npj | clean energy

Article

https://doi.org/10.1038/s44406-025-00011-7

Unveiling and estimating behind-the-meter rooftop solar self-consumption using explainable AI

Check for updates

Mizue Shimomura[1] ✉, Alexander Ryota Keeley[1,2], Ken'ichi Matsumoto[3], Kenta Tanaka[4] & Shunsuke Managi[1,2] ✉

As renewable energy adoption grows, rooftop solar for self-consumption is also increasing. This "behind-the-meter" self-consumption is usually unmeasured, making grid operation difficult. However, few studies have estimated self-consumption. This study analyzes the impact of weather and proposes a framework for estimating self-consumption at grid using readily available data. Employing machine learning with XAI, the downward impact of solar radiation on grid demand is quantified to estimate self-consumption. Using actual Australia data, self-consumption was estimated under different solar adoption levels, demonstrating the high accuracy and versatility. Additionally, the increase in self-consumption in summer was quantified by the interaction between temperature and radiation. In summer, peak hourly self-consumption increases linearly by 30% for every 10 °C rise in daily maximum temperature due to cooling demand. As cooling demand is expected to grow, this finding has significant implications. This study is globally applicable and valuable for grid operators and policymakers integrating renewables.

The global transition toward decarbonization has accelerated the adoption of renewable energy sources[1]. Among these, solar power has experienced the most significant growth and is projected to continue expanding rapidly[2–4]. Although utility-scale solar farms have traditionally led this trend, the surge in electricity prices caused by the Russia-Ukraine crisis, along with falling costs of solar systems, has shifted attention toward distributed energy systems, particularly rooftop solar[5,6]. Here, "rooftop solar" refers to small-scale, behind-the-meter systems installed on residential and commercial properties, where the generated electricity is primarily consumed on-site by the property owner. In 2023, rooftop solar installations accounted for 45% of global solar capacity additions[7,8].

Policy initiatives have further accelerated this shift. Spain's National Energy and Climate Plan aims to install 19 GW of rooftop solar capacity by 2030[9], and China's Whole-Country PV Program supports the widespread deployment of rooftop solar systems[10,11]. Community solar projects—which share solar generation among multiple users—are also contributing to increased adoption[12], and the growth of solar installations on commercial buildings, factories, and building-integrated photovoltaics is contributing to greater on-site solar self-consumption[4,13,14].

Using solar generation for on-site consumption offers several benefits, such as reducing electricity costs, supporting local energy production and consumption, and mitigating renewable energy curtailment[8,15,16]. However, significant challenges remain for the expansion of rooftop solar. A major issue is the lack of visibility into "behind-the-meter" data. Rooftop solar generation is primarily consumed on-site by the owner, with surplus power exported to the grid. When solar generation is insufficient, electricity is purchased from the grid. The amount of solar electricity that is consumed on-site (hereafter referred to as self-consumption), as well as the rooftop solar system owner's gross demand and total solar generation, are typically located behind-the-meter and are usually not recorded in available data[17,18].

This lack of visibility complicates efforts to ensure grid reliability, stability, and safety[19]. In the Netherlands, ~25% of rooftop solar systems are unregistered, making them invisible to system operators and complicating net demand forecasting and voltage regulation at the distribution level[20]. In Western Australia, the grid operator AEMO has reported that rapid rooftop solar growth has increased the need for real-time flexibility, while limited visibility into behind-the-meter generation complicates load forecasting and raises concerns for system stability[18]. Similarly, in Spain, transmission and distribution system operators note that the absence of data from small-scale solar introduces uncertainty and risk into secure grid operation, potentially increasing the likelihood of involuntary disconnection events[21]. In California, the increasing penetration of rooftop solar has contributed to the

[1]Departments of Civil Engineering, Graduate School of Engineering, Kyushu University, Fukuoka, Japan. [2]Urban Institute, Kyushu University, Fukuoka, Japan. [3]Faculty of Economics, Toyo University, Bunkyo-ku, Tokyo, Japan. [4]Faculty of Economics, Musashi University, Nerima-ku, Tokyo, Japan.
✉e-mail: shimomura.mizue.443@s.kyushu-u.ac.jp; managi@doc.kyushu-u.ac.jp

'duck curve' phenomenon, where daytime net demand drops sharply. While rooftop solar mitigates curtailment by enabling on-site consumption during the day, it also leads to a steep rise in grid demand in the evening and during sudden weather changes as solar generation declines, creating a need for rapid ramping and greater system flexibility. Since behind-the-meter self-consumption is unobservable, accurately estimating the required flexibility remains challenging[22]. Digitalization and the use of smart meters could help address these issues. However, smart meters are not yet widely adopted in many countries[23], especially in developing regions, and investments in such infrastructure are costly and time-consuming[23,24]. There is a growing need for data-driven approaches using artificial intelligence (AI) to improve the observability of distributed energy resources.

Another major challenge is the lack of reliable data on rooftop solar installations, including their capacity, location, tilt, and other technical characteristics. Most utilities do not have detailed records of rooftop solar systems[19]. In South Africa, 47% of rooftop solar installations are unregistered with the grid operator[20,25]; in the Netherlands, 25% of household solar systems are unregistered[20]; and in Spain, despite substantial deployment of rooftop solar, there are no official records of these installations[26]. These data gaps present serious challenges for grid operators and aggregators[18,20,24,27].

As rooftop solar system increases, grid demand declines during the day but rises sharply in the evening when solar generation drops[28,29]. Integrating renewable energy into power systems is already complex[30], and increased self-consumption combined with limited data further complicates the balancing of supply and demand[19]. These issues not only affect short-term grid operations but also have important implications for long-term energy system planning, including decisions regarding power plant construction, fuel procurement, and transmission upgrades—all of which require significant lead time. Thus, quantifying self-consumption and understanding its impact on the power system are crucial for effective grid operation and planning.

Despite its importance, self-consumption is rarely estimated directly, as it depends on both weather conditions and user behavior and is therefore difficult to quantify. One existing study estimates self-consumption using a mathematical model based on installed rooftop solar capacity and weather data[31], while another uses regression analysis with actual self-consumption data to extrapolate to the regional level[32]. In both cases, data on installed solar capacity are required, and estimation accuracy is not addressed. Other research has focused on estimating behind-the-meter solar generation or gross demand. In these studies, because rooftop solar capacity is often unknown, researchers have devised alternative estimation methods to address this practical challenge. For example, using methods such as machine learning to analyze weather patterns[33], disaggregating solar generation from smart meter net load using neighboring demand patterns[34], or comparing demand and generation between communities with and without solar installations[35]. These approaches typically use actual solar generation or gross demand data from neighboring areas as training data for machine learning, making data availability a critical limitation. Another method estimates solar generation and gross demand for net-zero buildings by using location data and weather conditions to estimate the panel angle and area, based on grid demand data[36]. This method requires the exact location of the building and is less accurate when the demand is significantly greater than solar generation[36]. When installation data are unavailable, some studies have attempted to identify rooftop solar panels using satellite imagery[17], but these approaches still face limitations in data availability, computational complexity, and scalability for grid management.

In this study, we aim to identify the factors influencing self-consumption and to develop a framework for its estimation that can be applied across broad regions. Understanding the drivers of self-consumption is important for both decarbonization and stable electricity system operation. Self-consumption is determined by both solar generation and the electricity demand of the owner, which vary in complex and non-linear ways and are influenced by multiple factors. The limited availability of behind-the-meter data further complicates estimation.

We hypothesize that self-consumption can be estimated by quantifying the impact of solar radiation on grid demand. To test this, we use machine learning models capable of capturing non-linear relationships, combined with Shapley Additive Explanations (SHAP), an explainable AI (XAI) tool. We build a machine learning model to predict grid demand and then use SHAP to quantify the contribution of solar radiation. SHAP applies game theory to decompose machine learning models and attribute the contribution of each feature to the target variable[37]. In recent years, SHAP has been widely adopted in fields such as healthcare, architecture, and energy to interpret complex models in an understandable and actionable manner[38–40]. This framework relies only on easily accessible data, such as grid demand and meteorological information. It does not require detailed data on solar installation capacity or actual solar generation, which are typically difficult to obtain.

This study uses real-world data from Australia to analyze the factors that influence solar self-consumption. Our results show that self-consumption is highly dependent on solar radiation. This is because rooftop solar generation increases with solar radiation, while household electricity demand is less sensitive to it. By isolating the direct effect of radiation using SHAP, the estimated self-consumption closely matches observed values. We also observe a linear increase in self-consumption during summer as temperatures rise. For every 10 °C increase in daily maximum temperature, the peak hourly self-consumption increases by 30% when the solar adoption rate is 30%. This finding suggests increased cooling demand during heatwaves can make self-consumption play a key role in reducing peak grid demand. However, it also highlights the sharp increase in evening grid demand and during sudden weather changes, emphasizing the need for additional flexibility and grid management.

This study clarifies the characteristics of rooftop solar self-consumption, develops a framework for its estimation, quantifies the increase in self-consumption driven by cooling demand, and demonstrates the potential of SHAP for estimating unobserved data. As rooftop solar installations continue to expand globally, self-consumption will become increasingly important for grid operators seeking to maintain supply stability. The proposed framework provides a simple and effective method for estimating self-consumption, offering a scalable solution that can be applied in various contexts. With ongoing climate change and growing cooling demand, especially in emerging economies, the installation of rooftop solar systems and the on-site consumption of their electricity are expected to have a significant impact on the energy system[41]. This study represents an important step toward improving renewable energy integration into electricity systems, particularly in regions with increasing solar adoption.

