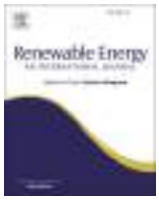
Evaluating uncertainty of shared energy in solar energy communities using a stochastic simulation framework

Renewable Energy 243 (2025) 122604

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene



Evaluating uncertainty of shared energy in solar energy communities using a stochastic simulation framework

F. De Bettin [a,b,*], F.D. Minuto [a,b], D.S. Schiera [a,b], A. Lanzini [a,b]

[a] Energy Department, Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129, Torino, Italy
[b] Energy Center Lab, Politecnico di Torino, Via Paolo Borsellino 38/16, 10129, Torino, Italy

ARTICLE INFO

Keywords:
Energy communities
Synthetic photovoltaic profiles
Stochastic simulation
Fokker-Planck
shared energy
Uncertainty evaluation

ABSTRACT

Properly sizing renewable energy sources is crucial for ensuring their techno-economic viability, especially under policies promoting solar power through photovoltaics (PV). The inherent variability of PV production requires assessing energy yield and self-consumption at different time scales, along with their associated uncertainties, to evaluate technical performance and financial risks. This challenge is critical for solar renewable energy communities (RECs), where energy sharing determines performance quality.

This work introduces a framework that quantifies the impact of PV's stochastic nature on energy sharing uncertainty in RECs. Tested across seven locations in Italy leveraging PVGIS data, the framework integrates path-integral and Fokker-Planck formalisms with a Monte Carlo approach, and is demonstrated to effectively capture production variability and energy yields.

For each location, 10,000 synthetic profiles were generated for a 50 kW peak power plant connected to the grid, serving 100 residential consumers with a typical consumption profile representative of the area. The relative uncertainty in yearly shared energy proved to range from 2 % to 3 %.

Comparisons with benchmark methods, like averaged hourly production (AHP) and typical meteorological year (TMY) profiles, revealed an systematic overestimation of shared energy during months of production surplus, underscoring the need of accounting for stochasticity in energy modeling.

1. Introduction

The importance of photovoltaic energy in Europe is growing, with projection indicating it will cover 15 % of its electrical demand by 2030 [1]. European member states are incentivizing rooftop photovoltaics through various legal paradigms, including the one of renewable energy communities (RECs). According to the Renewable Energy Directive (EU) 2018/2001, consumers and renewable energy prosumers can aggregate and establish and apply a collective framework to share renewable energy [2,3]. Prefeasibility studies are crucial for the photovoltaic installation process, as they assure stakeholders of the economic viability of their investments. Simulations of RECs rely on stochastic variables such as renewable energy production profiles, load profiles, and energy price profiles. However, the analysis and results are often not provided with a confidence interval.

REC studies are often approached as optimization problems with economic, energetic or environmental objective functions, aiming to retrofit existing buildings, as reported by Gjorgievski et al. [4], mostly using fixed production profiles and providing estimates without an associated uncertainty. As examples, we refer to the works by Sousa et al. [5], Jafari et al. [6] focusing on economic optimization functions; Cosic et al. [7] and Barone et al. [8] for both economic and environmental objective functions; Cielo et al. [9], Park et al. [10] and Weckesser et al. [11] for economic and energetic optimization strategies for technology sizing and energy distribution. Furthermore, Gomes et al. [12] published a software with the idea of providing a centralized and unified approach to REC analysis implementing different optimization algorithms for RECs, none of which provide a stochastic treatment of solar production. These approaches do not account for the inherent variability of the stochastic inputs, resulting in a lack of statistical evaluation of their variability. As a result, the long-term evaluation cannot be considered predictive due to the absence of associated statistical weights.

There are many software tools designed for PV planning [13], one of

* Corresponding author. Via Paolo Borsellino 38/16, 10129, Torino, Italy.
E-mail addresses: federico.debettin@polito.it (F. De Bettin), francesco.minuto@polito.it (F.D. Minuto), daniele.schiera@polito.it (D.S. Schiera), andrea.lanzini@polito.it (A. Lanzini).

https://doi.org/10.1016/j.renene.2025.122604
Received 17 August 2024; Received in revised form 15 January 2025; Accepted 2 February 2025
Available online 4 February 2025

the most famous being PVsyst. However, these tools do not include stochastic analysis and instead provide users either the typical meteorological year (TMY) or an hourly profile based on expected hourly production means. The TMY is constructed by assembling the most typical months from a time series dataset covering at least ten years of data, a procedure discussed by Huld et al. [14]. In contrast, the averaged hourly production profile (AHP) is composed of the hourly production means for each month, derived from a historical dataset. However, using these profiles as input data for REC simulation or optimization algorithms fails to capture the stochastic variability of solar production, thereby limiting the accuracy and robustness of the results. For these reasons, this work focuses on addressing extensively the preliminary step with respect to REC simulation and optimization, developing a methodology to choose PV generation inputs which have statistical significance.

While the introduction of stochasticity in self-consumption and energy sharing simulations has been explored in the literature [15], there has been less emphasis on modeling solar production. Existing models often face limitations, such as an imprecise treatment of solar production at fixed times, inadequate dynamics due to the lack of time correlation considerations, and the inability to generate a sufficient number of scenarios.

Zhang et al. [16] introduced stochasticity in the simulation, focusing mainly on demand profiles. Bashir et al. [17] included stochasticity in the wind and photovoltaic scenarios but did not account for time correlation when generating synthetic profiles, leading to potentially unrealistic synthetic time series. Kaplanis et al. [18], proposed a stochastic solar radiation model simplifying the problem by using one-mode probability distributions instead of a multi-modal one that would be necessary to fully capture the stochastic behavior of PV. Zhou et al. [19] used a Monte Carlo approach coupled with a MILP optimization algorithm to introduce solar production stochasticity; a similar approach was recently employed by Arun Rathore et al. [20,21]. Masoud Sharafi et al. [22] generated solar radiation scenarios by introducing Gaussian noise around the hourly mean. However, this approach does not adequately capture the complexity of the stochasticity of solar profiles. Georgios Mavromatidis et al. [23] generated multiple weather inputs reflecting possible future weather conditions and performed random sampling on such data, potentially losing time-correlation and limiting the number of profiles that can be generated by the dataset's size. Zheng et al. [24] added a stochastic component to wind and solar production profiles using the unscented transformation (UT) sampling method for day-ahead forecasting. Their model required weather forecast input variables and focused on short-term operational forecasting of the solar production, differing from our objective of long-term simulation for planning purposes.

On the other hand, since the 1970s in the field of photovoltaic generation, many works have been published with the objective of generating synthetic solar radiation time series, from which production profiles can be derived. Exell et al. [25] studied the probability distributions of daily radiations with a first-order Markov chain approach applied directly to historical radiation data from Thailand. They classified days into three categories, and obtained the probabilities of transitions between them. Liu et al. [26] recognized that while extra-terrestrial irradiance can be predicted accurately, it is the clearness index that encapsulates the stochasticity of solar radiation. In most of the literature the stochastic nature of solar radiation is indeed studied through the clearness index variable. Bartoli et al. [27] studied the clearness index behavior in different regions of Italy, finding that its time series is a first-order autoregressive stochastic process whose first autocorrelation coefficient is independent of locality. Aguiar et al. [28] used a one-step Markov Transition Matrix technique to generate solar radiation profiles, considering the clearness index as the stochastic variable. Graham et al. [29] provided an autoregressive time series method to generate hourly solar irradiation synthetic profiles using a statistical approach by modelling the clearness index. They also demonstrated that clearness index time series are non-stationary processes and that the correlation function depends on the type of day.

These studies provided valuable insights into the fundamental characteristics of solar radiation stochastic dynamics. However, they were conducted at a time when neither the amount of accessible data nor the computational power available today existed. In particular, the task of classifying day types proved to be problematic, since the number of day types is not easily defined in a formal way [30].

More recently, the field has advanced thanks to the publication of various solar radiation databases and the development of machine learning techniques [31]. Notable databases include NASA's MERRA and MERRA-2 [32], the ECMWF's ERA-Interim [33], the Japan Meteorological Agency's JRA-5 [34], the SARAH [35] and SARAH-2 [36] databases, which are used by PVGIS [37] to provide hourly surface solar radiation values. Pfenninger et al. [38], among others, validated the SARAH solar radiation database and provided correction factors for different geographical regions. Machine learning techniques have been used mainly to perform day type classification and regressions methods for short-term forecasting. Zhang et al. [39] reviewed recent long-term models for estimating and forecasting solar radiation over various timespans, such as monthly, daily and hourly. Mawloud Guermoui et al. [40] reviewed hybrid strategies for synthetic solar radiation data generation and forecasting, with timespan ranging from 15 min to three days. Similarly, Bazionis et al. [41], reviewed standard methods for short-term forecasting, typically ranging from 1 h to 1 day timespans. These reviews indicate that short-term and long-term forecasting require different strategies and serve distinct purposes.

