2022 IEEE PES/IAS PowerAfrica

Rwandan Pay-As-You-Go Solar Home System User Payment Behavioural Types

Vasco Mergulhão
EfID - The Bartlett School of Sustainable Construction
University College London
London, UK
vasco.mergulhao.20@ucl.ac.uk

Licia Capra
Department of Computer Science
University College London
London, UK
l.capra@ucl.ac.uk

Kostas Voglitsis
Lead Data Scientist
Bboxx Ldt
London, UK
k.voglitsis@bboxx.co.uk

Priti Parikh
EfID - The Bartlett School of Sustainable Construction
University College London
London, UK
priti.parikh@ucl.ac.uk

Abstract— **The pay-as-you-go (PAYGo) model is now the principal way through which solar home systems (SHSs) are distributed in Sub-Saharan Africa. By alleviating the upfront cost and providing flexible payment schemes, the PAYGo model helps tackle what is still the main barrier for SHS adoption – i.e., affordability. However, the scheme's design and evaluation are still largely guided by assumptions on user behaviour. This work provides a first evidence-based look into SHS PAYGo user payment patterns and behaviours, by using payment records of over 32,000 Rwandan SHS users. Three clustering algorithms are implemented to conduct a customer segmentation, employing an ensemble validation method which facilitates qualitative oversight. The analysis reveals five user payment behavioural profiles which serve to aid improvement in the current PAYGo model design.**

Keywords— solar home system, pay-as-you-go, user behaviour, customer segmentation, clustering

I. INTRODUCTION

Access to modern energy sources is a key element to ensure sustained welfare improvement, as highlighted by the United Nations designation of universal energy access as its 7th Sustainable Development Goal (SDG) [1]. However, 540 million people in Sub-Saharan Africa (SSA) remain without access to electricity (i.e., 68% of the global deprived population), and current projections see this figure unchanged by 2030 [2]. In addition, despite the rapid urbanization trend, 60% of SSA population still resides in rural areas, where access rates average on 27% in contrast to 77% in urban areas [3].

Standalone off-grid solutions, in particular Solar Home Systems (SHS), have become the most cost-effective solution to deliver energy access in remote rural areas with limited demand and affordability [4]. While some may argue these systems – typically under 100-watt capacity – are too small to produce a meaningful impact [5], mounting evidence reenforces the claim that the cost-benefit ratio delivered by SHSs justifies the investment [6]. The International Energy Agency (IEA) expects SHSs to serve up to 75% of the rural SSA population [7], while the Rwandan government foresees it to reach 38% of its unelectrified population by 2024 [8]. To achieve this, pay-as-you-go (PAYGo) financing schemes will be essential. These flexible payment plans have made SHSs affordable to an additional 40% larger population by reducing upfront costs, and now support 84% of global SHS sales [9] and 96% of Rwandan sales [10]. These schemes were largely made possible by the advent of remote lockout technology, which enforces compliance by locking the SHS unit in the absence of recurrent payments; thus, tying the loan repayment with the energy service [11].

In parallel to providing energy access, smart PAYGo SHSs with Internet of Things (IoT) capabilities have been generating large amounts of data over the past decade [12]. This real-time data detailing device usage and digital payment records constitutes a significant leap in behavioural information; however, despite their potential for improving the design of the PAYGo model, scarcely any studies have leveraged these resources to this end. In an exhaustive survey of studies dedicated to SHSs in the SSA context, the authors in [13] found that only 31% of the 139 studies applied quantitative methods, postulating that this was due to a lack of access to quantitative data.

Indeed, presently it was found that only three studies [14]–[16] leveraged payment records, with a few more examples focusing on energy consumption patterns. In particular, [14] leveraged a relatively rich dataset of 68,600 Rwandan and Kenyan SHS users to investigate customer recruitment strategies; conducting a customer segmentation analysis to identify links between customer types, and demographic and recruitment factors. [15] and [16] analyse the same records from less than 2,000 pico-solar users: in the first evidence was found linking energy consumption behaviour with payment patterns, revealing a lower average consumption in the period preceding a missed payment; while [16] demonstrated that most users followed either a broadly monthly or weekly regime. On the other hand, industry led efforts demonstrate the true scale of the resources the literature thus far has been deprived of; where in [17] over 450,000 users were investigated while in [18] 700,000 user records were accessed – representing 75% of the Ugandan market.