Results

This study investigates the impact of weather factors on rooftop solar self-consumption at the grid level and develops a framework for estimating self-consumption using available data. The overall analysis flow is summarized in Fig. 1. We focus on distributed power systems with rooftop solar, where generated electricity is primarily consumed on-site, surplus power is exported to the grid, and the system owners purchase electricity from the grid when their rooftop solar output is insufficient. Importantly, rooftop solar self-consumption, total solar generation, and gross demand are typically not directly measured because they are behind-the-meter[18,24]. In contrast, grid demand—the electricity purchased from the grid—is generally available. Therefore this framework is designed as a solution to estimate self-consumption without using any data on self-consumption, gross demand, or rooftop solar capacity. The central hypothesis of this study is that rooftop solar self-consumption can be estimated by quantifying the effect of solar radiation on grid demand.

To test this hypothesis, we use machine learning and SHAP. First, as a pre-analysis, we examine the impact of meteorological variables on grid demand, gross demand, and self-consumption to confirm that this hypothesis can be tested. Next, we develop a model to estimate self-consumption. In the estimation, a machine learning model is constructed with the grid demand as the target variable. This model is then decomposed

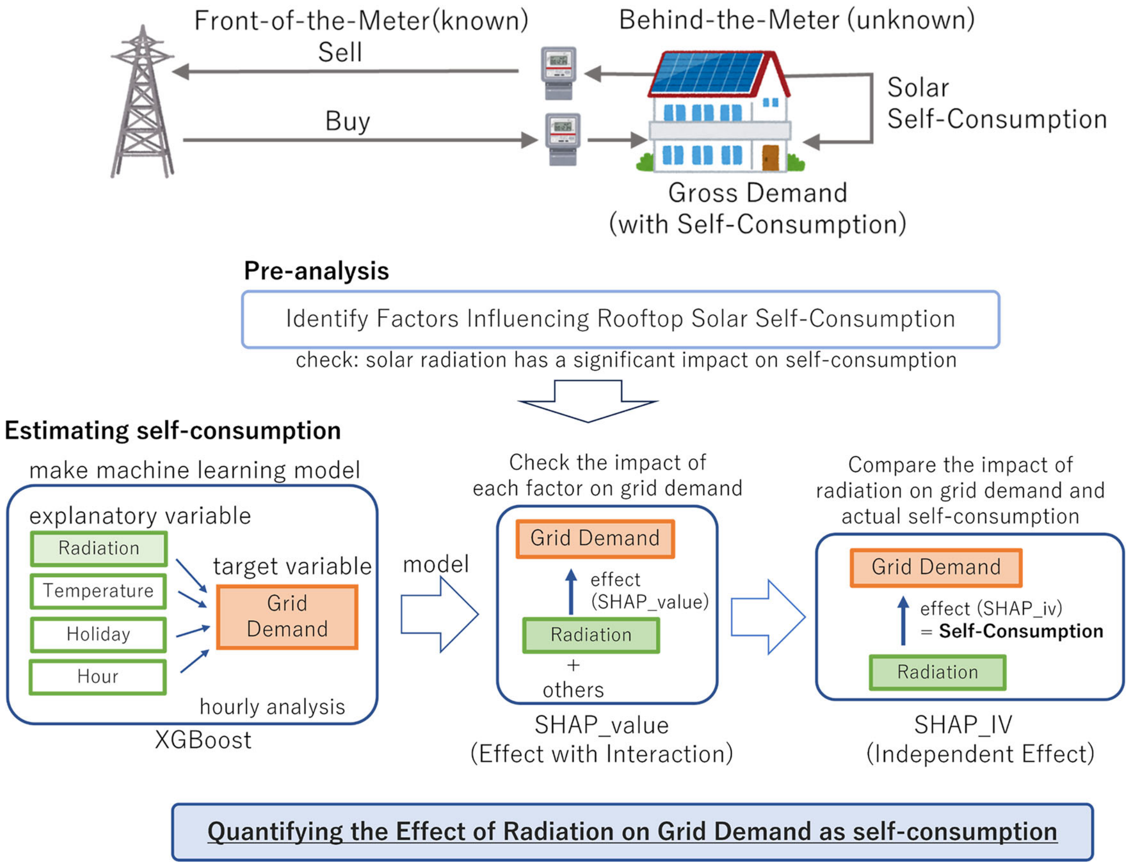

Fig. 1 | Framework of this study.

into its variable contributions using SHAP. Through this approach, the reduction in grid demand attributable to radiation is interpreted as self-consumption.

For this analysis, we use actual data from Australia. The dataset comprises three years of electricity demand (gross demand) and solar generation records from households with rooftop solar systems in the Sydney area[42]. To simulate grid behavior, we aggregate the household-level data to estimate grid-simulated demand and self-consumption. It is important to note that, although self-consumption data are included in this dataset, the use of such data is exceptional; in practical grid operation, actual self-consumption data are rarely available. In our analysis, self-consumption data are used only for factor analysis and accuracy validation, and are not used for model training. Furthermore, we do not use any information on rooftop solar systems, such as capacity, location, orientation, or efficiency, in either model training or self-consumption estimation because these data are often unknown.

Pre-analysis: Identifying the effects of weather on self-consumption and demand

As a prerequisite for our estimation framework, we first analyzed the effects of weather conditions on self-consumption and demand (gross demand and grid demand). This step establishes the basis for testing our central hypothesis. To identify the factors influencing each variable, we developed separate machine learning models with (a) gross demand (including self-consumption), (b) grid demand, and (c) self-consumption as the target variables. The independent variables used in all models were temperature, solar radiation (radiation), day type (weekday or holiday: holiday), and hour of the day (hour). We applied SHAP analysis to quantify the contribution of each input variable to the target. The results are summarized in Fig. 2.

As shown in Fig. 2a, temperature has the largest effect on gross demand, while solar radiation has a relatively small impact. In contrast, radiation is the primary driver of both grid demand (Fig. 2b) and self-consumption (Fig. 2c). The existing literature suggests that the impact of solar radiation on grid demand is small[43], possibly because the amount of rooftop solar installed is limited. In our analysis, as indicated in Fig. 2c, radiation accounts for most of the effect on self-consumption, while the effect of temperature is relatively minor. Figure 2d illustrates the relationships between gross demand, grid demand, self-consumption, and solar radiation. Since radiation has little impact on gross demand but a large impact on self-consumption, the observed reduction in grid demand directly reflects self-consumption. This result supports the idea that quantifying the impact of radiation on grid demand enables the estimation of self-consumption.

Figure 3 explores the effects of temperature and radiation on grid demand. In Fig. 3a, the effect of temperature, including its interactions, produces a U-shaped curve with a minimum around 20 °C. Grid demand increases at both low and high temperatures, as indicated by rising SHAP values, reflecting the sensitivity of electricity demand to temperature in regions with extensive cooling demand[43–45]. The increase in demand at higher temperatures is more pronounced than at lower temperatures. Because this dataset represents electricity demand in Sydney, Australia, it primarily captures high cooling demand during summer and relatively low heating demand during winter. The color scale in Fig. 3a represents radiation, showing that grid demand is lower when both temperature and radiation are high compared to when radiation is low.

Figure 3b shows the effect of radiation on grid demand, including interaction effects. The downward curve demonstrates that grid demand decreases as solar radiation increases. However, at very high temperatures, approximately above 30 °C, some data points deviate above the curve (indicated by blue arrows), suggesting that elevated cooling demand offsets the reduction in grid demand typically associated with higher solar radiation.