Frimane et al. [30] used a Dirichlet process Gaussian Mixture Model on the clearness index to classify day types and generate synthetic or downscaled profiles starting from 10-min time series training data depending on the day type. This method avoids bias by not pre-selecting the number of day types. Munkhammar et al. [42], generated clear sky index profiles using a linear copula model [43], assuming the clearness index behaves as a stationary process, while Ramírez et al. [44] refined the model by using a non-linear approach.

In this work, as done in Ref. [44], we avoid the cumbersome problem of day type classification and we do not use a downscaling approach. Instead, we present a novel mathematical formulation of the problem, utilizing path-integral [45] and Fokker-Planck [46] formalization to generalize the Markov-Chain approach. Rather than using discrete transition matrices, we employ continuous functions as propagators to mimic the stochastic dynamics of solar radiation. The approach is similar to the non-linear copula model used in Ref. [44], but with the addition that the propagator acts as a time-dependent copula.

Unlike most literature, our focus is not on the clearness index or on the solar radiation itself for the study of the dynamics of the time series, but directly on the solar production values retrieved from the PVGIS software, with an emphasis on long-term forecasting. Specifically, we study the stochastic effect of PV production on the shared energy within a REC composed of 100 residential consumers and a single photovoltaic generation plant of 50 kW peak power in seven different regions of Italy. This configuration is chosen to isolate and emphasize the stochastic impact of PV production on the shared energy dynamics within the REC. While it is true that many RECs feature multiple, smaller generation units located on households' rooftops, incorporating such a distributed generation setup would necessitate modeling the stochasticity of both generation and consumption profiles. By focusing on a single, centralized PV plant, we eliminate the complexity introduced by consumption stochasticity. Furthermore, this simplified configuration serves as a baseline case. Analyzing such a scenario is essential for building a foundational understanding of how PV production variability propagates through a community's energy-sharing mechanism.

We generate 10,000 different solar production time series to study the monthly and yearly probability distributions of photovoltaic production, adding a novel contribution to the work of [39]. Furthermore, we investigate the characteristics of RECs' shared energy probability

distributions. Performing an energy analysis on a REC is straightforward once the historical production and load profiles are available. However, in the literature, these calculations are often performed without knowing their statistical significance.

This work aims to propose a novel methodological framework in which the output can be considered predictive. This framework provides the most probable outcome along with its respective 0.05 and 0.95 quantiles, thereby obtaining the monthly probability distributions of shared energy.

The organization of the paper is as follows: Section 2 provides a theoretical background on the mathematical tools and methodologies used. Section 3 details the methodology, including data sources, framework implementation, and validation procedures. Section 4 presents the results, highlighting the accuracy and reliability of the synthetic profiles. Section 5 discusses the implications of the findings and potential areas for future research. Finally, Section 6 concludes the paper by summarizing the key contributions and the benefits of the proposed framework.

2. Theoretical background

The mathematical tools used in this framework are the path-integral [45] and Fokker-Planck [46] formalisms. While commonly employed in the fields of Physics and Economics, these methods can also be applied to study any time series of a stochastic variable. The path-integral formalism requires a shift in perspective: instead of treating solar radiation or consumption as isolated stochastic variables, it considers the entire time series as a stochastic variable. In this section, these tools are presented in the context of calculating the shared energy in a REC composed of n consumers and a photovoltaic energy plant connected to the grid.

Denoting time as t, the production profile for the photovoltaic energy plant is defined as $p(t)$ and the consumption profile of the REC as $l(t)$. For a given period of time y, the community self-consumption (CSC) for the REC can be written as

$$CSC = \int_0^y dt\, min(p(t), l(t)). \tag{1}$$

Since in our case there is no physical self-consumption in the REC, CSC corresponds to its shared energy. The production and consumption profiles at a fixed time t, respectively $p(t)$ and $l(t)$, are considered to be distributed as a probability density function $f(t)$ and $g(t)$:

$$\begin{cases} p(t) \sim f(p,t), l(t) \sim g(l,t) \\ \int_0^{+\infty} dp f(p,t) = 1, \int_0^{+\infty} dl g(l,t) = 1, \\ \phi(p,t) = \int_0^p dx f(x,t) \in [0,1], \psi(l,t) = \int_0^l dx g(x,t) \in [0,1] \end{cases} \tag{2}$$

with $\phi(p,t)$ and $\psi(l,t)$ being the cumulative density distribution functions (CDFs) of the production and consumption profiles at time t, respectively.

To calculate the average of the community self-consumption in equation (*1*) over a time-lapse y,

$$\langle CSC \rangle = \langle \int_0^y dt\, min(p(t), l(t)) \rangle \tag{3}$$

means calculating the average over all the possible production and load profiles. Equation (*3*) can be written within the path integral formalism [45] as

$$\langle CSC \rangle = \begin{cases} \frac{1}{Z} \int D[p(t), l(t)] \omega(p(t), l(t)) \int_0^y dt\, min(p(t), l(t)) \\ Z = \int D[p(t), l(t)] \omega(p(t), l(t)) \end{cases} \tag{4}$$

where $\omega(p(t), l(t))$ is the probability of the profiles $p(t)$ and $l(t)$ to occur; the measure $D[p(t), l(t)]$ accounts for all of the possible time series $p(t)$ and $l(t)$.

In this formalism, the function of interest is the probability distribution of the profiles, $\omega(p(t), l(t))$. Since the production does not depend on the consumption, the probability can be written as:

$$\omega(p(t), l(t)) = \omega_p(p(t)) \omega_l(l(t)|p(t)), \tag{5}$$

with $\omega_p(p(t))$ the probability of the production profile $p(t)$ to occur and $\omega_l(l(t)|p(t))$ the conditional probability of the consumption profile $l(t)$ given a production profile $p(t)$.

In a REC, the aggregate load profile is the sum of individual consumption profiles. As the number of consumers increases, the relative variability of this aggregate profile tends to decrease, as suggested in Ref. [15]. Thus, the load profile of a REC with hundreds of consumers can be approximated as a time-dependent Gaussian distribution with variance at any time t significantly smaller than its mean. Given these hypotheses, the aggregate profile's distribution closely approximates a Dirac Delta function [47] $\omega_l(l(t)|p(t)) = \delta(l - l(t))$, zero everywhere except for when $l = l(t)$.function

In simulations of RECs, the TMY production profile is often used, as seen in studies like [48]. While this provides a meaningful estimate, it lacks statistical weight, making it impossible to determine the position of the results within their probability distribution or the width of that distribution. In contrast, the path integral method considers the full spectrum of possible profiles.

Equation (4) can be tackled numerically using a Monte Carlo approach: by generating synthetic profiles $p(t)$ and $l(t)$ for production and consumption, the argument of the path integral in equation (4), $\int_0^y dt\, min(p(t), l(t))$, can be calculated n times, yielding a probability distribution of the outcome. While this approach seemingly circumvents the need to derive $\omega(p(t), l(t))$ in equation (4), the challenge lies in generating consistent and meaningful time series of hourly productions and consumptions, which ultimately means mimicking the behavior of $\omega(p(t), l(t))$ nonetheless.

This work shifts the focus from the production variable p to its CDF, $\phi(p,t)$, in particular from the paths $p(t)$ to the paths $\phi(t)$. With the change of variables $p \rightarrow \phi(p,t) = \int_0^p dx f(x,t)$, the mean of self-consumption in equation (4) becomes:

$$\langle CSC \rangle = \frac{1}{Z} \int D[\phi(t), l(t)] \omega_\phi(\phi(t)) \omega_l(l(t)|\phi(t)) \int_0^y dt\, min\left(\phi^{-1}(\phi(t)), l(t)\right). \tag{6}$$

Starting from this treatment, the Fokker-Planck formalism [46] can be used to obtain a time-dependent propagator of the CDF. The propagator is the function that yields the probability of the CDF taking a certain value ϕ at time t, given that at the previous time t_0 it had value ϕ_0. The propagator contains all the stochastic dynamics information of the CDF, and is denoted as:

$$P(\phi, t|\phi_0, t_0). \tag{7}$$

Through the propagator of the CDF, multiple CDF time series can be generated and utilized to solve the integral in equation (6) with a Monte Carlo approach.

A well-known example of the use of the propagator and of the Fokker-Planck formalism is the mathematical description of Brownian motion [46], where a particle in a heat bath and subject to a possibly position-dependent potential $V(x)$ follows the Fokker-Planck equation:

$$\frac{\partial}{\partial t} P(x, t|x_0, 0) = -\frac{\partial}{\partial x}\left[\frac{\partial}{\partial x} V(x) P(x, t)\right] + D \frac{\partial^2}{\partial x^2} P(x, t|x_0, 0), \tag{8}$$

with $P(x, t|x_0, 0)$ being the particle's propagator, and D its diffusion coefficient. Depending on the shape of the potential and the boundary conditions, this equation may be solved analytically.

The idea introduced in this work is to treat the CDF as a particle undergoing Brownian motion in the confined space $[0,1]$ to obtain an analytical form of its propagator.

