in commercial buildings precisely because of homogeneous, multi-state loads (like hundreds of identical PCs shifting CPU states) that create overlapping, indistinguishable electrical signatures. Our human-centric approach bypasses the electrical signature entirely, capturing the semantic intent of the user (e.g., “I am leaving for the day”). However, this introduces a different trade off: while we eliminate the need for high-frequency signal processing, our accuracy is entirely tethered to sustained user engagement and the localized effectiveness of the gamification mechanics. If user participation drops, the ‘sensor network’ effectively goes offline. Therefore, while our method circumvents the hardware and algorithmic bottlenecks of NILM, it requires ongoing behavioral maintenance. Future hybrid systems could optimally combine these paradigms: using sparse crowdsensing

to establish ground-truth labels for binary states (like lighting and occupancy), thereby simplifying the residual signal disaggregation left to algorithmic NILM.

Furthermore, it is important to contextualize these results against standard baseline approaches. Traditional time-feature-only models or simple occupancy-based models often rely on static operational schedules or broad presence detection, which inherently fail to capture the granular behavioral intent and ad hoc device interactions of individual occupants. Conversely, while typical office NILM schemes attempt to computationally bypass these limitations, they struggle with the aforementioned homogeneous loads and require substantial capital expenditure for high-frequency smart meters. By contrast, our sparse-infrastructure crowdsensing approach demonstrates that 62% of total building energy variance can be explained without relying on predictive time-schedules or exhaustive hardware instrumentation. While standard baseline models may offer fully automated operation, our human-in-the-loop method delivers highly competitive macro-level insights at a fraction of the deployment cost, effectively trading algorithmic complexity for behavioral engagement.

While our sensing infrastructure was sparse, verifying different claim types in very large or complex buildings can be made simple and thus scalable. Moreover, there is a certain user perception of action verifiability that makes them honest in their claims overall. The fact that ~25–30% of challenges were not verified to have been accomplished by the users is a positive finding for the system's integrity and provides the right incentives to users. Thus, a robust verification infrastructure (NFC, BLE) is necessary not just for points, but to filter out "noise" (unverified claims) that would otherwise degrade the 62% predictive power.

From another point of view, the failure of the 25–30% challenge failure rate observed across both pilot sites offers insight into the data quality problem that persists across the crowdsensing literature. In [37], the authors identify data quality as one of the main operational challenges in MCS, noting that participants may be unreliable, unevenly distributed, or contextually biased. This finding supports the design principle that crowdsensing systems should incorporate verification infrastructure as a structural component. Still, the cost of this verification layer (a few NFC stickers and BLE beacons per office area) is low compared to full sub-metering, yet sufficient to maintain data integrity.

While the study was conducted in two specific office cultures and contexts, and the results in this paper focused primarily on PCs and lights, more challenges were defined in the game, while the methodology to create challenges that target specific energy-consumption behaviours in other contexts is quite generic and comprehensive.

The users often lack knowledge and motivation to save energy and/or crowdsense energy-consumption data. In-game challenges guide the users on how to save energy to alleviate lack of knowledge. In addition, specific points per energy saving action are assigned taking into account: (a) the likelihood of the user to take the action and award more points to less likely behaviours and (b) the energy consumption impact of the specific devices associated to the energy-saving action. Monetary awards are given to the users based on a lottery where each participates with as many tickets as the points he aggregated in the game.

Few limitations should be considered when interpreting the findings of this study. First, a relatively small number of employees (80) participated in the game in the two office sites. Although the longitudinal design yielded 155 team-week observations, allowing the investigation of within-team changes over time and improving the efficiency of estimating temporal associations in comparison with a single cross-sectional assessment [65,66], these repeated observations do not compensate for the limited number of participating organizations or employees. Consequently, the external validity of the findings is limited, and

caution should be exercised when generalizing the results to other organizational settings, industries, or workforce populations. Second, the proposed approach was evaluated in two different office environments, yet not across a broad range of office workplace contexts. Future studies should examine the approach in more diverse organizational contexts, office layouts, and team structures to establish the robustness and generalizability of the reported results. A key limitation, and simultaneously a major theoretical finding, of the proposed approach is its inherent sensitivity to organizational context. The failure of cross-site model prediction, where models trained on one organization yielded negative correlations when tested on another, highlights that human-centric energy models are inherently non-transferable out-of-the-box. This is consistent with recent research emphasizing the importance of incorporating occupant behaviour explicitly to improve the transferability of building energy prediction models [67].