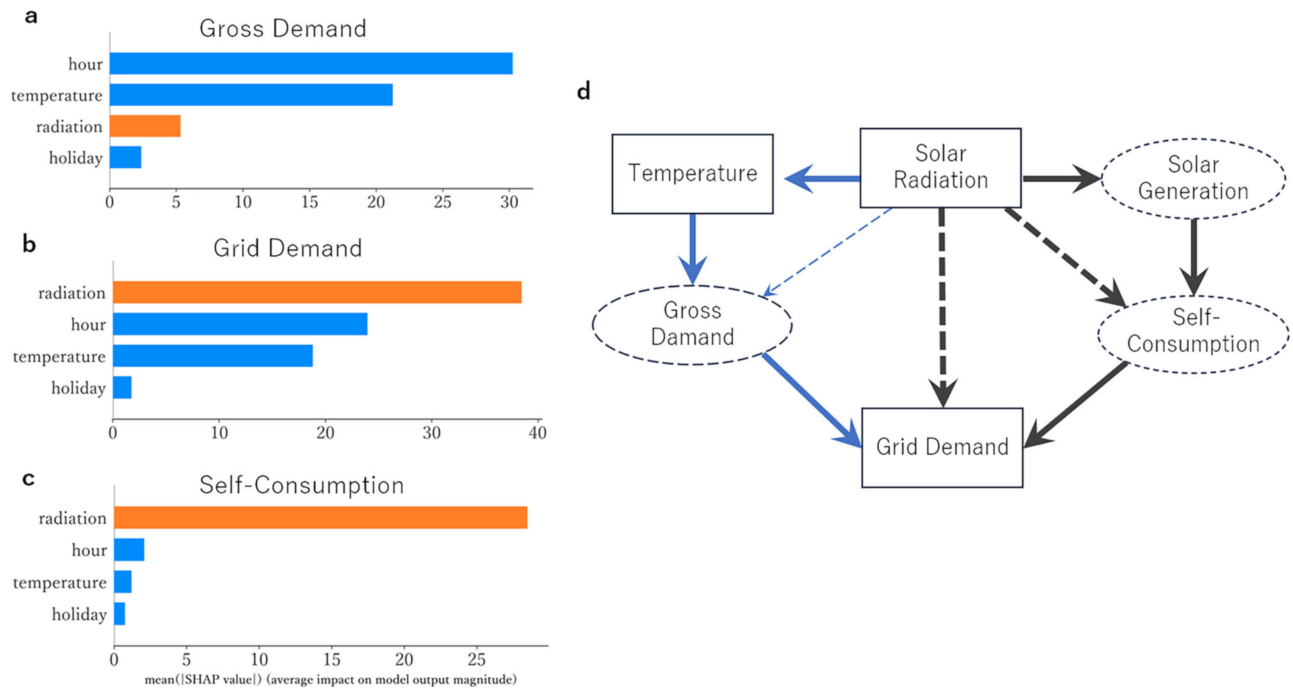

Fig. 2 | Impact of variables on gross demand, grid demand, and self-consumption (solar adoption rate 100%). **a**–**c** Feature importance determined by SHAP for models with **a** gross demand, **b** grid demand, and **c** self-consumption as the target variables. Orange bars represent the contribution of solar radiation. **d** Schematic diagram illustrating the relationships among radiation, temperature, gross demand, solar generation, self-consumption, and grid demand. Blue arrows indicate influences driven by temperature; black arrows indicate influences driven by radiation. Squares denote observable variables; dashed ellipses denote unobservable variables. Solid arrows represent direct influences, while dashed arrows represent indirect influences.

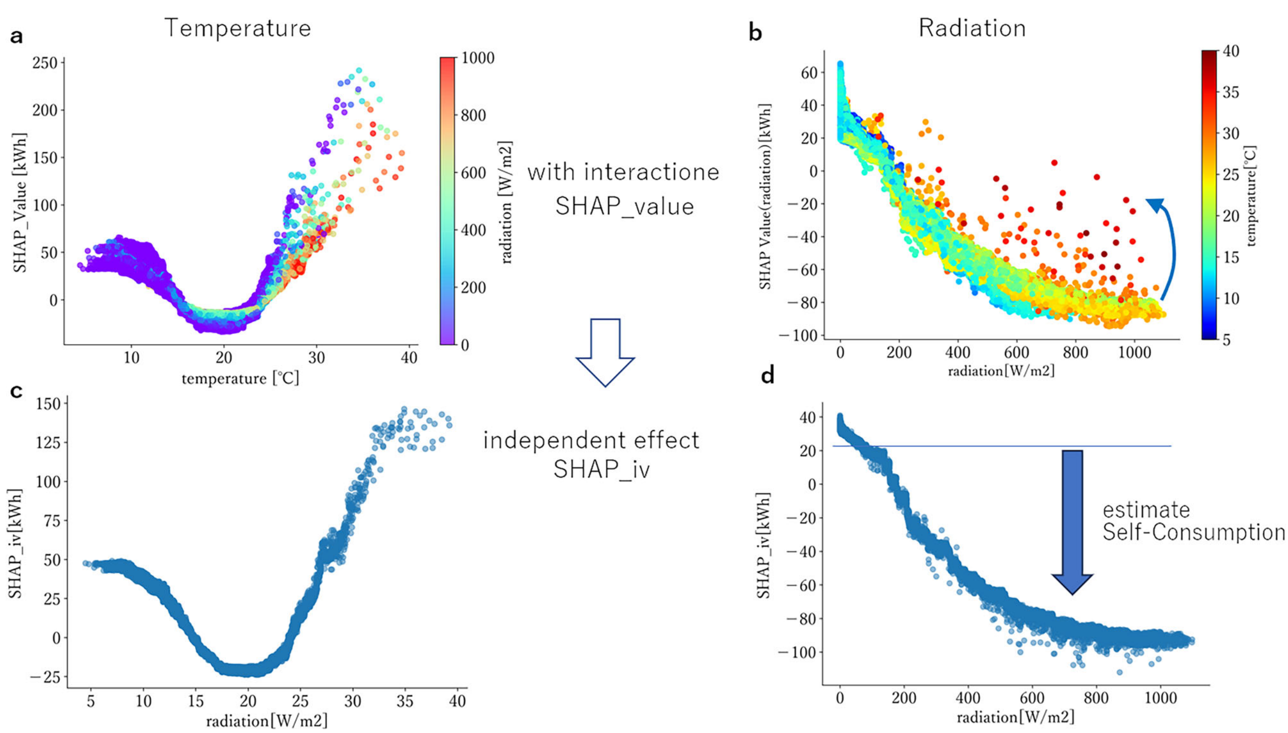

Fig. 3 | Effect of temperature and solar radiation on grid demand. Scatter plots show SHAP values and actual values for each variable. *x* axis: value of each variable. *y* axis: SHAP value of each variable (baseline y = 0 is the mean predicted value; positive indicates higher grid demand, negative indicates lower grid demand). **a**, **b** SHAP values including interaction effects; **c**, **d** individual SHAP values (no interaction). **a**, **c** *x* axis: temperature; *y* axis: SHAP or SHAP interaction value of temperature. **b**, **d** *x* axis: radiation; *y* axis: SHAP or SHAP interaction value of radiation. **c** color: radiation; **d** color: temperature.

Table. 1 | Evaluation metrics of self-consumption estimation.

Solar adoption rate	Result		
	Hourly (WAPE)	Daily maximum (MAPE)	Daily total (MAPE)
20%	0.39	0.32	0.45
30%	0.18	0.14	0.16
40%	0.22	0.23	0.20
50%	0.27	0.28	0.25
60%	0.29	0.29	0.27
70%	0.26	0.26	0.25
80%	0.22	0.19	0.20
90%	0.30	0.28	0.29
100%	0.29	0.26	0.28

Three-year average values from July 2010 to June 2013, calculated for the period from 9:00 a.m. to 5:00 p.m. Models were developed and evaluated separately for each year (July to the following June).

Figure 3c, d display the individual effects of temperature and radiation on grid demand based on SHAP interaction values (SHAP_iv). These lines are smoother compared to those in Fig. 3a, b, where interactions are included. Figure 3d illustrates the individual effect of radiation on grid demand: as radiation increases, grid demand decreases. At night or during periods with zero solar radiation, SHAP_iv is positive, indicating increased grid demand. Therefore, by using the SHAP_iv immediately before sunset as a baseline, the reduction in grid demand attributable to radiation can be interpreted as self-consumption.

Estimating self-consumption by quantifying the impact of radiation on grid demand (base estimation)

In the previous section, we demonstrated that the impact of solar radiation on grid demand can provide a basis for estimating rooftop solar self-consumption. Building on this finding, we now quantify the effect of solar radiation on grid demand and use it to estimate self-consumption at the simulated grid.

Our proposed framework is designed to be applicable to practical settings such as grid operation and policy planning, relying only on readily accessible data. Therefore, we do not use data from individual solar system owners—including gross demand, rooftop solar generation, or self-consumption itself—as training data, since these variables are generally unavailable. Similarly, detailed information on rooftop solar installations, such as system capacity, location, azimuth, panel tilt, or efficiency, is typically lacking in grid systems and is not used in our analysis.

For the estimation of self-consumption, we use the same household electricity data from Australia as in the previous section, which we aggregated to create a simulated grid-level dataset. In this analysis, self-consumption data were used only for accuracy evaluation and not for model training.

The estimation procedure is described in detail in the Method section and summarized here. First, we develop a machine learning model to predict grid demand using the same explanatory variables as in the previous section—hourly temperature, radiation, hour, and holiday. SHAP analysis is then applied to the trained model to isolate the effect of radiation, which is interpreted as the amount of self-consumption. This approach allows us to estimate self-consumption without requiring direct measurement of household demand or solar generation.

In Western Australia, for example, rooftop solar is installed in approximately one-third of households[18]. Based on this, our main analysis assumes a 30% solar adoption rate as the baseline scenario for self-consumption estimation. The objective of this study is to develop a framework for estimating rooftop solar self-consumption that is applicable across diverse regions with varying levels of solar adoption. To evaluate the generalizability of the framework under different conditions, we estimate self-consumption across a range of adoption rates, from 20%—representing lower penetration—up to 100%, to reflect possible future growth. For each scenario, solar-owning and non-owning households were randomly assigned from the dataset to match the designated adoption rate. (see Method section for details).

The solar adoption rate was not used in model training; it was applied only in scenario evaluation, so estimation was independent of the adoption rate. The individual effect of solar radiation on grid demand, measured by the SHAP interaction value for radiation (SHAP_iv_radiation), was interpreted as self-consumption. Estimation accuracy was evaluated by comparing predicted values with observed hourly values, daily maximum (hourly), and daily totals using weighted absolute percentage error (WAPE) and mean absolute percentage error (MAPE). Lower values indicate better estimation performance. The results are summarized in Table 1.