The present work leverages a partnership with Bboxx Ltd – a large-scale SHS provider – to evaluate the PAYGo model through the users' behavioural payment patterns. In the absence of previous large-scale data-driven research, the model's design has so far relied on assumptions which may not reflect the users' reality. This disconnect may in turn be hindering the efficient deployment of SHSs. We find that one of the chief assumptions relates to the main user behavioural architype(s) and the expectation that these can follow daily payment regimes, as implied in industry reports such as [19]. Therefore, this research aims to address this gap by identify the main user behavioural groups found in a sample of over 32,000 Rwandan SHSs users.

II. DATA

This study is made possible through a collaboration with Bboxx Ltd which distributes solar home systems in Rwanda and ten other SSA countries. As with other providers, users interact with the PAYGo scheme by pre-purchasing time-credits to keep the SHS unlocked. The data analysed contain the daily interactions of 32,816 Rwandan SHS users through a record of the amount and date of each payment, as well as a log of their time-credit balance. To avoid the behavioural disturbances likely induced by the COVID-19 crisis, only records prior to April 2020 were considered. In addition, a minimum one year of usage threshold was introduced to ensure users were well acquainted with the service.

Presently we consider a customer has defaulted after 120 consecutive days without credit and for the purposes of this research we have discarded any records collected after this event. In addition to the payment records, we also include information about a customer's maturity, which is calculated based on how much time had passed between the first scheduled payment date and the 1st of April 2020.

III. METHODOLOGY

This work implements a customer segmentation of SHS PAYGo users based on their payment behavioural patterns. This is achieved through a clustering analysis which draws on a set of five aggregate features derived from the time-series payment records, exploring solutions from three different clustering algorithms, and employing an ensemble validation method - which combines validation indices with qualitative inputs - for selecting the preferred solution for further analysis.

A. Feature Engineering

Five aggregate features were designed to capture the most relevant aspects of PAYGo behaviour. All features were calculated for each individual user by combining different aspects of the payment records (i.e., size and date) and the corresponding time-credit balances. One should note that the payment records were normalized by their respective daily rate, thus transforming their monetary value into equivalent days of credit purchased, with so decoupling payment behaviour from SHS system cost and characteristics.

Because Bboxx Ltd implements a seven-day minimum payment penalty fee for late payments, there may be a discrepancy between a user's preferred payment regime and their actual one. Therefore, two features were included to differentiate these behaviours:

Average Payment Size [Pay] – captures the preferred payment regime by providing the overall average payment size for each user, excluding late payments.
Frequency [Freq] – reveals the actual payment frequency, by calculating the average interval of days between payments.

The remaining three features were designed to capture different elements of how users incur into late payment periods:

Percentage of Late Payments [PLP] – computes the percentage of top-up payments made at a point in time when there was no time-credit left in the balance (i.e., late).
Average Consecutive Days Late [avg(CDL)] – computes, on average, how many consecutive days pass with a user having no time-credit balance.
Maximum Consecutive Days Late [max(CDL)] – first introduced in [17], also leverages the records of zero time-credit balances but instead highlights the longest late period length.

Table I displays the features name, units, and summary statistics considering their distribution for the full user sample presently analysed.

TABLE I. AGGREGATE PAY-AS-YOU-GO BEHAVIOURAL FEATURES

Feature Name	Units	[Min, Max]	Average
Average Payment Size [Pay]	[days]	[0.2, 39.8]	10.2
Frequency [Freq]	[days]	[7.9, 41.5]	11.7
Percentage of Late Payments [PLP]	[%]	[2.1, 100]	50
Avg. Consec. Days Late [avg(CDL)]	[days]	[0, 51.5]	10.1
Max. Consec. Days Late [max(CDL)]	[days]	[0, 118]	43.1

It should be noted that all features were rescaled to a [0, 1] range prior to clustering, since some of the algorithms used are sensitive to the variable scales.