3. Methodology

The methodological framework proposed consists of a training and a generative step, relying on dependable historical data. We use PVGIS SARAH and SARAH-2 databases, corrected by a factor of 1.098 following Pfenninger et al. [38], as our examples are based in Italy. Utilizing PVGIS offers the advantage of directly incorporating photovoltaic production data. Specifically, the photovoltaic generation profiles obtained using PVGIS (version 5.1) with the SARAH database from year 2005–2016 are used for the training part, while the SARAH-2 time series from 2017 to 2020 are used for validation.

Specifically, we focus our attention on seven locations in Italy, reported in Table 1 and depicted in Fig. 1.

The model training, depicted in Fig. 2 and detailed in sections 3.1, 3.2 and 3.3, can be summarized in five high level steps.

1. Cluster historical data: Group historical data by month and hour, obtaining 288 clusters.
2. Fit marginal distributions: Fit the marginal distribution of each cluster with a multi-modal probability density function, obtaining 288 fits;
3. Calculate historical CDFs: Calculate the historical CDFs using the obtained fixed time marginal distributions with the definitions in equation (*2*), resulting in a set of 12 CDF yearly time series, $\phi(t)$, one for each year of the training data set;
4. Pair time-adjacent CDFs: Pair time-adjacent CDFs belonging to the same month (i.e., for a fixed hour h, forming a pair comprising the CDF value at hour h and the value at hour h+1), obtaining 12 CDF transition matrices, one for each month*;*
5. Fit CDF transition matrices: Fit the 12 CDF transition matrices with a solution to the Fokker-Planck equation. This function is the propagator of the CDF.

All fits are performed using the maximum likelihood method with the CERN Iminuit library [49].

The generative part, depicted in Fig. 3, involves generating n CDF Markov Chains to calculate n production time series. The approach follows a typical Monte Carlo simulation.

1. Initialize propagator parameters: Generate a number from a uniform distribution and use it to select the parameters of the propagator of the CDF;
2. Generate CDF Markov chains

Loop.

a. Generate random number from a uniform distribution
b. Use the generated random number to obtain a new CDF value by calculating the inverse of the integral of the propagator up to that number.
c. Add the value to the CDF time series.
d. Update the propagator parameters based on the last generated CDF value.

3. Generate photovoltaic production profile: Use the CDF time series within the photovoltaic production marginal distributions to obtain the synthetic PV production profiles.

Table 1
Locations of the studied energy plants.

Name	kWp	Latitude	Longitude	Province
NORTH	10	45.67541	12.275577	Treviso
CENTER NORTH	15	43.443115	11.885566	Arezzo
CENTER SOUTH	15	42.344274	13.405452	L'aquila
SOUTH	5	40.836071	17.35847	Brindisi
SICILY	40	37.897836	13.560399	Palermo
CAL	15	39.364244	16.03256	Cosenza
SARD	10	39.25636	8.947062	Cagliari

3.1. Fixed time production PDFs calculation

Once the characteristics of the photovoltaic energy plant, including latitude and longitude, are determined, hourly production data spanning from 2005 to 2016 is obtained from PVGIS and organized into monthly and hourly clusters. Each cluster contains around 360 data points. Subsequently, this data is fitted for each hour, resulting in 288 curves per energy plant, obtaining the distribution $f(p,t)$ cited in equation *(2)*. For this fitting, we use a trimodal distribution as the prior. Such distribution should take into account that at a given hour the PV plant may not produce; that productions may only take positive values; that more than one peak may be present, as shown for instance by Ref. [44].Specifically, the function is chosen to be

$$f(p,t) = p_0(t)\delta(p) + p_\Gamma(t)\Gamma(p,t) + (1 - p_0(t) - p_\Gamma(t))N(p,t). \quad (9)$$

In equation (9), $\Gamma\ (p,t)$ represents the gamma distribution [50], which models non-negative values, accounting for the first peak of the distribution; $N(p,t)$ the normal distribution, accounting for the higher peak of the distribution; $\delta(p)$ the Dirac Delta distribution [47], which is non-zero only when $p\ = 0$. The Dirac Delta distribution represents the probability of zero production, capturing instances when the energy plant is inactive. Unlike standard probability distributions, the Dirac Delta lacks continuity in its CDF because of its degeneracy. Both p_0 and p_Γ are positive, and their sum is ≤ 1, making the function a valid probability distribution. The parameter p_0 is directly derived from the raw data points as the ratio of zero-production data points to the number of points in the cluster. Clusters with fewer than 10 non-zero data points are considered to have $p_0 = 1$ and to follow a pure Dirac Delta function.

To test the quality of the fits, the p-value is calculated for each fit using the Kolmogorov-Smirnov test, with a threshold p-value of 0.05.

3.2. Propagator Visualization

Within this framework, the weather's behavior is entirely encapsulated in the CDF derived from the marginal distributions $f(p,t)$ in equation (9), calculated with equation (2). High values of the CDF (close to 1) indicate a high clearness index, whereas low CDF values indicate a low clearness index. For example, if the production p takes the maximum value the energy plant may produce at a time t, the corresponding CDF $\phi(p,t)\ = 1$. During nighttime, when the marginal distribution $f(p,t)$ is a pure Dirac Delta, the CDF cannot be determined due to the degeneracy of the function. This approach remains agnostic with respect to the peak power of the plant and the hour of the day, as that information is encoded within the parameters of the fixed time probability density functions $f(p,t)$.

The historical CDFs are calculated to obtain an analytical form of the propagator in equation (7). For each specific location, CDFs are organized in pairs composed of the CDF value at hour h and the value at hour h+1, and such pairs are visualized in the heatmap in Fig. 4. Brighter colors indicate a higher frequency of appearance of the pair. The frequency peaks are mostly in the diagonal of the matrix and display a width, indicating that weather changes do not typically occur abruptly within short periods and tend to persist. At the bottom left, there is an off-diagonal peak, suggesting a different behavior compared to the rest of the transition matrix's domain. This organization of the data enables to quantify the probability of a CDF taking value ϕ at hour h+1 given that at hour h it had value ϕ_0. This is the definition given to the propagator in equation (7), $P(\phi,t|\phi_0,t_0)$, which encapsulates the stochastic

Fig. 1. Locations of the studied photovoltaic energy plants.

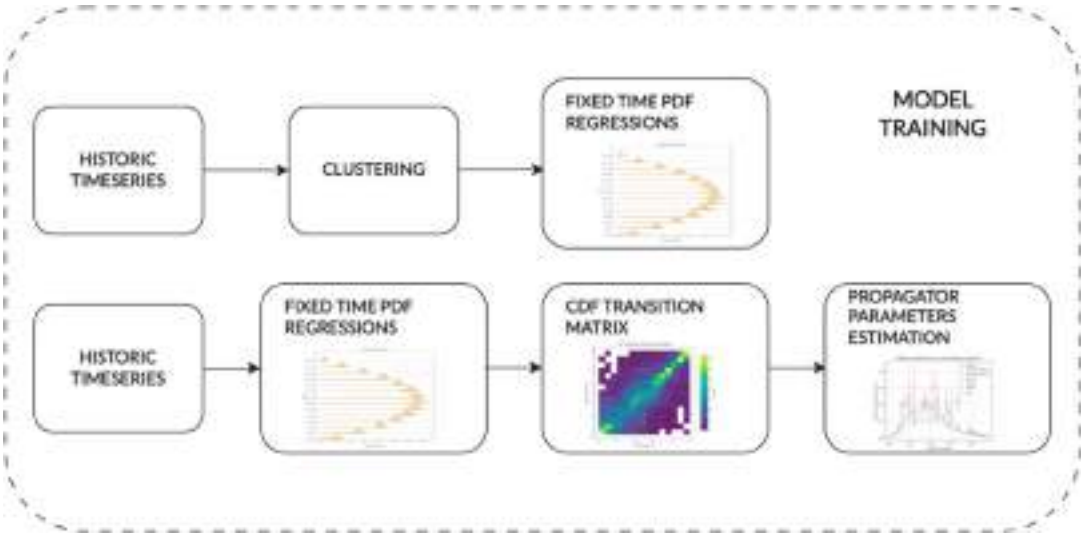


Fig. 2. Training procedure figurative description. Historic data is clustered by month and hour to obtain fixed-time marginal distributions. The same data is then used to calculate its corresponding CDFs with respect to the previously calculated marginal distributions. The CDFs are clustered by month and organized in CDF transition matrices, and fit with a specific solution of the Fokker-Planck equation.

dynamics of the production CDF.

The computed CDFs are clustered by month as shown in Fig. 5 (NORTH location), yielding 12 different transition matrices. Each month contains a different quantity of data points due to the variation in number of hours with solar radiation. White squares indicate the absence of data in the corresponding intervals. The shape of the matrix changes month by month, indicating different weather patterns. The same organization of the data is shown in section 3 of the supplementary materials for all other studied locations.

Fig. 6 shows two projections of the matrix in Fig. 4 (SOUTH location) at fixed initial CDF. The plots show the probability distributions of the CDF at time $h+1$ given that its value at hour h is 0.12 (Fig. 6-left) and 0.92 (Fig. 6 – Right). The left plot reveals two peaks, while the right plot shows a single, well-defined peak. Both plots exhibit long tails, proving that it is unsuitable to fit the data with a normal or a beta distribution.