The self-consumption estimates varied slightly depending on the solar adoption rate. At a 20% adoption rate, the WAPE for hourly values reached 0.39, and the MAPE for daily totals was 0.45, indicating lower accuracy in this scenario. This is likely due to greater demand variability among non-solar households at low adoption rates. Estimation accuracy improved at adoption rates above 30%, with all metrics below 0.30. Residual analysis confirmed that the residuals showed no apparent bias and were approximately normally distributed. These results demonstrate that the proposed framework achieves high estimation accuracy under practical data constraints.

Although direct quantitative comparison with previous studies is not possible because prior work typically assumes known rooftop solar capacity and does not report accuracy metrics for self-consumption[31,32], papers estimating behind-the-meter rooftop solar generation typically report MAPE values around 30%[33,34,36], and our self-consumption estimation achieves similar or better accuracy. Importantly, our framework is designed for grid operation settings where accurate data on self-consumption and rooftop solar capacity are often unavailable. Unlike conventional machine learning or linear regression approaches that treat self-consumption as a target variable, our method is fundamentally different and novel, requiring no direct solar generation or capacity data while still delivering robust and interpretable results via SHAP.

Figure 4 compares actual and estimated self-consumption at a solar adoption rate of 30%. The estimates closely match the actual values overall. Figure 4a, b show that the hourly estimates align well with observations, even on days with unfavorable weather. However, the result tends to underestimate daily maximum self-consumption on extremely hot days in February 2011 and 2013 (Fig. 4d). This underestimation may be due to self-consumption being affected not only by solar radiation but also by temperature during extreme heat events. Therefore, we conducted additional estimations that included interactions between radiation and temperature during summer.

Verification of self-consumption on extremely hot days

In the previous section, self-consumption was estimated by isolating only the effect of solar radiation on grid demand. While this approach provided good accuracy under a range of conditions, the model tended to underestimate self-consumption during periods of extreme heat. This likely occurs because rooftop solar owners increase their use of cooling during very hot conditions, which raises their gross demand. Since gross demand serves as an upper bound for self-consumption (as shown in Fig. 1), increased cooling demand can lead to higher self-consumption. The previous estimation, which relied solely on radiation, did not capture this effect.

To address this issue, we incorporated the interaction between radiation and temperature into the estimation of self-consumption during summer. SHAP_iv was used to quantify both the individual effects and the interaction effects of these variables.

Figure 5 shows the SHAP_iv for the interaction between radiation and temperature. The interaction term is high during periods of extreme summer heat. In contrast, during January and February 2012, when there were

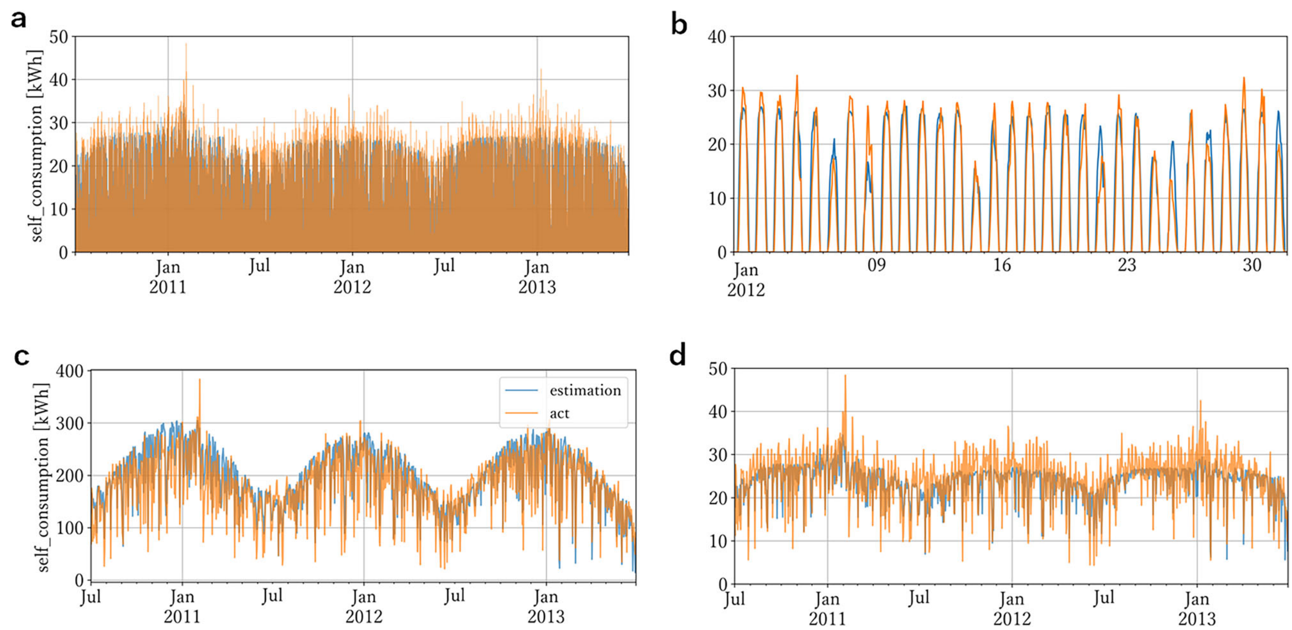

Fig. 4 | Comparison between actual and estimated self-consumption (rooftop solar adoption rate 30%) blue: estimate self-consumption, orange: actual self-consumption. a, **c**, **d** July 2010–June 2013, **b** January 2012, **a**, **b** hourly values, **c** daily totals, **d** daily maximum values.

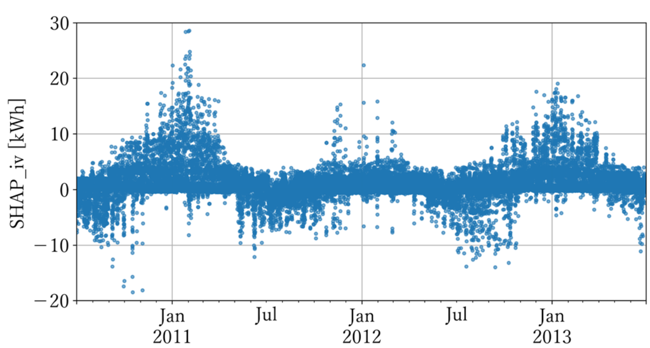

Fig. 5 | Interaction between temperature and radiation (solar adoption rate 30%) *x* axis: date(June 2010–July 2013), *y* axis: SHAP interaction value (SHAP_iv) for the interaction between temperature and radiation.

fewer extreme heat events, there was little increase in the radiation-temperature interaction. In summer, when cooling demand is substantial, higher solar radiation increases temperatures and results in elevated gross demand. Since self-consumption is limited by gross demand, an increase in gross demand due to higher cooling demand during periods of intense summer radiation results in increased self-consumption.

As shown in Fig. 3, the effect of temperature on grid demand generally follows a U-shaped pattern, and the independent effect of solar radiation decreases exponentially. However, Fig. 3b highlights an exception at high temperatures, where grid demand increases due to interactions between radiation and temperature, representing additional self-consumption during extreme summer heat. These interaction values were added to the baseline estimation based on the effect of radiation alone. Because winter heating demand is limited in Sydney, Australia, only summer interactions were included.

Figure 6 presents the self-consumption estimates with and without the interaction between radiation and temperature. As shown in Fig. 6a, b, incorporating this interaction improves the accuracy of self-consumption estimates during periods of extreme heat. Figure 6c illustrates the relationship between the daily maximum self-consumption and the daily maximum temperature, showing that self-consumption increases linearly when the temperature exceeds 25 °C. In Fig. 6c, the estimation accuracy for heatwave days (maximum temperature above 30 °C) improved from root mean squared error (RMSE) of 8.00 for the base estimation (based solely on radiation) to an RMSE of 4.76 when the interaction between temperature and radiation was included. During extreme summer heat, self-consumption rises above the base estimation as the interaction between solar radiation and temperature further increases self-consumption. In this case, for every 10 °C rise in daily maximum temperature, daily maximum self-consumption increases by ~30% at a solar adoption rate of 30%. This linear relationship is observed during the summer period and within the temperature range of the dataset, up to a maximum of 39 °C.

These results demonstrate that self-consumption can be estimated as the reduction in grid demand attributable to the downward effect of solar radiation. Furthermore, the increase in self-consumption during extreme heat can be quantified by incorporating the interaction between radiation and temperature. This additional increase reflects the effect of summer cooling demand and is expected to vary depending on the penetration of cooling systems.

Machine learning models are well-suited for capturing complex, non-linear relationships involving multiple factors within the range of the training data, but they are limited in their ability to extrapolate beyond observed data. Therefore, the results presented here are specific to the conditions observed in this dataset. However, by combining machine learning with SHAP analysis, our approach successfully captures the pattern of increased self-consumption during extreme heat based on grid demand behavior. This pattern-based approach may also be useful for projecting increases in self-consumption under future climate conditions or in other regions with different climate characteristics. As future work, it will be important to validate this framework using data from other regions and under different meteorological conditions.

Discussion

Rooftop solar systems, where the electricity generated is self-consumed by the owner, have seen rapid deployment in recent years. However, the actual self-consumption remains unobservable, as it is located behind-the-meter. In many cases, data on rooftop solar installations—including system capacity and location—are also unavailable or incomplete. This lack of data visibility makes it difficult to accurately assess system behavior, presenting a growing challenge for grid operation and planning.