B. Clustering Algorithms

The goal of the customer segmentation is to highlight the main behavioural heterogeneities found in the user sample with no basis on a priori knowledge. Therefore, three different clustering algorithms were selected to produce solutions with a diverse set of implicit algorithmic biases.

The first of the three is the k-Means algorithm, which partitions the dataset by assigning datapoints to their nearest centroid, producing balanced and homogenous groups. However, by defining clusters based on Euclidean distances, the k-Means algorithm has a strong bias towards spherical clusters regardless of the underlying data structure [20]. To counteract this, a Spectral Clustering algorithm was also selected, which – through a graphical data transformation and decomposition – can identify clusters of arbitrary shapes [21]. Presently, the nearest neighbour algorithm is used to define the graph for the Spectral Clustering algorithm. Lastly, the Hierarchical Clustering (HC) algorithm, with ward-linkage, which defines clusters through sequential hierarchical structures [20] was also added to increase algorithmic bias diversity.

Each algorithm delivered eight solutions, ranging the number of clusters (i.e., **k**) from 2 to 9, inclusive, resulting in a total of 24 solutions.

C. Ensemble Validation

A two-step validation procedure - composed of a quantitative ensemble method and a qualitative oversight element - has been applied to select the preferred solution. The first step adapts the ensemble method found in [22] to rank and highlight a subset of clustering solutions. After which, the final selection is guided by a qualitative oversight which accounts for algorithmic biases and favours interpretable value for the present context.

The ensemble method leverages three internal Clustering Validation Indices (CVI) to produce a consensus ranking of the clustering solutions: namely, the Silhouette (Sil) index, the Davies-Bouldin (DB) index, and the Caliński-Harabasz (CH) index. Each index scores all solutions from their preferred choice (i.e., 24 points) to least desirable (i.e., 1 point). The combined rank serves to attenuate inevitable index biases, while access to the individual index rankings also allows for an additional level of scrutiny. For instance, it is expected that the CH index should prefer k-Means solutions, while the Sil index should be biased towards solutions with lower **k** [23].

IV. Results

Table II shows that the k-Means algorithm produces the most favoured solutions, with the first alternative emerging only at rank 6. As expected, the CH index had a clear favouritism towards the latter, and apparently so did the Sil index. Only the DB index disagreed, preferring the Spectral Clustering (**k**=6) solution above all. Both top two solutions for the k-Means and Spectral Clustering algorithms are for **k** equal four and five, respectively; demonstrating that a partition within the range of four to five clusters has the best quantitative value, within which the k-Means algorithm dominates. Considering this subset of solutions, then through a qualitative assessment the k-Means (**k**=5) partition emerges as the one which delivers the highest analytical value.

TABLE II. Summary Ensemble CVI Ranking Results

V. Discussion

Clustering Algorithm	Rank	Total	Sil	DB	CH
k-Means (4)	1	68	22	22	24
k-Means (5)	2	61	23	17	21
k-Means (2)	3	58	24	11	23
…					…
Spect. Clust. (4)	6	54	17	18	19
…					…
Spect. Clust. (6)	11	37	7	24	6
HC-ward (2)	11	37	20	1	16

Each of the clusters defined by the k-Means (**k**=5) solution are displayed in Fig. 1 according to the distribution of the five features defined in Table I for their respective user subsamples, with all features in a [0, 1] range. The present findings suggest that Rwandan SHS users broadly follow either a weekly or monthly payment regime, as witness in [16]; beyond this, the main distinguishing factor relates to late period characteristics. Based on these two elements we can describe each cluster as follows:

Cluster 1 – captures the first of three weekly user types – i.e., with a Pay averaging close to seven days. These users are also the ones which are on average late the least often and for the shortest time.

Cluster 2 – contains the group of weekly payers that tend to be late often – i.e., with a distinctly high PLP averaging on 70% – but for short periods at a time.

Cluster 3 – highlights the last weekly payers group, which is characterized by long-lasting late periods.