Our supplementary Linear Mixed-Effects analysis fundamentally corroborated this localized phenomenon, while providing a mathematically rigorous separation of nested data effects. The LMM revealed a massive baseline variance explained entirely by the random intercept of the *Team* cluster (Group Var = 24,872.95). This statistical layout provides empirical proof for our assertion that a “one-size-fits-all” model cannot be deployed across varied teams or organizations without prior calibration; the overriding majority of total energy variance stems from baseline physical and organizational differences unique to each team environment rather than temporal shifts. Crucially, the fixed effect for *Week* was shown to be completely insignificant ($p = 0.728$), mathematically proving that the gamified challenges operate as independent weekly events, clean from temporal carry-over dependencies or structural habituation.

Deepening this analysis, the LMM results suggest that building energy networks must be treated as socio-technical systems rather than purely physical ones. The requirement for a localized calibration period is not a flaw of the crowdsensing methodology, but rather a fundamental characteristic of human-in-the-loop IoT systems. Because teams operate with distinct micro-cultures, the “sensor calibration” in this context equates to measuring a specific team’s engagement baseline.

Ultimately, even when subjected to strict controls for time and the massive baseline differences between team clusters, the primary human crowdsensing indicator remained resilient, with accomplished PC challenges maintaining true statistical significance ($p = 0.047$) and Lights challenges establishing marginal significance ($p = 0.073$). This proves that the behavioral layer captures a highly robust signal capable of proxying key occupant loads. From a practical standpoint, this suggests that building managers can obtain actionable, macro-level insights on occupant-driven loads, and verify the impact of energy-saving interventions, simply by calibrating a gamified app to local teams for a short period, entirely circumventing the prohibitive costs of device-level smart metering retrofits.

7. Conclusions

This study establishes a paradigm shift in energy disaggregation, proving that the computational challenges of traditional Non-Intrusive Load Monitoring (NILM) can be successfully bypassed by moving the disaggregation task to the behavioral layer. By utilizing a carefully designed, risk-based gamified incentive mechanism (Impact $\times$ Likelihood), we effectively transformed building occupants into reliable “human sensors”. This user-centred methodology successfully elicited high-quality data, enabling the explanation of the majority of the variance in total building energy consumption based solely on verified user claims for PC and lighting use. Consequently, this human-in-the-loop framework provides building managers with a highly scalable, privacy-friendly, and cost-effective alternative to exhaustive hardware sub-metering for understanding occupant-driven energy loads.

The pilot experiments with real users proved that the summed PC and lighting load explained 98% ($R^2 = 0.98$) of the total consumption variance. Thus, lights and PCs are significant proxies of human energy-consumption activities. Moreover, we established that when challenges for lighting and PCs are employed in office buildings, they can explain a significant fraction of 65% of the variation in aggregate energy consumption for these specific devices, i.e., PCs and lights, and at least 62% of the variance for the total building energy consumption. We have found that Lights challenges correlate -0.533 to -0.578 with total energy consumption, while the respective correlation for PC challenges is only 0.047 to 0.114 . Our findings indicate that Lights challenges are a more reliable indicator of building-wide trends than PC usage, likely because lighting acts as a binary state of occupancy, whereas PC consumption fluctuates based on workload.

Analysis across different organizations revealed that human-centric energy models are highly sensitive to local office culture and the specific effectiveness of motivation within a group. The drop in correlation when moving between pilot sites proves that a one-size-fits-all regression model cannot be applied without local calibration. Furthermore, the fact that 25% to 30% of challenges were not accomplished by users indicates that a robust verification infrastructure, such as NFC and BLE, is necessary to maintain system integrity and filter out the noise of unverified claims.

To validate the robustness of this crowdsensing signal against the inherent statistical biases of nested group dynamics and time, this study employed mixed-effects modelling. This rigorous analysis mathematically confirmed that while specific user claims remain significant predictors of energy consumption, the overwhelming majority of baseline energy variance is dictated by local team micro-cultures rather than temporal habituation. Rather than a limitation, this reveals that organizational energy networks must be treated as socio-technical systems. Therefore, this study concludes that human-centric crowdsensing is a highly viable disaggregation tool, provided that deployments structurally incorporate a localized baseline calibration period to capture the unique operational footprint of each specific team. Once calibrated, building managers can continuously monitor macro-level energy trends directly through user engagement, drastically lowering the barrier to entry for continuous commissioning in legacy buildings.