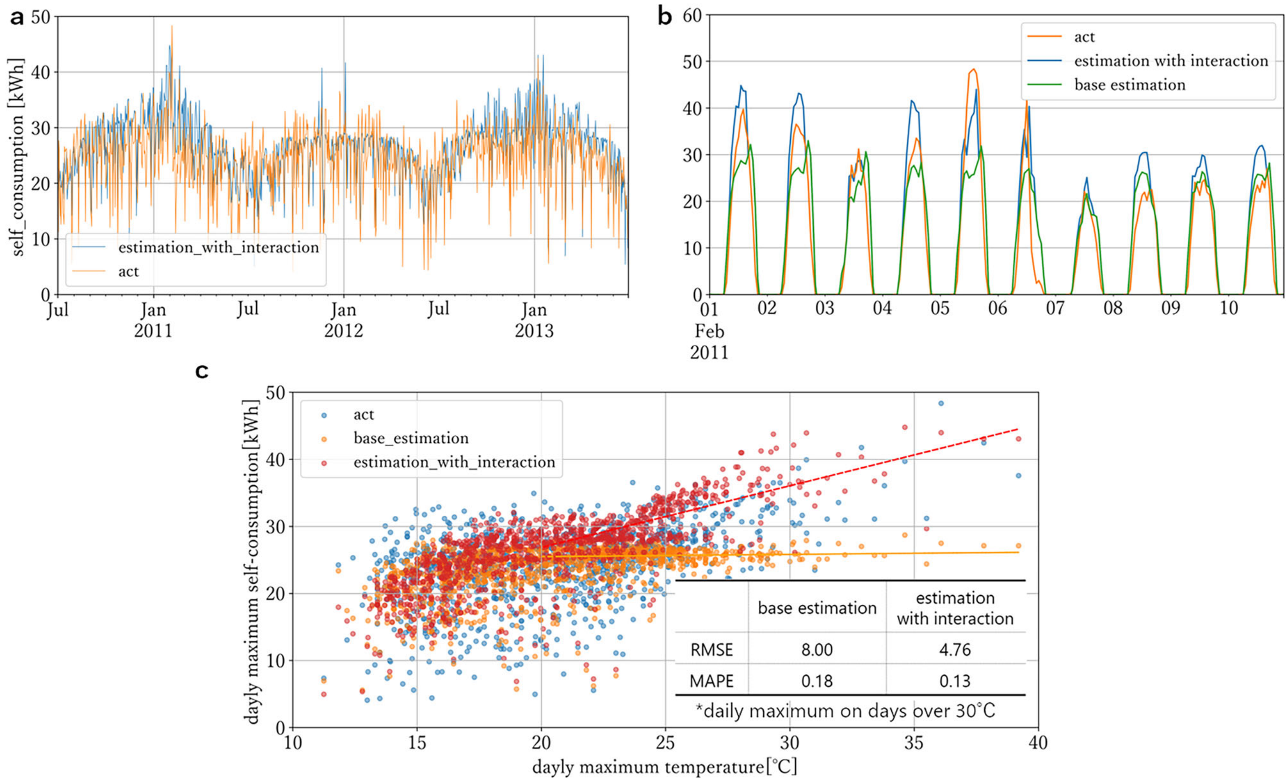

Fig. 6 | Self-consumption estimation results incorporating the interaction between temperature and solar radiation in summer (solar adoption rate 30%). **a** Daily maximum values, June 2010–July 2013; blue: estimated self-consumption with interaction, orange: actual self-consumption. **b** Hourly estimates, January 2013; blue: estimated self-consumption with interaction, green: base estimate (radiation only), orange: actual self-consumption. **c** Relationship between daily maximum temperature and daily maximum self-consumption; blue: actual self-consumption, orange: base estimate, red: estimate with interaction in summer. Evaluation metrics for daily maximum self-consumption are shown for days with daily maximum temperature over 30 °C (33 average, July 2010–June 2013, 9:00 a.m. to 5:00 p.m.).

In this study, we identified the impact of meteorological factors (temperature and solar radiation) on both electricity demand and rooftop solar self-consumption at the simulated grid level. We then proposed a framework to estimate self-consumption by quantifying the impact of solar radiation on grid demand using machine learning and SHAP, one of the XAI.

The proposed framework enables the estimation of rooftop solar self-consumption using only readily accessible variables, such as solar radiation, temperature, day type, and hour. By avoiding reliance on difficult-to-obtain data, this approach is widely applicable to practical grid operation and planning. Testing with real data from Australian households suggests that the framework is both effective and scalable across different solar adoption rates. Some limitations were observed, such as lower estimation accuracy for daily maximum self-consumption during periods of extreme heat.

To address the underestimation of self-consumption during periods of extreme heat, this study considered the interaction between solar radiation and temperature in the estimation framework. This addition improved the accuracy of summer self-consumption estimates. The analysis also revealed that peak hourly self-consumption increases linearly in summer—by ~30% for every 10 °C rise in daily maximum temperature—at a rooftop solar adoption rate of 30%, likely due to increased cooling demand. This linear trend was fitted within the observed temperature range of the dataset (up to 39 °C) and may not hold outside this range; the extent of this increase could also vary depending on the prevalence of cooling systems in different regions. Self-consumption is inherently limited by either the owner's electricity demand or by available solar generation. In regions with widespread adoption of cooling systems, extremely hot conditions can raise owners' electricity demand, effectively lifting the demand-side cap on self-consumption. Under such circumstances, the upper limit of self-consumption is determined by solar generation, and the rate of increase is expected to slow when cooling demand exceeds available solar generation. These results suggest that, on extremely hot days with increased cooling demand, the rise in self-consumption can help mitigate the increase in grid demand, as also indicated by previous studies[46–48]. Nevertheless, the sharp rise in evening grid demand following high daytime consumption, as well as during periods of sudden adverse weather, highlights the ongoing need for additional electricity supply and system flexibility. The framework presented here, by providing hourly estimates of on-site solar use, supports improved grid management.

While many studies have focused on the relationship between temperature and electricity demand[38–42], the growing adoption of rooftop solar requires consideration of the influence of solar radiation. As the share of distributed and variable renewable energy increases, the power sector is becoming more sensitive to weather and climate fluctuations[45,49]. This study shows that self-consumption in buildings equipped with a solar system is not small in terms of overall energy use. Existing studies have also described the importance of understanding self-consumption hourly[50]. Additionally, climate change causes dynamic changes in weather situations. Some studies have already pointed out that future climate change greatly impacts energy supply and demand[51,52]. Future adjustments to electricity demand and supply will become increasingly difficult due to the combined effects of climate change and the widespread adoption of solar power. To achieve a low-carbon society that is resilient to climate change, understanding self-consumption is essential for managing the overall electricity system. The approach of this study could provide a more accurate understanding of trends in self-consumption and support the realization of power supply configurations that use renewable energy more effectively.

The framework presented in this study simulates increased daytime self-consumption and the corresponding evening demand surge, providing valuable insights for grid operators and policymakers. For grid operators, this approach offers a practical method to estimate rooftop solar self-consumption using only readily available data, which can improve demand forecasting and operational planning, especially as solar penetration increases. For policymakers, the ability to quantify self-consumption without detailed system-level data supports the evaluation of solar adoption impacts and the design of effective support schemes. Our findings highlight that self-consumption increases during extreme heat due to rising cooling demand—an important insight for planning future electricity systems under climate change.

The data used in this study are from the Sydney area in Australia, where high cooling demand occurs on days with temperatures exceeding 39 °C. As climate change progresses, cooling demand is expected to rise globally, and rooftop solar is becoming increasingly prevalent, including in developing countries with hot climates[41,53]. It is essential to quantify the impact of increased cooling demand, self-consumption, and evening grid demand on grid stability. By contrast, our dataset showed minimal impact from heating demand in winter, indicating a need for further investigation in regions with high winter heating demand. Additionally, patterns of self-consumption may vary depending on the regional prevalence of cooling systems and the capacity of installed rooftop solar. Future research using data from other regions will be important for understanding how self-consumption responds to diverse climatic conditions, particularly under climate change scenarios.

This study established a framework for estimating behind-the-meter solar self-consumption using commonly available meteorological data and grid-level demand. Typically, estimation of self-consumption requires detailed data on solar capacity, panel efficiency, or gross demand, which are often unavailable[17,33,36]. Recently, meteorological reanalysis data from satellites have become widely accessible, even in developing countries where data constraints are most significant[54,55]. Our approach combines machine learning models with SHAP analysis to decompose the contribution of each variable, estimating the impact of solar radiation on grid demand reduction as self-consumption. While SHAP does not establish causality, it provides transparent and interpretable insights into model predictions and the relative importance of variables. The results of this study demonstrate that this method can be used to estimate unobserved variables in the power sector. The framework is not only applicable to estimating self-consumption, but can also serve as a general method for inferring unobservable variables. We anticipate that this approach will be widely used for analyzing hidden dynamics in a range of contexts.

The findings of this study regarding self-consumption, its relationship with weather, and its increase due to cooling demand, together with the proposed estimation framework, can help improve grid operation and inform policy for electricity systems with high shares of renewables. Rooftop solar self-consumption is expected to continue increasing, and the expansion of cooling demand due to climate change and economic development in emerging economies is also anticipated. This study contributes to revealing and managing behind-the-meter solar self-consumption and supports the integration of renewable energy into power systems.

Methods

Framework

In this study, the framework is developed to ensure applicability in practical settings, such as grid operation and policy planning. To reflect practical data constraints, only readily accessible information is used for analysis. This approach was chosen to address the typical data limitations associated with behind-the-meter consumption.