More visualizations of the propagator are reported in section 1 of the supplementary materials.

3.3. Propagator modeling

The CDF is modelled as a particle confined in a finite one-dimensional space [0,1] undergoing Brownian Motion, as suggested in

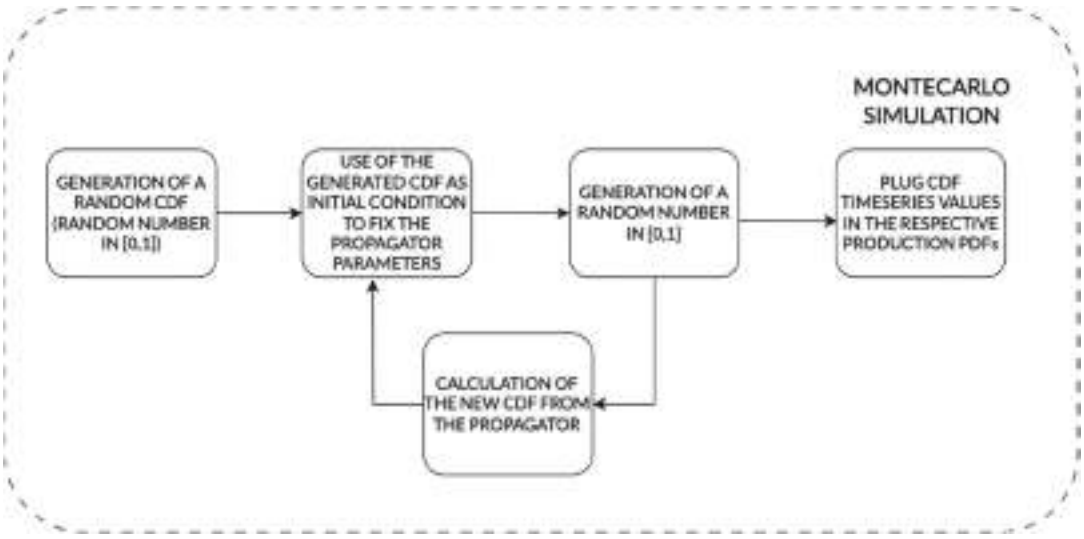


Fig. 3. Synthetic profile generation workflow.

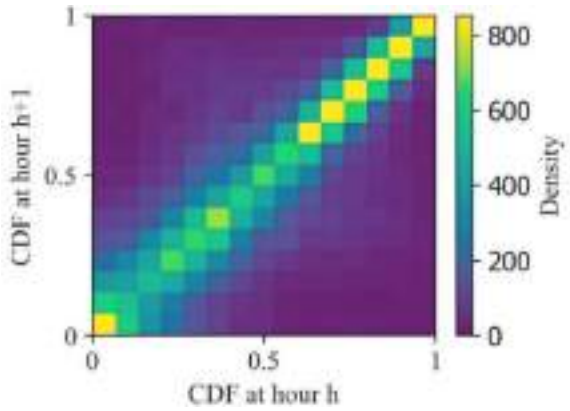


Fig. 4. Transition matrix considering data from all months for the NORTH PV plant.

section 2. This behavior is indicated by the tendency of the transition matrix peaks in to be on the diagonal, resembling a particle with inertia.

The propagator of a free particle undergoing Brownian motion follows the Fokker-Planck equation (8) reported in section 2. Since the CDF's domain is $[0, 1]$, an infinite well potential $V(\phi)$ is used, where $V(\phi) = 0$ for $\phi \in [0, 1]$. The solution of this equation, with the initial condition:

$$P(\phi, 0|\phi_0, 0) = \delta(\phi - \phi_0), \tag{10}$$

with $\delta(\phi - \phi_0)$ the Dirac Delta function [47] equal to zero everywhere except when $\phi = \phi_0$, can be written as [51]:

$$P(\phi, t|\phi_0, 0) = \frac{1}{z(D,t)} \left(\sum_{n=0}^{+\infty} e^{-n^2\pi^2 Dt} (\sin(n\pi\phi_0)\sin(n\pi\phi) + k\cos(n\pi\phi_0)\cos(n\pi\phi)) \right) \tag{11}$$

with $z(D, t)$ the normalization factor. The infinite sum ensures the condition in equation *(10)* is satisfied [52]. D determines the width of the peak at a given time, and k is a factor introduced to account for the height of the long tails of the distribution. The challenge is that the diffusion coefficient, which determines the width of the peak at a given time, is position-dependent. Additionally, the presence of $k > 0$ implies that $P(\phi, t|\phi_0, 0) \neq 0$ at the boundaries, which breaks the usual initial conditions imposed for the problem of a particle in a box [53].

To address these issues, fits are performed by fixing the value of the initial CDF and keeping only the pairs on the transition matrix that contain this initial condition. For each slice of data, a maximum likelihood fit with the prior in equation *(11)* is performed to estimate D and k. This way, the diffusion coefficient depends on the initial condition.

3.4. Synthetic data generation and validation

To generate Markov chains of the CDFs, the procedure detailed in Fig. 3 is followed by exploiting the obtained propagators. Once the CDF time series are generated, they are plugged into their respective production hourly marginal distributions to calculate their inverse CDF, thereby obtaining the final production time series.

For validation, The simulated profiles are first compared to the time series in the training dataset, and then to those in the test set. Following [54], the standard deviation of increments (SDI) is chosen as a variability metric to compare the simulated and real data:

$$SDI = \sqrt{\frac{\sum_{h=0}^{N-1} \left(|\Delta p_h| - \frac{1}{N} \sum_{h=0}^{N-1} |\Delta p_h| \right)^2}{N-1}}, \tag{12}$$

where N is the number of hours in the month; $\Delta p_h = p_{h+1} - p_h$ is the difference between the production at hour $h+1$ and the production at hour h. For each simulated monthly profile, the SDI is calculated and compared to the corresponding SDI of real data using:

$$SDI_{ratio}(timespan) = \frac{SDI_{sim}(timespan)}{SDI_{real}(timespan)}, \tag{13}$$

with the variable timespan being a month. The same test is performed

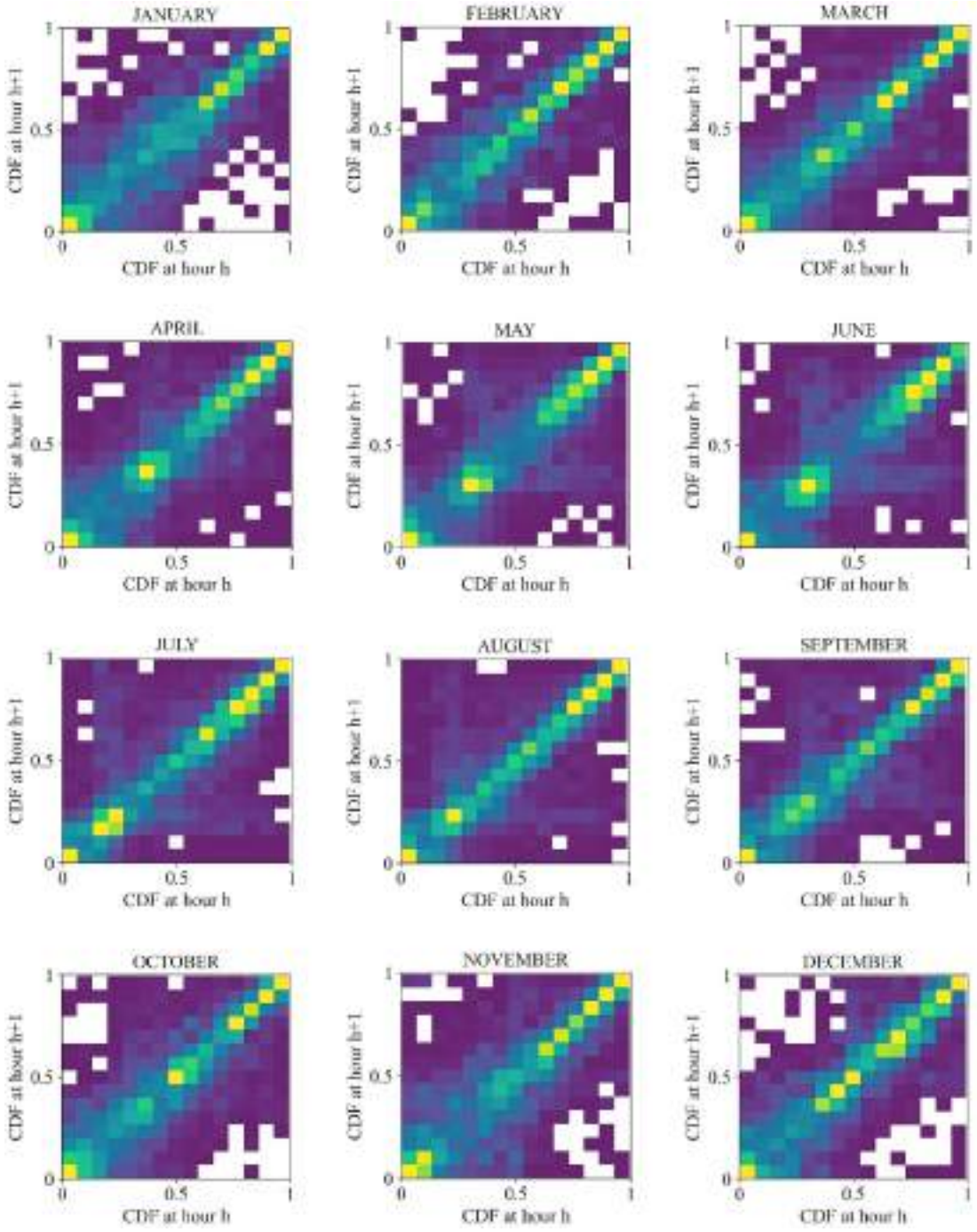


Fig. 5. CDF transition matrices of the NORTH PV plant built with monthly clustered data.

with a timespan of one year. With 10,000 simulations and comparisons to 12 years of training data (2005–2016) in the first step and 4 years of validation data (2017–2020) in the second step, the analysis generates 1,440,000 and 480,000 data points, respectively.