Cluster 4 – captures the first monthly payers type – with a Pay mode of thirty days – which tend to be late more often and for longer periods.

Cluster 5 – reflects the second monthly payers type which is late less often and for shorter periods of average.

The top-ranking solution (i.e., k-Means (**k**=4)) offered a similar characterisation, apart from the distinction between clusters 4 and 5; this valuable differentiation is what prompted the selection of the present solution.

Table III provides additional user details on the clusters identified by the clustering: the user count and relative share, the percentage of defaulted users, and the user maturity breakdown. Defaults are presented by the absolute share of defaulters per cluster, but also by their relative shift from the overall share of defaulters – which in the given dataset under the present definition of default is 18%. For the maturity level columns only the relative shifts are given, indicating whether a given maturity level is more or less common than in the overall sample.

Table III highlights how new customers appear to prefer weekly payments, since cluster 1 and 2 are the youngest, and that this regime is by far the most popular, followed by 84% of users. From the latter we also find that two factors appear to incur into higher default rates. Firstly, monthly regime followers are more prone to defaulting, with the best of the two – cluster 5 – having the same rate as the worst of the former three – cluster 3. Secondly, defaults also seem to be associated with greater maturity, given how the eldest clusters (i.e., 3 and 4) also have the highest default rates of their respective payment regimes. On the other hand, cluster 2 indicates that a high frequency of late payments does not necessarily imply high default rates if these late periods tend to be short; perhaps revealing that this use of the payment

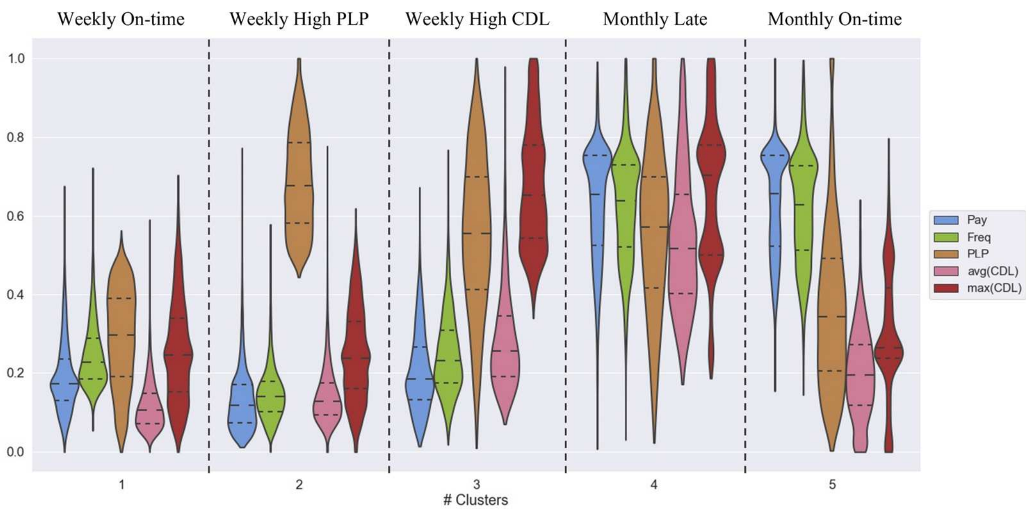

Fig. 1. Customer Segmentation.

TABLE III. CLUSTER PROPERTIES OVERVIEW

Cluster	Description	Count	Defaulted	Maturity Level [a]		
				Young	*Middle*	*Old*
1	Weekly On-time	10509 (32%)	16% (-2pp)	+2.6pp	-0.6pp	-1.9pp
2	Weekly High PLP	10273 (31%)	15% (-3pp)	+5pp	+0.5pp	-5.5pp
3	Weekly High CDL	6113 (19%)	21% (+3pp)	-9.6pp	+2.1pp	+7.5pp
4	Monthly Late	2431 (7%)	28% (+10pp)	-8.8pp	-2.6pp	+11.3pp
5	Monthly On-time	3490 (11%)	21% (+3pp)	+0.5pp	-1.4pp	+1pp

a. Young: 1 to 2 years; Middle: 2 to 3 years; Old: over 3 years. These groups each contain one third of users in the original sample.

flexibility awarded by PAYGo schemes is not a negative trait. Lastly, a strong correlation between default rates and high max(CDL) values was expected, as reported in [17]. However, while clusters 3 and 4 may be the respective worst weekly and monthly performers, their default rates still differ substantially, despite similar max(CDL) values.