As future work, we intend to extend the challenges to include additional loads such as HVAC, printers, and other shared appliances to improve the comprehensiveness of the disaggregation model. We also plan to investigate whether user engagement and model accuracy persist in the long run and employ machine learning models to dynamically adapt challenge points and provide personalized feedback. This is in accordance with recent studies [51,63], outlining that future research should aim to further explore personalized and adaptive incentive mechanisms and their importance for long-term engagement and sustaining behavioural change. Finally, we aim to test and refine the approach in other building types, such as schools, laboratories, and retail spaces.

Author Contributions: Conceptualization, T.G.P. and D.K.; methodology, T.G.P., D.K. and C.B.; software, T.G.P., D.K. and C.B.; validation, T.G.P., D.K. and C.B.; formal analysis, T.G.P. and D.K.; investigation, T.G.P., D.K. and C.B.; resources, T.G.P., D.K. and C.B.; data curation, T.G.P.; writing—original draft preparation, T.G.P.; writing—review and editing, T.G.P., D.K. and C.B.; visualization, T.G.P. and D.K.; supervision, T.G.P. and D.K.; project administration, T.G.P., D.K. and C.B.; and funding acquisition, T.G.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Horizon Europe project RESONANCE, grant number 101096200.

Data Availability Statement: The data employed in this study are collectively owned by the H2020 ChArGED project (grant number 696170) partners and cannot be publicly disclosed at this time due to privacy restrictions. Nevertheless, the authors will make every effort to satisfy individual data requests from interested parties.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Commission, Directorate-General for Energy. Energy Performance of Buildings Directive. 2026. Available online: https://energy.ec.europa.eu/topics/energy-efficiency/energy-performance-buildings/energy-performance-buildings-directive_en (accessed on 10 June 2026).
2. European Environment Agency (EEA). Addressing the Environmental and Climate Footprint of Buildings. Report. 2024. Available online: <https://www.eea.europa.eu/en/analysis/publications/addressing-the-environmental-and-climate-footprint-of-buildings> (accessed on 30 May 2026).
3. European Environment Agency (EEA). Greenhouse Gas Emissions from Energy Use in Buildings in Europe (EEA Indicator Report). 2025. Available online: <https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emissions-from-energy> (accessed on 30 May 2026).
4. European Climate Neutrality Observatory (ECNO). 2025 Flagship Report: State of EU Progress to Climate Neutrality—An Indicator-Based Assessment Across 13 Building Blocks for a Climate Neutral Future. 2025. Available online: https://climateobservatory.eu/sites/default/files/2025-09/Flagship-Report-2025_online.pdf (accessed on 10 June 2026).
5. European Commission. Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the Energy Performance of Buildings (Recast). Official Journal of the European Union. Document 02024L1275-2026052. 2024. Available online: <http://data.europa.eu/eli/dir/2024/1275/2026-05-24> (accessed on 10 June 2026).