Additionally, the monthly and yearly production of the simulated profiles is compared to real data. For each simulated hourly profile, the monthly and yearly productions are calculated and divided by the corresponding values from PVGIS data, yielding 40,000 data points per test.

3.5. *Shared energy probability distribution function estimation*

Synthetic production profiles are utilized to estimate the shared energy in seven RECs located in the provinces listed in Table 1 depicted in Fig. 1. Each REC case study is composed of 100 residential consumers and a 50 kW peak power photovoltaic energy plant connected to the grid. This simple configuration decouples consumption and production stochasticity, isolating the effects of PV production variability on shared energy while maintaining generality in modeling PV plants. Indeed, given the localized nature of RECs, it is reasonable to assume that all PV plants experience nearly identical weather conditions, resulting in highly correlated production profiles. Consequently, the same CDF time series generated with the framework described in Fig. 3 can be applied within the marginal distributions of all plants to derive perfectly correlated synthetic production profiles through inverse CDF sampling.

The REC demand profiles are generated using the standard normalized consumption profiles for residential consumers provided by the

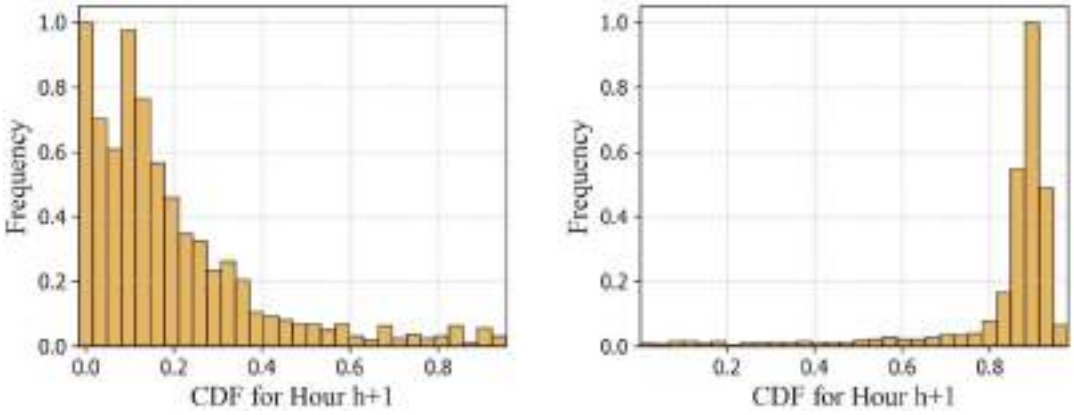


Fig. 6. – On the left: low end projection of the transition matrix for the SOUTH energy plant (initial CDF fixed at 0.12); two peaks are present for this initial condition. On the right: high end slice of the transition matrix for the SOUTH energy plant (initial CDF fixed at 0.92); in both cases, long tails are present.

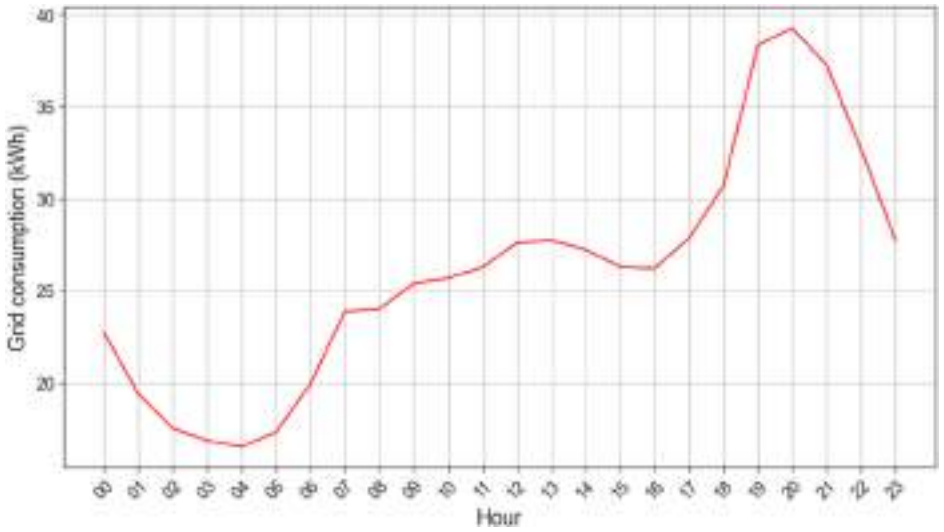


Fig. 7. Average daily profile of the NORTH renewable REC as simulated with the data from ARERA [56] and GSE [55].

Italian national energy agency (Gestore Servizi Energetici - GSE) [55] and data from the Italian national energy authority (Autorità di Regolazione per Energia Reti e Ambiente - ARERA) on the average monthly consumptions of residential users in different provinces in Italy [56]. The ARERA data includes information about monthly average consumption and percentage consumed at different times of the day, which are the input used to compute the consumption profiles starting from the normalized ones provided by GSE. Fig. 7 presents the generated average consumption profile of one day for the NORTH REC as an example. The REC's consumption profile will be assumed to be fixed, as suggested in section 2, due to the high number of REC consumers.

For each synthetic production profile, the shared energy is calculated on an hourly basis. From these computations, the probability distributions of the monthly shared energy are derived by summing the hourly values of each simulation. The relative width of each obtained distribution is calculated with respect to its quartiles and 0.05 and 0.95 quantiles as:

$$w = \frac{q_{max} - q_{min}}{mode}, \tag{14}$$

where q_{max} is either the upper quartile or the 0.95 quantile, and q_{min} the lower quartile or the 0.05 quantile of the distribution. The *mode* is the mode of the distribution. To quantify whether a value ν calculated from a benchmark profile is acceptable, a normalized relative distance from the mode is calculated as:

$$d = \frac{2|\nu - mode|}{mode} - w, \tag{15}$$

with w calculated as in equation (14). If d value is greater than zero, one may state that the value is outside the distributions' quartiles. The same analysis is performed for both the production and shared energy probability distributions.

The resulting distributions are compared as benchmark with the forecast obtained using the TMY and the AHP, obtained both from PVGIS. Calculations for an entire year are performed, too, and shared energy results are compared with the benchmark profiles. The final estimates are considered to be the mean of the yearly distributions, while the uncertainties Δ_{year} correspond to their sigmas. Finally, a 20-year projection is calculated. The 20-year estimate is achieved by multiplying the yearly estimates by the number of years, while the uncertainty $\Delta_{20-years}$ is obtained through error propagation [50] as

$$\Delta_{20-years} = \sqrt{20}\,\Delta_{year}, \tag{16}$$

since yearly distributions are expected to be Gaussian.

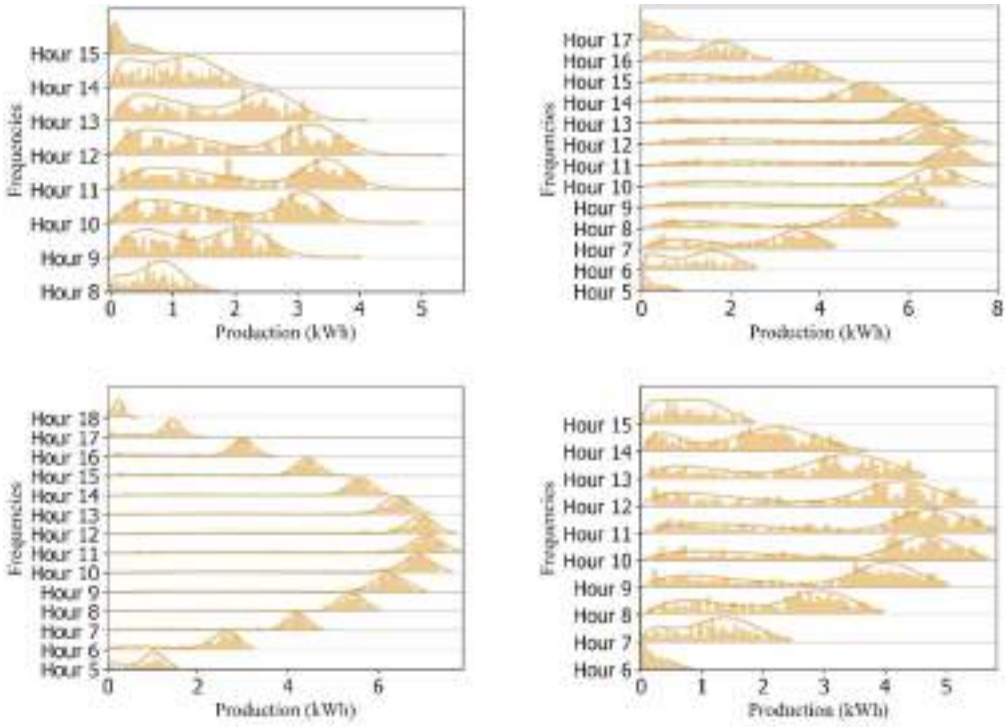


Fig. 8. Distributions for the NORTH energy plant in the months of January, April, July and October.