This consistent predominance of higher default rates amongst monthly payers may be a sign of how the current PAYGo model design is failing to account for the behavioural heterogeneity of SHS users. Two factors suggest that a monthly regime is not only a preference for these user types, but it is actually a constraint. Firstly, the max(CDL) distributions for clusters 4 and 5 are heavily dominated by local maxima – which correspond to multiples of thirty in the original scale – while for weekly payers the distribution is significantly smoother. Secondly, the avg(CDL) is also markedly higher in monthly payers if we compare the good and worse performers separately – i.e., cluster 4 versus the first two and cluster 5 versus cluster 3, respectively. Both CDL values suggest that monthly payers are often unable to react (e.g., amend a missed payment) in cycles shorter than a month.

However, current PAYGo distributer practices design evaluation metrics, incentives and penalties, and default definitions based on the assumption that all customers operate on a daily cycle, or that at least they can react on this scale. Considering CDL > 120 days as the definition for default, by then a monthly user will have had missed four payment cycles, a weekly payer would miss seventeen, meanwhile the distributer considers they both have missed 120 chances. This discrepancy in opportunities to react before penalization may be partially to blame for higher default rates amongst monthly payers. Regardless, the behavioural heterogeneity found in the sample motivates further investigations into how the PAYGo model design may be improved.

VI. CONCLUSION

Affordability is still the main barrier for SHS adoption [24], and while flexible payment schemes have helped alleviate this problem, the assumptions that guide PAYGo model design and evaluation have received little attention. This despite the data passively collected at scale by smart SHS detailing user interactions with the service.

This work leverages PAYGo data from 32.816 Rwandan SHS users to provide a first characterisation of user types according to their payment behaviours. It finds that, rather than daily payments, users follow either weekly or monthly regimes. Most importantly, however, it highlights how the current PAYGo model design may be failing to accommodate a significant portion of its users. With so, ultimately fomenting the need for further evidence-based analysis of SHS PAYGo user payment patterns and behaviours.

ACKNOWLEDGMENT

We wish to acknowledge funding from the Royal Academy of Engineering and Bboxx through the Smart Solar Solutions for All (S34ALL) fellowship. We would additionally like to thank Bboxx for the access and support needed to navigate the data.

REFERENCES

[1] G. A. UN, “Transforming our world: the 2030 Agenda for Sustainable Development,” Oct. 2015.

[2] IEA, IRENA, UNSD, World Bank, and WHO, “Tracking SDG 7: The Energy Progress Report,” World Bank, Washington, DC, Washington DC, 2020.

[3] UNECE, “Policy Brief on Accelerating achievement of SDG7 in the time of COVID-19 | UNECE,” Dec. 2020.

[4] I. Bisaga, P. Parikh, J. Tomei, and L. S. To, “Mapping synergies and trade-offs between energy and the sustainable development goals: A case study of off-grid solar energy in Rwanda,” Energy Policy, vol. 149, p. 112028, Feb. 2021, doi: 10.1016/J.ENPOL.2020.112028.

[5] M. P. Blimpo, A. Postepska, and Y. Xu, “Why is household electricity uptake low in Sub-Saharan Africa?,” *World Development*, vol. 133, p. 105002, Sep. 2020, doi: 10.1016/J.WORLDDEV.2020.105002.

[6] O. Urgessa Ayana, J. Degaga, O. U. Ayana, and J. Degaga, “Effects of rural electrification on household welfare: a meta-regression analysis,” *International Review of Economics 2022*, pp. 1–53, Apr. 2022, doi: 10.1007/S12232-022-00391-7.

[7] IEA, “Africa Energy Outlook 2019,” Paris, 2019.