6. Buildings Performance Institute Europe (BPIE). *EU Buildings Climate Tracker*, 3rd ed.; BPIE: Brussels, Belgium, 2024; Available online: <https://www.bpie.eu/publication/eu-buildings-climate-tracker-3rd-edition/> (accessed on 10 June 2026).
7. Schirmer, P.A.; Mporas, I. Non-intrusive load monitoring: A review. *IEEE Trans. Smart Grid* **2023**, *14*, 769–784. [CrossRef]
8. Xu, X.; Yu, H.; Sun, Q.; Tam, V.W.Y. A critical review of occupant energy consumption behavior in buildings: How we got here, where we are, and where we are headed. *Renew. Sustain. Energy Rev.* **2023**, *182*, 113396. [CrossRef]
9. Kollmuss, A.; Agyeman, J. Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* **2002**, *8*, 239–260. [CrossRef]
10. Simon, J.; Jahn, M.; Al-Akkad, A. Saving energy at work: The design of a pervasive game for office spaces. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM 2012)*; ACM: New York, NY, USA, 2012. [CrossRef]
11. Apostolou, G.; Krinidis, S.; Ioannidis, D.; Tzovaras, D.; Borges, C.E.; Casado-Mansilla, D.; López-de-Ipiña, D. GreenSoul: A novel platform for the reduction of energy consumption in communal and shared spaces. In *4th International Symposium on Environment Friendly Energies and Applications (EFEA 2016)*; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]
12. Orland, B.; Ram, N.; Lang, D.; Houser, K.; Kling, N.; Coccia, M. Saving energy in an office environment: A serious game intervention. *Energy Build.* **2014**, *74*, 43–52. [CrossRef]
13. García, Ó.; Alonso, R.S.; Prieto, J.; Corchado, J.M. Energy efficiency in public buildings through context-aware social computing. *Sensors* **2017**, *17*, 826. [CrossRef] [PubMed]
14. Abdurahmanovic, N.; Cadenbach, A. Enhancing energy efficiency through user engagement and behaviour change: A review on gamification approaches and serious games in energy systems. *Energy* **2025**, *334*, 137496. [CrossRef]
15. Papaioannou, T.G.; Dimitriou, N.; Vasilakis, K.; Schoofs, A.; Nikiforakis, M.; Pursche, F.; Deliyski, N.; Taha, A.; Kotsopoulos, D.; Bardaki, C.; et al. An IoT-based gamified approach for reducing occupants' energy wastage in public buildings. *Sensors* **2018**, *18*, 537. [CrossRef] [PubMed]
16. Ganti, R.K.; Ye, F.; Lei, H. Mobile crowdsensing: Current state and future challenges. *IEEE Commun. Mag.* **2011**, *49*, 32–39. [CrossRef]
17. Yu, Z.; Ma, H.; Guo, B.; Yang, Z. Crowdsensing 2.0. *Commun. ACM* **2021**, *64*, 76–80. [CrossRef]
18. Papaioannou, T.G.; Vasilakis, K.; Dimitriou, N.; Garbi, A.; Schoofs, A. A sensor-enabled rule engine for changing energy-wasting behaviours in public buildings. In *2018 IEEE International Energy Conference (ENERGYCON)*; IEEE: Piscataway, NJ, USA, 2018. [CrossRef]
19. Angelis, G.-F.; Timplalexis, C.; Krinidis, S.; Ioannidis, D.; Tzovaras, D. NILM applications: Literature review of learning approaches, recent developments and challenges. *Energy Build.* **2022**, *261*, 111951. [CrossRef]
20. Meier, A.; Cautley, D. Practical limits to the use of non-intrusive load monitoring in commercial buildings. *Energy Build.* **2021**, *248*, 111308. [CrossRef]