4. Results

4.1. Fixed time production PDFs and validation results

In this section, we present the results of fitting the fixed-time production probability density functions (PDFs) for the photovoltaic energy plants. We use the photovoltaic plant at NORTH location as a representative example.

Fig. 8 shows the fits for the NORTH PV plant during four months, illustrating its behavior in different seasons. The results are obtained following the procedure detailed in section 3.1.The plots are divided into hours of the day, shown as ticks on the y-axis. The x-axis represents the power in kWh, and for each hour, the normalized frequency of the data is shown. The histograms are built from the raw data, while the solid lines represent the fits of the histograms obtained with equation *(9)*. The trimodal nature of the distribution seems to be more significant during the winter and autumn seasons, while in the summer it is almost negligible, with the Gaussian peak always strongly dominant compared to the gamma peak. Pure Dirac Delta distributions are not shown. Figures regarding the other studied locations are reported in section 2 of the supplementary materials.

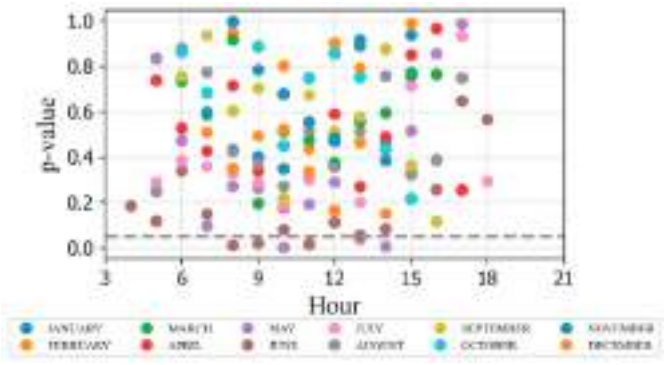


Fig. 9. – Kolmogorov-Smirnov tests for all the fits performed for the NORTH energy plant.

To test the quality of the fits, the p-value was calculated for each one using the Kolmogorov-Smirnov test, as shown in Fig. 9. The y-axis represents the p-values for all fits of the NORTH PV plant, while the x-axis indicates the hour of the day at which the fit is performed. Different colors correspond to different months. Values for which the p-value is higher than 0.05 (grey dotted line) are the ones for which the fit passes the Kolmogorov-Smirnov test. The test fails very seldom, indicating that the chosen trimodal prior in equation *(9)* is sufficiently flexible to capture almost all the fixed-time stochastic behavior of solar production. Similar results about the other locations can be found in section 2 of the supplementary materials.

4.2. Transition matrix fits results

Following section 3.3, fits of the transition matrix were performed for each location of the PV plant using equation (11) as a prior to obtain the analytical forms of the CDF propagator. Fig. 10 shows the diffusion coefficient estimates, clustered by month to account for statistically significant seasonal variations. The y-axis represents the diffusion coefficients fitted for fixed initial conditions, with error bars obtained using CERN's Iminuit library [49]. The x-axis shows these fixed initial conditions. Different colors correspond to different locations. Plots for January and July are presented as examples. The value of the diffusion coefficient becomes higher in the middle of the CDFs' domain for all locations and months, indicating increased weather variability. This effect is especially pronounced in January. The diffusion coefficient varies in a statistically significant way among the different months in each location.

Fig. 11 illustrates the time evolution behavior of the NORTH PV plant's July CDF transition matrix, shown both as a heatmap and as a series of projections with fixed initial condition. The top plots display the CDF transition matrices for the month of July at different time-steps with raw data: CDF pairs are separated by 1 h, 2 h and 3 h respectively. The diffusive behavior is evident in the three plots below, where the same transition matrices are shown together with the propagator in equation (11) calculated at different time-steps. Indeed, the width of the

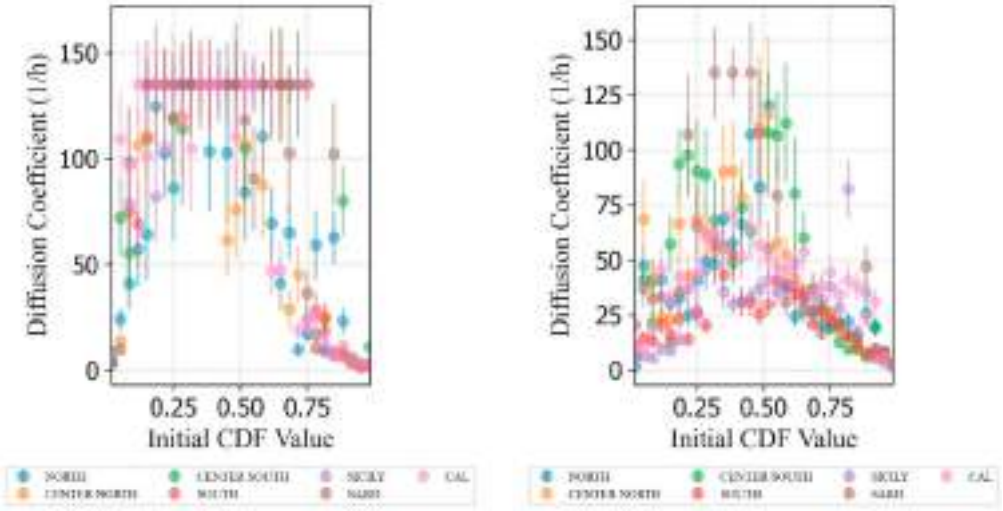


Fig. 10. – January and July diffusion coefficient estimations for all considered locations.

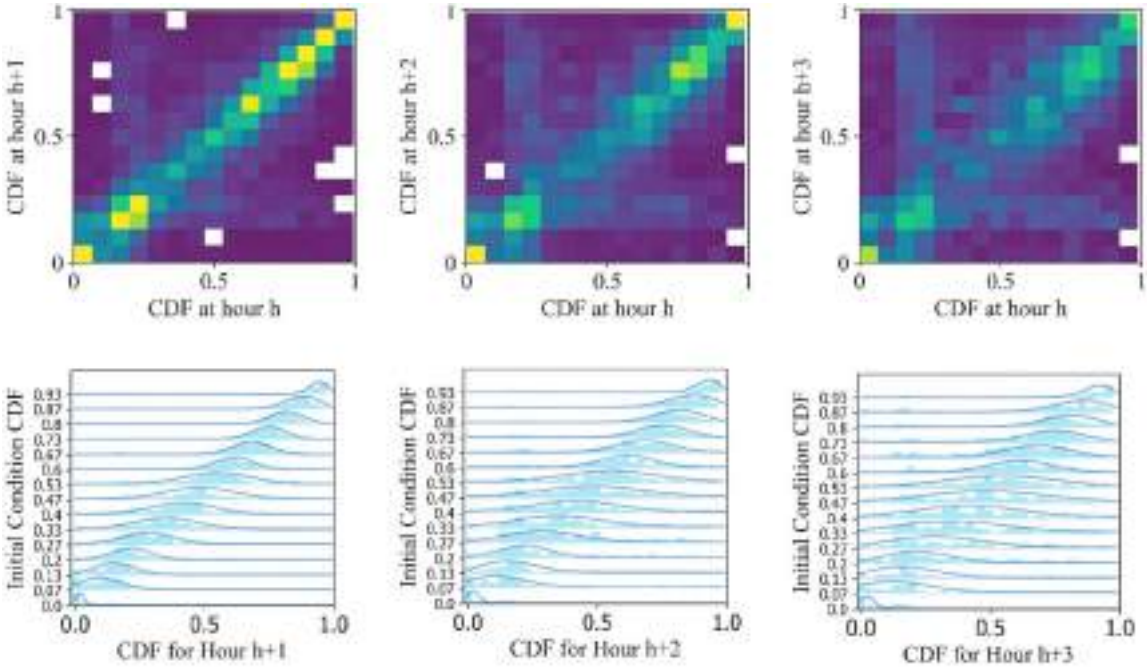


Fig. 11. Evolution of the CDF transition matrix in time.

data increases with the timespan. The propagators shown have the fixed parameters found during the fitting procedure, and evolve in time mimicking raw data's diffusive behavior. The propagator follows the data more thoroughly in the upper region of the CDF's domain, and with less precision in the lower region. The time evolution behavior of the propagator fits for the other locations are shown in the supplementary materials.