This study aims to estimate behind-the-meter self-consumption on the electricity grid using only easily accessible data. In behind-the-meter settings, not only self-consumption but also the owner's gross electricity demand (including self-consumption), rooftop solar generation, and the exact location and capacity of the solar systems are typically unobserved. Furthermore, self-consumption is influenced by both solar generation and the solar system owner's individual demand, and it exhibits non-linear characteristics. To address these challenges, we adopted a unique approach. The main hypothesis of this study is that rooftop solar self-consumption can be estimated by quantifying the effect of solar radiation on grid demand, which is readily available. To achieve this, we employed machine learning and used SHAP to quantify the influence of each factor within grid demand.

The overall analysis flow is summarized in Fig. 1. As a pre-analysis, we examine whether solar radiation has a strong influence on self-consumption and whether it has a limited or no effect on gross demand. These relationships are critical assumptions that underpin the core hypothesis. To assess this, we first construct machine learning models with gross demand, grid demand, and self-consumption as target variables, respectively. By applying SHAP to each model, we analyze the contribution of weather and temporal factors to each type of demand and confirm whether solar radiation can be isolated as the primary driver of self-consumption, without significantly affecting gross demand.

After verifying these assumptions, we proceed to test the main hypothesis of our framework for estimating self-consumption. Specifically, we build a machine learning model with grid demand as the target variable. The downward effect of solar radiation on grid demand, as identified through SHAP, is then interpreted as rooftop solar self-consumption. In this process, we do not use self-consumption data for model training; it is used only for accuracy evaluation.

Although the dataset used in this study does include self-consumption data, this represents an exceptional case. In power system operations, accurate self-consumption and rooftop solar capacity data are often unavailable. Consequently, in this study, we only use those data for exploratory feature analysis and for validating estimation accuracy—not as inputs for estimating self-consumption.

Data

This study used solar home electricity data provided by Ausgrid, an Australian distribution network service provider, to obtain household electricity usage and solar power generation data[33]. The dataset covers three years, from July 2010 to June 2013, for households with rooftop solar systems near Sydney, Australia. It includes measurements of household electricity demand (gross demand) and rooftop solar generation, recorded at 30-minute intervals.

The dataset does not contain detailed information about each rooftop solar system, such as panel tilt, azimuth, efficiency, or location. Although installed solar capacity is available in the dataset, we exclude capacity data from the analysis. This is because our study assumes a realistic setting in which solar capacity is unknown at the grid level. The dataset also lacks information on the presence of household battery systems. Since the data were collected between 2010 and 2013, when household batteries were not widely adopted, households are assumed not to have battery storage.

The dataset contained a substantial number of missing values. To balance data quality and sample size, only households with missing values for either demand or solar generation for an entire 24-hour period were excluded. This approach allowed us to retain as many households as possible for analysis. After this exclusion, 194 households remained in the final dataset.

To facilitate the analysis, the 30-minute interval data were aggregated into hourly data. For gross demand, hourly zero values were linearly interpolated. To simulate the grid, self-consumption, grid demand, and gross demand were aggregated across households, representing the demand and generation profile of a modeled grid composed of the target households. These values were calculated using Eq. 1–4.

$$self-consumption_{i,t} = \min(solar\_generation_{i,t}, gross\_demand_{i,t}) \quad (1)$$

$$self-consumption_{t} = \sum_{i} self-consumption_{i,t} \quad (2)$$

$$grid\ demand_{t} = \sum_{i} (gross\_demand_{i,t} - self-consumption_{i,t}) \quad (3)$$

Table. 2 | Descriptive statistics of the data (July 2010 to June 2013), solar adaptation rate 100%, hourly values

	Unit	Mean	Std	Min	Max
Radiation	W/m^2	190.14	264.10	0.00	1098.50
Temperature	°C	17.32	4.51	4.40	39.20
Solar_generation	kWh	46.50	68.61	0.00	260.66
Gross_demand	kWh	147.64	53.42	63.94	560.85
Grid_demand	kWh	121.46	69.23	6.92	518.45
Self-consumption	kWh	26.18	34.39	0.00	165.88

$$gross\ demand_t = \sum_i (gross\_demand_{i,t}) \quad (4)$$

where *self-consumption*$_{i,t}$, *solar_generation*$_{i,t}$, *gross_demand*$_{i,t}$ represent each household's rooftop solar self-consumption, rooftop solar generation, and actual electricity demand (including self-consumption), respectively. *self-consumption*$_t$, *grid demand*$_t$, *gross demand*$_t$ are defined as the aggregated self-consumption, aggregated grid demand (i.e., demand without self-consumption), and aggregated gross demand at time t. t denotes a time step of one hour. i is an individual household's values. Self-consumption is limited by the gross demand of each household, meaning that any solar generation exceeding gross demand is exported to the grid. This relationship is illustrated in Fig. 1. Grid demand is defined as gross demand minus self-consumption.

Meteorological data, including temperature, direct solar radiation, and diffuse solar radiation for Sydney, were obtained from the ERA5 satellite reanalysis dataset[54,56,57] using the Open Meteo API[58]. Historical meteorological data for Sydney were collected by specifying Sydney's latitude and longitude in the API. The ERA5 data generally align well with local meteorological observations, with only slight deviations observed during extreme heat events[59]. Global horizontal irradiance (GHI) was calculated from direct and diffuse solar radiation, and the GHI value was used as the measure of radiation in the analysis. A statistical summary of the dataset is presented in Table 2. Seasonally, Sydney experiences extremely high temperatures exceeding 39 °C in summer, while winter temperatures are mild, resulting in low heating demand (see Fig. 3a).

Structure of the machine learning model and SHAP interpretation

This study aims to estimate behind-the-meter self-consumption using only easily accessible data, specifically weather data and simulated grid demand. Because self-consumption is a non-linear outcome influenced by multiple factors, we employ machine learning and SHAP to capture and interpret these complex relationships. In this study, we use gradient boosted decision trees (GBDT) as the machine learning method. GBDT is an ensemble learning algorithm that integrates multiple decision trees using the gradient boosting technique. Each decision tree partitions the data based on input features, with nodes representing threshold-based splits[55]. The gradient boosting approach iteratively trains new trees to correct the prediction errors of previous ones, progressively improving model accuracy. GBDT is particularly suitable for modeling non-linear relationships, such as those observed in electricity demand[55]. The XGBoost package in Python was used to implement the GBDT model.

For the machine learning model, the explanatory variables used are temperature, solar radiation (radiation), weekday/weekend (holiday), and hour of day (hour). The selection of explanatory variables is intentionally minimal to ensure that only readily available data are used. Difficult-to-obtain data—such as solar panel installation details —are excluded. Similarly, data on neighboring solar generation and household gross demand, which have been used as training data in previous studies[33,35], are not included in this study. The general form of the machine learning model is shown in Eq. 5.

$$target\ value_t = f\left(radiation_t, temperature_t, hour_t, holiday_t\right) \quad (5)$$

where *radiation*, *temperature*, *hour*, *holiday* are solar radiation (GHI), temperature, hour, holiday, respectively. *target_value* is either gross_demand, grid_demand, or self-consumption for pre-analysis, and grid demand for estimating self-consumption. All of these variables are aggregated values of the target households used for grid simulation. t denotes a time step of one hour. For the machine learning training, 80% of the data was randomly selected as training data on a daily basis, and the remaining 20% was used for validation. Cross-validation and early stopping were employed to mitigate the risk of overfitting. In cross-validation, the data were divided into four segments in chronological order. One segment was used for validation, and the remaining three were used for training. This approach helps to prevent excessive fitting to specific subsets of data and improves the generalizability of the model. Additionally, the number of iterations was capped at 100 to further reduce the risk of overfitting. The analytical framework used in this study is illustrated in Fig. 1.

To quantify the contribution of each explanatory variable to the machine learning model, we employed SHAP, a method from XAI. SHAP, which was proposed by Lundberg and Lee[37], applies concepts from game theory to enhance the interpretability of machine learning models. It has been widely adopted in various research fields[38–40]. SHAP quantifies the contribution of each variable using the SHAP value. The average estimated value serves as the baseline (where SHAP value = 0). A positive SHAP value indicates an upward effect from the baseline, while a negative value indicates a downward effect on the target variable.

The relationship between estimated values and the contribution of each variable as determined by SHAP is shown in Eq. 6.

$$\begin{aligned} estimated\_target\_value_t &= baseline + radiation\_SHAP_t \\ &+ temperature\_SHAP_t + hour\_SHAP_t + holiday\_SHAP_t \end{aligned} \quad (6)$$

where *baseline*, *radiation_SHAP*$_t$, *temperature_SHAP*$_t$, *hour_SHAP*$_t$, *holiday_SHAP*$_t$ is the average estimated value, radiation's SHAP_value, temperature's SHAP_value, hour's SHAP_value, holiday's SHAP_value, respectively.

In SHAP quantification, the SHAP value includes interaction effects among variables, while the individual effects or the specific interactions between variables are represented as the SHAP interaction value (SHAP_iv)[37]. In this study, the SHAP value is used to analyze the impact of weather on self-consumption and demand, and the SHAP_iv is used to estimate self-consumption.

Method for pre-analysis: Identifying the effects of weather on self-consumption and demand

To investigate the impact of weather on gross demand, grid demand, and self-consumption, separate machine learning models were developed for each of these target variables. This analysis also serves as a prerequisite for testing the main hypothesis.