21. Carrie Armel, K.; Gupta, A.; Shrimali, G.; Albert, A. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* **2013**, *52*, 213–234. [CrossRef]
22. Gupta, A.; Chakravarty, P. Impact of energy disaggregation on consumer behavior. In Proceedings of the 2013 BECC Conference, Sacramento, CA, USA, 18–20 November 2013. Available online: <https://escholarship.org/uc/item/62d3456p> (accessed on 10 June 2026).
23. Faustine, A.; Mvungi, N.H.; Kaijage, S.; Michael, K. A survey on non-intrusive load monitoring methods and techniques for energy disaggregation problem. *arXiv* **2017**, arXiv:1703.00785.
24. Iqbal, H.K.; Malik, F.H.; Muhammad, A.; Qureshi, M.A.; Abbasi, M.N.; Chishti, A.R. A critical review of state-of-the-art non-intrusive load monitoring datasets. *Electr. Power Syst. Res.* **2021**, *192*, 106921. [CrossRef]
25. Amenta, V.; Tina, G.M. Load demand disaggregation based on simple load signature and user's feedback. *Energy Procedia* **2015**, *83*, 380–388. [CrossRef]
26. Sundhu, A.A.; Rottondi, C.; Verticale, G. Real-time disaggregation of residential energy consumption enhanced with user feedback. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)*; IEEE: Piscataway, NJ, USA, 2019. [CrossRef]
27. Miyasawa, A.; Matsumoto, M.; Fujimoto, Y.; Hayashi, Y. Energy disaggregation based on semi-supervised matrix factorization using feedback information from consumers. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*; IEEE: Piscataway, NJ, USA, 2017. [CrossRef]
28. Kormos, C.; Gifford, R. The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *J. Environ. Psychol.* **2014**, *40*, 359–371. [CrossRef]
29. Suomalainen, K.; Eyers, D.; Ford, R.; Stephenson, J.; Anderson, B.; Jack, M. Detailed comparison of energy-related time-use diaries and monitored residential electricity demand. *Energy Build.* **2019**, *183*, 418–427. [CrossRef]
30. Brown, R.; Ghavami, N.; Siddiqui, H.U.R.; Adjrad, M.; Ghavami, M.; Dudley-McEvoy, S. Occupancy based household energy disaggregation using ultra wideband radar and electrical signature profiles. *Energy Build.* **2017**, *141*, 134–141. [CrossRef]
31. Flett, G.; Kelly, N. A disaggregated, probabilistic, high resolution method for assessment of domestic occupancy and electrical demand. *Energy Build.* **2017**, *140*, 171–187. [CrossRef]
32. Rafsanjani, H.N.; Ahn, C.R.; Chen, J. Linking building energy consumption with occupants' energy-consuming behaviors in commercial buildings: Non-intrusive occupant load monitoring (NIOLM). *Energy Build.* **2018**, *172*, 317–327. [CrossRef]
33. Tang, G.; Ling, Z.; Li, F.; Tang, D.; Tang, J. Occupancy-aided energy disaggregation. *Comput. Netw.* **2017**, *117*, 42–51. [CrossRef]
34. Rafiq, H.; Manandhar, P.; Rodriguez-Ubinas, E.; Qureshi, O.A.; Palpanas, T. A review of current methods and challenges of advanced deep learning-based non-intrusive load monitoring (NILM) in residential context. *Energy Build.* **2024**, *305*, 113890. [CrossRef]
35. Kostakos, V.; Rogstadius, J.; Ferreira, D.; Hosio, S.; Goncalves, J. Human sensors. In *Participatory Sensing, Opinions and Collective Awareness*; Loreto, V., Haklay, M., Hotho, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 69–92. [CrossRef]
36. Ma, H.; Zhao, D.; Yuan, P. Opportunities in mobile crowd sensing. *IEEE Commun. Mag.* **2015**, *52*, 29–35. [CrossRef]
37. Liu, J.; Shen, H.; Narman, H.S.; Chung, W.; Lin, Z. A survey of mobile crowdsensing techniques: A critical component for the Internet of Things. *ACM Trans. Cyber-Phys. Syst.* **2018**, *2*, 18. [CrossRef]
38. Capponi, A.; Fiandrino, C.; Kantarci, B.; Foschini, L.; Kliazovich, D.; Bouvry, P. A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2419–2465. [CrossRef]
39. Wang, L.; Zhang, D.; Wang, Y.; Chen, C.; Han, X.; M'Hamed, A. Sparse mobile crowdsensing: Challenges and opportunities. *IEEE Commun. Mag.* **2016**, *54*, 161–167. [CrossRef]
40. Dabirian, S.; Panchabikesan, K.; Eicker, U. Occupant-centric urban building energy modeling: Approaches, inputs, and data sources—A review. *Energy Build.* **2022**, *257*, 111809. [CrossRef]
41. Froehlich, J.E. Promoting energy efficient behaviors in the home through feedback: The role of human-computer interaction. In Proceedings of the HCIC 2009 Winter Workshop, Fraser, CO, USA, 4–8 February 2009.
42. Lesic, V.; Bruine De Bruin, W.; Davis, M.C.; Krishnamurti, T.; Azevedo, I.M.L. Consumers' perceptions of energy use and energy savings: A literature review. *Environ. Res. Lett.* **2018**, *13*, 033004. [CrossRef] [PubMed]
43. Attari, S.Z.; DeKay, M.L.; Davidson, C.I.; Bruine De Bruin, W. Public perceptions of energy consumption and savings. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16054–16059. [CrossRef] [PubMed]
44. Cottafava, D.; Magariello, S.; Ariano, R.; Arrobio, O.; Baricco, M.; Barthelmes, V.M.; Baruzzo, G.; Bonansone, M.; Console, L.; Contin, L.; et al. Crowdsensing for a sustainable comfort and for energy saving. *Energy Build.* **2019**, *186*, 208–220. [CrossRef]
45. Jayathissa, P.; Quintana, M.; Abdelrahman, M.; Miller, C. Humans-as-a-sensor for buildings—Intensive longitudinal indoor comfort models. *Buildings* **2020**, *10*, 174. [CrossRef]
46. Cruz-Rangel, D.; Ocampo-Martinez, C.; Diaz-Rozo, J. Online non-intrusive load monitoring: A review. *Energy Nexus* **2025**, *17*, 100348. [CrossRef]