4.3. Synthetic data generation and validation

In this section, we validate the synthetic data generated for the photovoltaic energy plants following section 3.4.

Fig. 12 shows the SDI ratio for the NORTH PV plant, demonstrating that it is near the value of 1 when calculated on different hourly profiles on a monthly basis. This indicates that the variability of the simulated profiles closely matches that of the real profiles. Table 2 reports the average and standard deviation of the SDI ratios for all locations, for both the training and testing datasets. The value of 1 is always within less than one sigma from the mean for datasets, indicating that the synthetic hourly profiles mimic the monthly variability of real data well for all studied locations (see Fig. 13).

The same calculation was performed for hourly profiles across a year timespan for all locations. In this case, the widths of the distributions become narrower, but the means almost always stay within one sigma from their expected value of 1, as reported in Table 3. This indicates that weather variability is well captured by the model. The tests for the yearly SDI ratio for the CENTER-SOUTH and SARD PV plants for the test dataset produced a mean respectively three-sigma and two-sigma away from the expected value of 1. This result may indicate slightly different weather behavior in such locations in the test dataset with respect to the training dataset, which is within the statistics of the model. Fig. 14 show the example plots of the CENTER-NORTH PV plant.

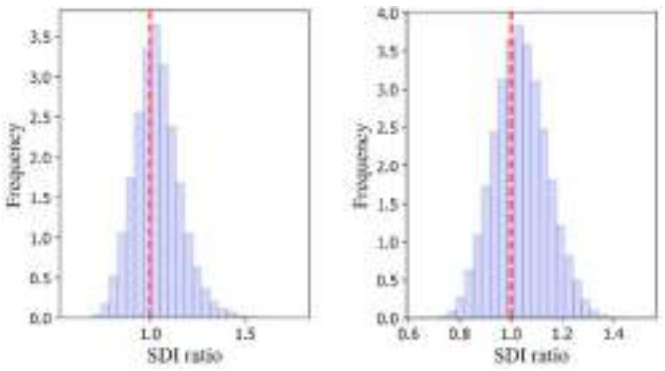


Fig. 12. – Density histograms displaying 144,000,000 (training dataset) and 480,000 (test dataset) SDI ratio data points for the NORTH PV plant. Each datapoint is obtained from equation (13). Both distributions are peaked aound 1.

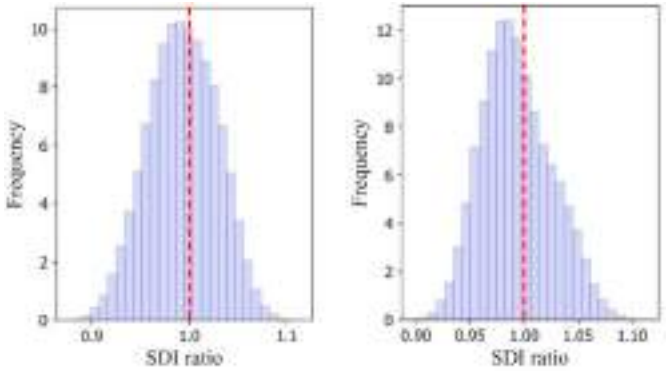


Fig. 13. Density histograms displaying 120,000 (training dataset) and 40,000 (testing dataset) SDI ratio data points for the CENTER-NORTH PV plant. Each datapoint is obtained from equation (13). Both distributions are peaked around 1.

Table 2
– Average and standard deviation of SDI ratios for all studied locations.

Region	Training Dataset (Mean ± SD)	Testing Dataset (Mean ± SD)
NORTH	**1.04 ± 0.12**	**1.04 ± 0.10**
CENTER-NORTH	**1.01 ± 0.11**	**1.02 ± 0.11**
CENTER-SOUTH	**0.96 ± 0.11**	**0.95 ± 0.13**
SOUTH	**1.00 ± 0.13**	**1.02 ± 0.13**
SICILY	**1.02 ± 0.17**	**1.03 ± 0.16**
CAL	**0.97 ± 0.09**	**0.98 ± 0.10**
SARD	**1.03 ± 0.15**	**1.04 ± 0.15**

Table 3
Average and standard deviation of yearly SDI ratios for all studied locations.

Region	Training Dataset (Mean ± SD)	Testing Dataset (Mean ± SD)
NORTH	**1.02 ± 0.03**	**1.03 ± 0.03**
CENTER-NORTH	**0.99 ± 0.04**	**0.99 ± 0.03**
CENTER-SOUTH	**0.97 ± 0.04**	**0.92 ± 0.03**
SOUTH	**1.01 ± 0.04**	**1.02 ± 0.03**
SICILY	**1.03 ± 0.03**	**1.03 ± 0.03**
CAL	**0.97 ± 0.04**	**0.97 ± 0.03**
SARD	**1.02 ± 0.03**	**1.06 ± 0.03**

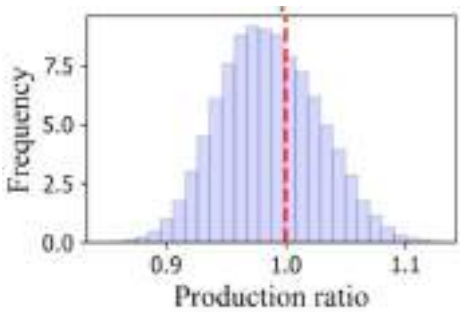


Fig. 14. Density histogram of the yearly production simulated ratio and the recorded yearly productions from the year 2017–2020 for the SOUTH energy plant. The peak is within less than 1-sigma from the value of 1.

Table 4
Monthly production ratios' means and standard deviations for training and test datasets across all locations.

Region	Training Dataset (Mean ± SD)	Testing Dataset (Mean ± SD)
NORTH	**1.04 ± 0.21**	**1.02 ± 0.20**
CENTER-NORTH	**1.07 ± 0.20**	**1.02 ± 0.16**
CENTER-SOUTH	**0.99 ± 0.21**	**0.92 ± 0.23**
SOUTH	**1.07 ± 0.19**	**1.02 ± 0.18**
SICILY	**1.01 ± 0.18**	**0.99 ± 0.18**
CAL	**1.07 ± 0.19**	**1.03 ± 0.16**
SARD	**1.02 ± 0.18**	**1.00 ± 0.15**

Table 5
Yearly Monthly production ratios' means and standard deviations for training and test datasets across all locations.

Region	Training Dataset (Mean ± SD)	Testing Dataset (Mean ± SD)
NORTH	**1.00 ± 0.05**	**0.99 ± 0.04**
CENTER-NORTH	**1.03 ± 0.06**	**0.99 ± 0.04**
CENTER-SOUTH	**0.98 ± 0.05**	**0.91 ± 0.04**
SOUTH	**1.01 ± 0.04**	**0.99 ± 0.04**
SICILY	**0.97 ± 0.04**	**0.96 ± 0.05**
CAL	**1.03 ± 0.04**	**1.01 ± 0.05**
SARD	**0.98 ± 0.03**	**0.98 ± 0.05**

Fig. 14 shows the yearly production ratios of the SOUTH energy plant obtained between the 10,000 simulated profiles and the test dataset. The bell-curve is peaked around the value of 1, indicating a close match between simulated and real profiles. Table 4 reports the means and standard deviations of the monthly production ratios for the training and test datasets across all locations. Table 5 provides the corresponding yearly results, which are consistently peaked near the value of 1 within less than 1-sigma in all cases except for the CENTER-SOUTH energy plant in the yearly validation of the test dataset. This finding aligns with the SDI yearly validation result. Since the corresponding results for the CENTER-SOUTH PV plant in training dataset are within one-sigma for both the SDI and the production ratio, it suggests that the data in the test dataset exhibits a slightly different behavior compared to the training dataset. Specifically, the test dataset appears to contain profiles with smaller variability and smaller producibility.