The dataset consisted of the 194 households described in the previous section, and the data were aggregated to simulate the behavior of a grid composed of these households. For this analysis, the solar adoption rate was set to 100%, meaning all households were assumed to have their own rooftop solar system. The explanatory variables used in all models are radiation, temperature, hour of the day, and holiday, as specified in Eq. 5. The machine learning models were constructed as described in the previous section.

All three years of data were combined and used for training to develop a comprehensive model for each target variable. SHAP was applied to interpret each model and to assess the importance of each feature in predicting the target variable. Feature importance—used to assess the relative contribution of each input variable—was analyzed to understand the influence of weather conditions, especially solar radiation, on gross demand,

Table. 3 | Descriptive daily statistics by solar adoption rate (July 2010 to June 2013): self-consumption ratio is defined as the mean gross_demand/self-consumption

Solar adoption rate	Self-consumption share	Gross_demand (kWh/day)		Grid demand (kWh/day)		Self-consumption (kWh/day)		Solar generation (kWh/day)	
		Mean	Max	Mean	Max	Mean	Max	Mean	Max
20%	3.5%	3543	8159	3420	7883	123	276	208	375
30%	5.1%	3543	8159	3362	7771	181	388	327	591
40%	7.0%	3543	8159	3294	7617	249	542	439	803
50%	8.5%	3543	8159	3242	7516	301	642	517	951
60%	10.6%	3543	8159	3168	7389	375	770	658	1207
70%	12.3%	3543	8159	3109	7258	434	901	737	1341
80%	13.6%	3543	8159	3062	7148	481	1010	834	1521
90%	16.0%	3543	8159	2975	6957	568	1201	1004	1835
100%	17.7%	3543	8159	2915	6832	628	1327	1116	2036

grid demand, and self-consumption. Additionally, SHAP values for grid demand were used to visualize the effects of temperature and radiation on grid demand, as well as their interaction. By analyzing these SHAP values and feature importances, we assess whether self-consumption can be estimated based on the effect of solar radiation on grid demand. This approach allows us to confirm that solar radiation is the primary driver of self-consumption and to establish a foundation for the estimation method.

Method for estimating self-consumption (Base estimation)

After confirming the prerequisite assumptions of the framework, which are that solar radiation has a strong influence on self-consumption and a limited effect on gross demand, we proceed to estimate self-consumption based on simulated grid demand. In this estimation step, we assume that actual self-consumption, gross demand, and rooftop solar information are unknown. Such data are typically difficult to obtain in grid operations, so they are not used as training data in the machine learning model.

In this model, self-consumption is estimated based on the reduction in grid demand attributable to solar radiation. Equations 6–10 show the estimating steps.

$$grid\_demand_t = f\left(radiation_t, temperature_t, hour_t, holiday_t\right) \tag{6}$$

$$\begin{aligned} lestimated\_grid\_demand_t &= baseline + radiation\_SHAP_t \\ &+ temperature\_SHAP_t + hour\_SHAP_t + holiday\_SHAP_t \end{aligned} \tag{7}$$

$$SHAP\_iv\_radition_t = radiation\_SHAP_t - \sum_{j \neq \text{radiation}} SHAP\_iv_{j,t} \tag{8}$$

$$grid\_demand\_baseline = \frac{1}{N}\sum_{\text{day}=1}^{N} \max_{17 \le t < 20}\left\{SHAP\_iv\_radition_t, |, radition_t > 0\right\} \tag{9}$$

$$estimated\_self-consumption_t = grid\_demand\_baseline - SHAP\_iv\_radition_t \tag{10}$$

where $SHAP\_iv\_radiation_t$ is radiation's SHAP_iv, j represents each explanatory variable. N denotes the total number of days to analyze.

To estimate self-consumption, we first developed a machine learning model with grid demand as the target variable (Eq. 6). This model was then decomposed using SHAP, which quantifies the contribution of each factor as a SHAP_value (Eq. 7). At this stage, the SHAP_value includes interactions among the different features. The independent effect of each factor, excluding interactions, is represented by the SHAP_iv (Eq. 8).

To establish a reference point, we use the average daily highest SHAP_iv for radiation from 17:00 to 20:00, which is the period when solar radiation is near zero but may still be present (Eq. 9). The reduction in grid demand attributable to radiation, as identified by SHAP_iv, is interpreted as self-consumption (Eq. 10).

SHAP_iv quantifies the isolated effect of each explanatory variable. For example, the influence of hour factors, such as the evening increase in electricity demand, is captured by SHAP_iv_hour. The effect of temperature is reflected in SHAP_iv_temperature. The SHAP_iv for radiation, SHAP_iv_radiation, captures the main effect of radiation itself, which appears predominantly during daylight hours and remains nearly constant at night.

In Western Australia, approximately one-third of households have rooftop solar systems[18]. Based on this fact, the study uses a 30% rooftop solar installation rate as the baseline scenario. To verify the generality of the framework, estimation is also conducted at rooftop solar adoption rates ranging from 20% to 100% in 10% increments, resulting in nine scenarios.

For each scenario, households with rooftop solar systems are randomly selected from the 194 valid households (out of 300 Australian households) described in the Data section to match the designated solar adoption rate. The remaining households are treated as non-solar households. For solar-owning households, grid demand and self-consumption are calculated from gross demand and solar generation using Eq. 1–3, while non-solar households contribute only gross demand data. The combined dataset is used to construct a synthetic grid that includes both solar and non-solar households for each adoption scenario. The machine learning models are trained without including the solar adoption rate as an input variable. Estimation is conducted independently, and the adoption rate is used only to generate reference values for evaluating estimation accuracy. A statistical summary of electricity data across different solar adoption rates is provided in Table 3.

The methodology for creating machine learning models and interpreting them is consistent with the previous section. Separate models are created for each year in order to account for annual variations. The year is defined as running from July to June of the following year, which results in three distinct models.

For self-consumption, it is important to evaluate not only the hourly values but also the daily maximum value, which is related to peak power, and the daily total value, which is associated with total power supply. Therefore, the hourly, daily maximum, and daily total self-consumption estimates are each compared with the actual self-consumption values. As comparison indicators, the MAPE is used for the daily maximum and daily total values. Since the minimum self-consumption is zero for hourly values and MAPE cannot be calculated in such cases, the WAPE is used for hourly evaluation. Both MAPE and WAPE express the deviation from the actual values as percentages.

Method for estimating self-consumption in summer

Additional analysis was conducted for summertime, when cooling demand may significantly affect self-consumption. This step was chosen to address potential underestimation during periods of extreme heat.

In regions where cooling systems are widely adopted, particularly hot summer days lead to increased cooling demand and higher electricity consumption. Because self-consumption is inherently constrained by demand, an increase in cooling demand leads to a corresponding increase in self-consumption (Fig. 1, Eq. 1). In the base estimation, self-consumption is derived from the reduction in grid demand attributable to solar radiation, quantified using SHAP_iv_radiation. However, this approach captures only the effect of radiation on demand and does not account for increases in gross demand caused by high temperatures resulting from cooling needs. As a result, estimating self-consumption based solely on the effect of radiation can lead to underestimation on extremely hot days when cooling demand is elevated.

To address this limitation, we incorporate the interaction between radiation and temperature (SHAP_iv_radiation × temperature) into the estimation of self-consumption during summer daytime hours, defined as 9:00 to 15:00 from December to February. This adjustment enables the model to reflect the increase in self-consumption driven by higher temperature-induced demand. The modified estimation method is presented in Eq. 11.

$$estimated\_summer\_self - consumption_t = estimated\_self - consumption_t + SHAP\_iv\_(radition \times temperature)_t \tag{11}$$

where $SHAP\_iv\_(radiation \times temperature)_t$ is SHAP_iv interaction between radiation and temperature.

To compare the increase in self-consumption during periods of extreme summer heat, we evaluated the accuracy of the base estimate and the estimate that incorporates the interaction between temperature and solar radiation by comparing daily maximum temperature and daily maximum self-consumption. This comparison was used to validate the effectiveness of the proposed method.

Data availability

In this paper, household electricity demand and rooftop solar generation were used from Ausgrid "Solar Home Electricity Data", downloaded through Australia NSW Government Data.NSW (https://data.nsw.gov.au/data/dataset/solar-home-electricty-data/resource/d2dc76f0-22e3-4efc-bed9-bb4e0e50f0db). For weather data, Temperature and solar radiation data were downloaded from open-meteo (https://open-meteo.com/).