[8] M. GoR, “Energy Sector Capacity Development Strategy,” Nov. 2019.

[9] GOGLA, “Global Off-Grid Solar Market Report Semi-Annual Sales and Impact Data Public Report ,” 2020.

[10] EnDev, “Rwanda Off-grid Sector Status Report 2018,” Aug. 2019.

[11] C. Muchunku, K. Ulsrud, D. Palit, and W. Jonker-Klunne, “Diffusion of solar PV in East Africa: What can be learned from private sector delivery models?,” *Wiley Interdisciplinary Reviews: Energy and Environment*, vol. 7, no. 3, p. e282, May 2018, doi: 10.1002/WENE.282.

[12] I. Bisaga and P. Parikh, “To climb or not to climb? Investigating energy use behaviour among Solar Home System adopters through energy ladder and social practice lens,” Energy Research and Social Science, vol. 44, pp. 293–303, Oct. 2018, doi: 10.1016/j.erss.2018.05.019.

[13] V. Kizilcec and P. Parikh, “Solar Home Systems: A comprehensive literature review for Sub-Saharan Africa,” Energy for Sustainable Development, vol. 58. Elsevier B.V., pp. 78–89, Oct. 01, 2020. doi: 10.1016/j.esd.2020.07.010.

[14] R. Kennedy, S. Numminen, J. Sutherland, and J. Urpelainen, “Multilevel customer segmentation for off-grid solar in developing countries: Evidence from solar home systems in Rwanda and Kenya,” Energy, vol. 186, p. 115728, Nov. 2019, doi: 10.1016/j.energy.2019.07.058.

[15] J. A. Guajardo, “How Do Usage and Payment Behavior Interact in Rent-to-Own Business Models? Evidence from Developing Economies,” Production and Operations Management, vol. 28, no. 11, pp. 2808–2822, Nov. 2019, doi: 10.1111/POMS.13067.

[16] J. A. Guajardo, “Repayment performance for pay-as-you-go solar lamps,” Energy for Sustainable Development, vol. 63, pp. 78–85, Aug. 2021, doi: 10.1016/J.ESD.2021.06.001.

[17] N. Khaki, R. Borst, K. Kennedy, and M. Mattern, “PAYGo PERFORM: Financial, Operational, and Portfolio Quality KPIs for the PAYGo solar industry,” Washington, D.C., 2021.

[18] M. Jain, R. Gravesteijn, A. Jacobson, E. Gamble, and N. Scarrone, “Digital Finance for Energy Access in Uganda: Putting mobile money big data analytics to work,” Feb. 2020.

[19] A. G. Sotiriou, P. Bardouille, D. Waldron, and G. Vanzulli, “Strange Beasts: Making Sense of PAYGo Solar Business Models,” Washington, D.C, Jan. 2018.

[20] C. C. Aggarwal, Data Mining: The Textbook, 1st Edition. Springer, 2015. doi: 10.1007/978-3-319-14142-8.

[21] J. Liu and J. Han, “Spectral Clustering,” in Data Clustering, 1st Edition., Chapman and Hall/CRC, 2014, pp. 177–200. doi: 10.1201/9781315373515-8.

[22] M. G. R. Sause, A. Gribov, A. R. Unwin, and S. Horn, “Pattern recognition approach to identify natural clusters of acoustic emission signals,” Pattern Recognition Letters, vol. 33, no. 1, pp. 17–23, Jan. 2012, doi: 10.1016/J.PATREC.2011.09.018.

[23] M. Halkidi, M. Vazirgiannis, and C. Hennig, “Method-Independent Indices for Cluster Validation and Estimating the Number of Clusters,” in Handbook of Cluster Analysis, 1st Edition., Chapman and Hall/CRC, 2015, pp. 616–639. doi: 10.1201/B19706-35.

[24] S. M. Aarakit, J. M. Ntayi, F. Wasswa, M. S. Adaramola, and V. F. Ssennono, “Adoption of solar photovoltaic systems in households: Evidence from Uganda,” Journal of Cleaner Production, p. 129619, Nov. 2021, doi: 10.1016/J.JCLEPRO.2021.129619.