47. Sobot, T.; Stankovic, V.; Stankovic, L. Human in the loop active learning for time-series electrical measurement data. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108589. [CrossRef]
48. Yang, D.; Xue, G.; Fang, X.; Tang, J. Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones. *IEEE/ACM Trans. Netw.* **2016**, *24*, 1732–1744. [CrossRef]
49. Yang, G.; He, S.; Shi, Z.; Chen, J. Promoting cooperation by the social incentive mechanism in mobile crowdsensing. *IEEE Commun. Mag.* **2017**, *55*, 86–92. [CrossRef]
50. Zhang, X.; Yang, Z.; Sun, W.; Liu, Y.; Tang, S.; Xing, K.; Mao, X. Incentives for mobile crowd sensing: A survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 54–67. [CrossRef]
51. Li, W.-T.; Iuorio, O.; Fang, H.; Mak, M.W.T. Gamification approaches and assessment methodologies for occupants' energy behavior change in buildings: A systematic review. *Buildings* **2024**, *14*, 1497. [CrossRef]
52. Johnson, D.; Deterding, S.; Kuhn, K.A.; Staneva, A.; Stoyanov, S.; Hides, L. Gamification for health and wellbeing: A systematic review of the literature. *Internet Interv.* **2016**, *6*, 89–106. [CrossRef] [PubMed]
53. Lee, C.-I.; Chen, I.-P.; Hsieh, C.-M.; Liao, C.-N. Design aspects of scoring systems in game. *Art Des. Rev.* **2017**, *5*, 26–43. [CrossRef]
54. Malone, T.W. Toward a theory of intrinsically motivating instruction. *Cogn. Sci.* **1981**, *5*, 333–369. [CrossRef] [PubMed]
55. Darejeh, A.; Salim, S.S. Gamification solutions to enhance software user engagement—A systematic review. *Int. J. Hum.-Comput. Interact.* **2016**, *32*, 613–642. [CrossRef]
56. Kotsopoulos, D.; Bardaki, C.; Lounis, S.; Papaioannou, T.; Pramatar, K. Designing an IoT-enabled gamification application for energy conservation at the workplace: Exploring personal and contextual characteristics. In Proceedings of the 30th BLED eConference (BLED 2017), Bled, Slovenia, 18–21 June 2017.
57. Kotsopoulos, D.; Bardaki, C.; Lounis, S.; Papaioannou, T.; Pramatar, K. Gamification at work: Employee motivations to participate and preferences for energy conservation. In Proceedings of the 12th Mediterranean Conference on Information Systems (MCIS 2018), Corfu, Greece, 27–30 September 2018.
58. Garvey, P.R. *Analytical Methods for Risk Management: A Systems Engineering Perspective*; Chapman-Hall/CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2008; ISBN 1584886374.
59. Dumbravă, V.; Iacob, V.-S. Using probability–impact matrix in analysis and risk assessment projects. *J. Knowl. Manag. Econ. Inf. Technol.* **2013**, *3*, 1–7.
60. MITRE Corporation. Risk impact assessment and prioritization. In *MITRE Systems Engineering Guide*; The MITRE Corporation: McLean, VA, USA, 2013. Available online: <https://www.mitre.org/our-impact/mitre-labs/systems-engineering-innovation-center/risk-impact-assessment> (accessed on 31 May 2026).
61. Kneebone, S.; Smith, L.; Fielding, K. The Impact-Likelihood Matrix: A policy tool for behaviour prioritisation. *Environ. Sci. Policy* **2017**, *70*, 9–20. [CrossRef]
62. Mayhorn, E.; Butzbaugh, J.; Meier, A. A field study of nonintrusive load monitoring devices and implications for load disaggregation. *Sensors* **2023**, *23*, 8253. [CrossRef] [PubMed]
63. Wang, T.; Xin, Z.; Ma, X.; Wang, S.; Ametefe, D.S.; John, D.; Darboe, T. Gamification as a catalyst for energy efficiency in buildings: Behavioural design, engagement dynamics, and implementation pathways. *Results Eng.* **2025**, *22*, 106807. [CrossRef]
64. Kaselimi, M.; Protopapadakis, E.; Voulodimos, A.; Doulamis, N.; Doulamis, A. Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring. *Sensors* **2022**, *22*, 5872. [CrossRef] [PubMed]
65. Singer, J.D.; Willett, J.B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*; Oxford University Press: Oxford, UK, 2003.
66. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Cambridge University Press: Cambridge, UK, 2007.
67. Zhang, C.; Luo, Z.; Rezgui, Y.; Zhao, T. Enhancing building energy consumption prediction introducing novel occupant behavior models with sparrow search optimization and attention mechanisms: A case study for forty-five buildings in a university community. *Energy* **2024**, *294*, 130896. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.