Received: 11 May 2025; Accepted: 19 September 2025;
Published online: 31 October 2025

References

1. IEA. Renewable Energy Market Update - June 2023. *International Energy Agency*, *Paris, France* (2023).
2. Jäger-Waldau, A. Snapshot of photovoltaics-February 2024. *EPJ Photovolt.* **15**, 21 (2024).
3. IEA. Renewables 2024. *International Energy Agency*, *Paris, France* (2024).
4. Szabo, L. et al. Impacts of large-scale deployment of vertical bifacial photovoltaics on European electricity market dynamics. *Nat. Commun.* **15**, 6681 (2024).
5. IEA. *Renewables 2023*. www.iea.org. (2023).
6. SolarPower Europe. *Global Market Outlook For Solar Power 2023-2027*. www.solarpowereurope.org (2023).
7. IEA. Electricity 2024 - Analysis and forecast to 2026. *International Energy Agency*, *Paris, France* (2024).
8. van Ouwerkerk, J. et al. Quantifying benefits of renewable investments for German residential Prosumers in times of volatile energy markets. *Nat. Commun.* **15**, 8206 (2024).
9. Ministry for Ecological Transition and the Demographic Challenge (MITECO). Integrated national energy and climate plan. *Government of Spain* (2024).
10. Ye, Y. Are rooftop solar panels the answer to meeting China's challenging climate targets? *Nature* https://doi.org/10.1038/d41586-023-02991-x (2023)
11. Zhang, C., Yan, X. & Nie, J. Economic analysis of whole-county PV projects in China considering environmental benefits. *Sustain. Prod. Consum.* **40**, 516–531 (2023).
12. O'Shaughnessy, E., Barbose, G., Kannan, S. & Sumner, J. Evaluating community solar as a measure to promote equitable clean energy access. *Nat. Energy* **9**, 955–963 (2024).
13. Best, R. & Burke, P. J. Small-scale solar panel adoption by the non-residential sector: the effects of national and targeted policies in Australia. *Econ. Model.* **120**, 106164 (2023).
14. I. E. A. Technology Collaboration Programme by International Energy Agency Photovoltaic Power Systems Programme PVPS Task 1 Strategic PV Analysis and Outreach. *IEA PVPS* (2024).
15. Roldán Fernández, J. M., Burgos Payán, M. & Riquelme Santos, J. M. Profitability of household photovoltaic self-consumption in Spain. *J. Clean. Prod.* **279**, 123439 (2021).
16. Zhang, Z. et al. Worldwide rooftop photovoltaic electricity generation may mitigate global warming. *Nat. Clim. Change* **15**, 393–402 (2025).
17. Erdener, B. C., Feng, C., Doubleday, K., Florita, A. & Hodge, B.-M. A review of behind-the-meter solar forecasting. *Renew. Sustain. Energy Rev.* **160**, 112224 (2022).
18. AEMO. DER participation framework recommendations for policy and rule changes to encourage and facilitate participation of aggregated DER in the WEM. ARENA Work Package 7 Report. *Australian Energy Market Operator* (2023).
19. SIEMENS. Seeing behind the meter. *Oxford Economics* (2024).
20. Stedin. Kwart van de zonnepanelen niet in beeld. https://www.stedin.net/over-stedin/pers-en-media/persberichten/kwart-van-de-zonnepanelen-niet-in-beeld (2018).
21. redeia. *Annual_Accounts_2024*. https://www.redeia.com/sites/default/files/publication/2025/02/downloadable/Consolidated_Annual_Accounts_2024.pdf (2025).
22. California Public Advocates Office. *The Rooftop Solar Dilemma: Rising Electricity Rates and the Diminishing Value of Rooftop Solar*. (2024).
23. Steinbach, S. A. & Blaschke, M. J. How grid reinforcement costs differ by the income of electric vehicle users. *Nat. Commun.* **15**, 9674 (2024).
24. IEA. Digital tools will help keep distributed solar PV growing strongly. https://www.iea.org/commentaries/digital-tools-will-help-keep-distributed-solar-pv-growing-strongly (2023).
25. SALGA. Status of Embedded Generation in South African municipalities. *South African local government association* (2023).
26. appa. *2024 Informe Anual Del Autoconsumo Fotovoltaico*. *Spanish Renewable Energy Association* https://www.appa.es/wp-content/uploads/2025/02/Informe-Autoconsumo-Fotovoltaico-2024.pdf (2025).
27. Huxley, O. T. et al. The uncertainties involved in measuring national solar photovoltaic electricity generation. *Renew. Sustain. Energy Rev.* **156**, 112000 (2022).
28. Bowers, R., Fasching, E. & Antonio, K. As solar capacity grows, duck curves are getting deeper in California. EIA (2023).
29. Olczak, P. et al. Analyses of duck curve phenomena potential in polish PV prosumer households' installations. *Energy Rep.* **7**, 4609–4622 (2021).
30. IEA. Integrating Solar and Wind Global experience and emerging challenges. https://www.iea.org/reports/integrating-solar-and-wind (2024).
31. Garrido-Herrero, M., Jaramillo-Moran, M. A., Carmona-Fernandez, D. & Ozcariz-Arraiza, I. M. The impact of photovoltaic self-consumption on the daily electricity demand in Spain: Definition of a model to estimate it. *Heliyon* **10**, e32581 (2024).

32. McKenna, E., Pless, J. & Darby, S. J. Solar photovoltaic self-consumption in the UK residential sector: New estimates from a smart grid demonstration project. *Energy Policy* **118**, 482–491 (2018).
33. Li, K. et al. Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation. *Appl. Energy* **253**, 113595 (2019).
34. Cheung, C. M., Kuppannagari, S. R., Srivastava, A., Kannan, R. & Prasanna, V. K. Behind-the-meter solar generation disaggregation at varying aggregation levels using consumer mixture models. *IEEE Trans. Sustain. Comput.* **8**, 43–55 (2023).
35. Pylorof, D. & Garcia, H. E. Situational awareness-enhancing community-level load mapping with opportunistic machine learning. *Appl. Energy* **366**, 123291 (2024).
36. Chen, D. & Irwin, D. SunDance: Black-box behind-the-meter solar disaggregation. In *e-Energy 2017 - Proceedings of the 8th International Conference on Future Energy Systems,* 45–55 (Association for Computing Machinery, Inc, 2017).
37. Lundberg, S. M., Allen, P. G. & Lee, S.-I. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (2017).
38. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
39. Ruan, Y., Zhang, X., Zhang, M., Sun, F. & Chen, Q. Nonlinear and synergistic effects of demographic characteristics on urban polycentric structure using SHAP. *Nat. Sci. Rep.* **14**, 29861 (2024).
40. Shimomura, M., Keeley, A. R., Matsumoto, K., Tanaka, K. & Managi, S. Beyond the merit order effect: Impact of the rapid expansion of renewable energy on electricity market price. *Renew. Sustain. Energy Rev*. **189**, 114037 (2024).
41. IEA. World Energy Outlook 2024. *International Energy Agency*, *Paris, France* (2024).
42. Ausgrid. Ausgrid Solar Home Electricity Data. https://near.csiro.au/assets/42966a8f-bc3c-4bde-91d6-91bc5826aa21 (2021).
43. Gallo Cassarino, T., Sharp, E. & Barrett, M. The impact of social and weather drivers on the historical electricity demand in Europe. *Appl. Energy* **229**, 176–185 (2018).
44. Chen, S., Fang, X. & Khayatnezhad, M. Forecasting for electricity demand utilizing enhanced inception-V4 using improved Osprey optimization. *Sci. Rep.* **14**, 30832 (2024).
45. Staffell, I. & Pfenninger, S. The increasing impact of weather on electricity supply and demand. *Energy* **145**, 65–78 (2018).
46. Luthander, R., Widén, J., Nilsson, D. & Palm, J. Photovoltaic self-consumption in buildings: a review. *Appl. Energy* **142**, 80–94 (2015).
47. Fernández Bandera, C., Bastos Porsani, G. & Fernández-Vigil Iglesias, M. A demand side management approach to increase self-consumption in buildings. *Build Simul.* **16**, 317–335 (2023).
48. Calpa, M., Castillo-Cagigal, M., Matallanas, E., Caamaño-Martín, E. & Gutiérrez, Á. Effects of large-scale pv self-consumption on the aggregated consumption. *Procedia Comput. Sci.* **83**, 816–823 (Elsevier B.V., 2016).
49. Richardson, D., Pitman, A. J. & Ridder, N. N. Climate influence on compound solar and wind droughts in Australia. *NPJ Clim. Atmos. Sci.* **6**, 184 (2023).
50. Saez, R., Boer, D., Shobo, A. B. & Vallès, M. Self-consumption potential and surplus compensation policy impact on rooftop photovoltaic systems in Spain. *Renew. Energy* **229**, 120713 (2024).
51. Bartos, M. D. & Chester, M. V. Impacts of climate change on electric power supply in the Western United States. *Nat. Clim. Change* **5**, 748–752 (2015).
52. Solaun, K. & Cerdá, E. Climate change impacts on renewable energy generation. A review of quantitative projections. *Renew. Sustain. Energy Rev.* **116**, 109415 (2019).
53. Staffell, I., Pfenninger, S. & Johnson, N. A global model of hourly space heating and cooling demand at multiple spatial scales. *Nat. Energy* **8**, 1328–1344 (2023).
54. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
55. Gelaro, R. et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **30**, 5419–5454 (2017).
56. Mullainathan, S. & Spiess, J. Machine learning: an applied econometric approach. *J. Econ. Perspect.* **31**, 87–106 (2017).
57. Athey, S. & Imbens, G. W. Machine learning methods that economists should know about. *Annu. Rev. Econ.* **11**, 685–725 (2019).
58. Zippenfenig, P. Open-Meteo.com Weather API. Preprint at https://doi.org/10.5281/zenodo.14582479 (2023).
59. Su, C. H. et al. Preliminary assessment of regional moderate-resolution atmospheric reanalysis. *Australian Bureau of Meteorology* (2023).

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science 25K03322.

Author contributions

Design of the research: M.S., A.R.K., K.M., K.T., S.M., Performance of the research: M.S., A.R.K., Data collection and analysis: M.S. Writing of manuscript: M.S., A.R.K., K.M., K.T., S.M., Supervision: S.M.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Mizue Shimomura or Shunsuke Managi.