4.4. Evaluating uncertainty in solar energy communities

In this section, we analyze the impact of production stochasticity on the shared energy distribution of the seven RECs listed in Table 1. Indeed, an accurate estimation of community shared energy is fundamental to correctly assess the economic feasibility of a REC, as it could represent economic revenue. For example, in Italy, to support the

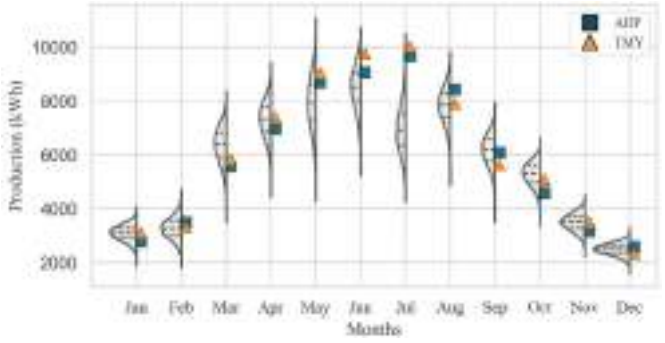


Fig. 15. Monthly production probability distributions for the SICILY PV plant. The blue square markers represent the results from the average hourly profile, and the orange triangle markers represent the TMY profile. Each distribution is based on 10,000 synthetic production profiles, with quartiles marked by black dotted lines.

diffusion of RECs, the legislator recently proposed an implementing decree contemplating contributions on capital funds and an incentive mechanism consisting of a premium tariff of 110 €/MWh recognized for shared energy, plus a correction linked to the level of insolation of the geographical area.[1] Therefore, the proposed framework could help estimating the average community shared energy in a certain area and for a certain time horizon and properly assess its uncertainty as a fundamental measure of economic risk.

4.4.1. Production stochasticity

The intrinsic variability of solar energy plays a crucial role in estimating the photovoltaic generation. Fig. 15 presents the monthly production probability distributions of the SICILY location, compared with the results obtained with the AHP, (blue square marker) and the TMY (orange triangle marker). The probability distributions are obtained by summing the hourly productions of the 10,000 synthetically generated production profiles separately for each month. The y-axis represents the monthly production of the individual simulations in kWh, while the x-axis displays the months as x-ticks. For each x-tick, the frequency of the data is shown. The quartiles for each distribution are highlighted with black dotted lines. The width of the distributions tends to be higher during the spring and autumn seasons, indicating greater variability in production during these times. Numerical values of the modes of the monthly and yearly distributions, 0.05 and 0.95 quantiles and quartiles are reported for each location in section 4 of the supplementary materials, together with the AHP and TMY estimates.

In Fig. 16, the relative width of the probability distributions calculated with equation (14), both in terms of quartiles and 0.05 and 0.95 quantiles, are shown for all the chosen locations. The y-axis represents the locations, while the x-axis the months. The color of the squares indicates the modulus of the relative width: the warmer colors signify higher values.

The relative variations range from as low as 8 % to up to 18 % with respect to the quartiles. On the other hand, the relative variations in the quantile plot range from 19 % to 47 %, reflecting the color scale in the plot. These plots together provide a comprehensive picture of the variability in solar production. The quartile plot focuses on the central tendency and consistency, while the quantile plot captures the potential for extreme variations.

Considering only a single profile (TMY or AHP), even if precise in estimating the most probable producibility, omits a diverse set of scenarios which may significantly differ from the mode and does not quantify how much the estimate may actually deviate from a real future measurement. Moreover, these widths are relative, meaning that this consideration becomes increasingly important as the size of the plant grows.

In Fig. 17, the relative distance between the AHP and TMY profiles estimations and the mode of the probability distribution as calculated in equation (15) is shown for all months and studied locations. The x-axis of the heatmaps represents the month, while the y-axis indicates the location. Red squares represent the location-month pairs for which the profiles fall outside the quartiles, while blue squares correspond to ones for which the value falls inside the quartile. There is no clear pattern indicating a systematic deviation for any particular month or location, suggesting that the deviations are random.

These results highlight that the estimates derived from the benchmark can be significantly different from the most typical value and can even fall outside the quartiles.

4.4.2. Uncertainty of the community shared energy

This section examines the variability and distribution of monthly shared energy within the case studies. By comparing the simulated shared energy with benchmark profiles, we gain insights into the alignment and discrepancies between different modeling approaches.

In Fig. 18, the monthly probability distributions of shared energy are shown for the SICILY case. Similar to Fig. 15, the x-axis displays the months as x-ticks, and for each x-tick the frequency of the monthly shared energy is shown, forming the probability distributions. The y-axis reports the monthly shared energy of the single simulations in kWh. The

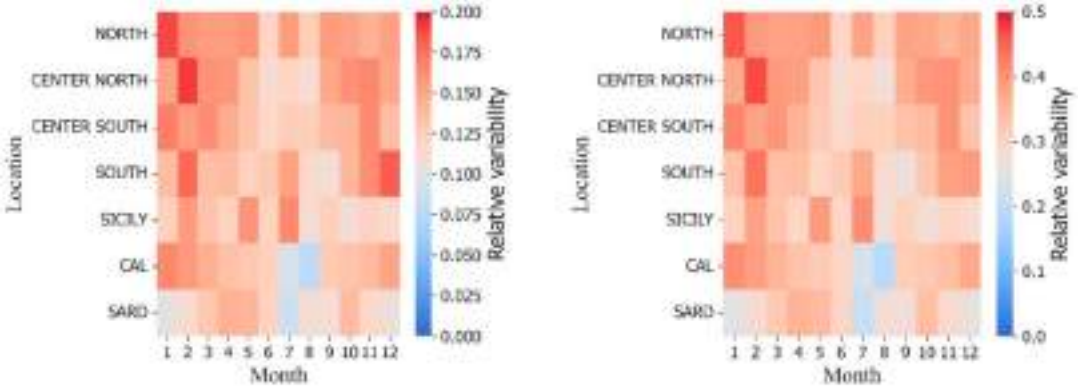


Fig. 16. Production Relative variability with respect to the probability distribution quartiles and 0.05 and 0.95 quantiles.

[1] Legislative decree DM n. 414 07.12.2023 "Decreto CER", https://www.mase.gov.it/sites/default/files/Decreto%20CER.pdf.

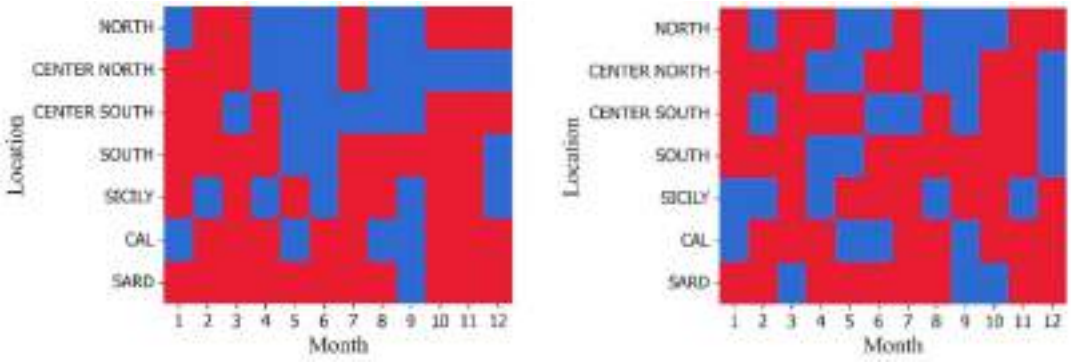


Fig. 17. Relative distance of the AHP and TMY profiles with respect to the mode compared to the relative variability of the distributions with respect to the quartiles. Red squares represent the pairs for which the profile fall out of the quartiles, while blue squares correspond to a location-month pair for which the value falls inside the quartile.

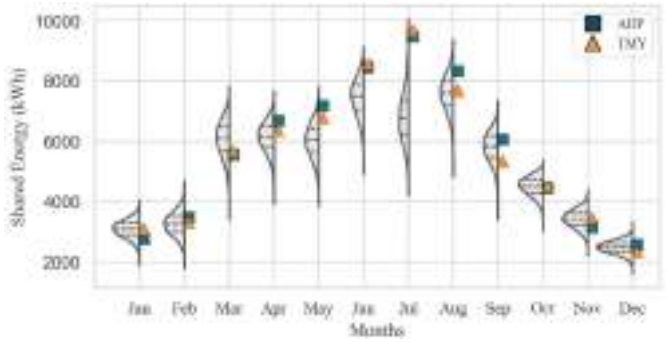


Fig. 18. – Monthly shared-energies' probability distributions for the SICILY REC compared with results obtained with benchmark profiles.

quartiles for each distribution are highlighted with black dotted lines. The blue squares markers represent the shared energy calculated using the AHP data, while the orange triangle markers represent the shared energy from the TMY.

During the winter, the energy consumption is always higher than the production, so the shared energy probability distribution mirrors the production. Consequently, the benchmark profiles' shared energy estimates behave similarly to those of the production in terms of relative position within the corresponding probability distributions. However, when the PV plant starts producing more than what is consumed during peak hours, the distribution becomes limited at the top compared to the production distribution. For the SICILY case, the effect becomes visible in the month of April. Comparing Figs. 18 and 15, the benchmark profiles shift upwards relative to their position in the production distribution, signaling a bias in the estimation. This behavior is systematic for the AHP, which falls out of the quartiles for all summer months at all locations. While the effect is less pronounced for the TMY profile, it is still present.

In Fig. 19, the differences between the shared energies obtained with the benchmark production profiles and the higher quartile of the shared energy probability distributions are shown. Data from all the seven locations is displayed for all the 12 months. The two heatmaps refer respectively to the results obtained from the AHP and the TMY profiles. Dark blue squares indicate that the benchmark estimate for the specific month and location is lower than the higher quartile of the corresponding distribution, while other colors indicate that the value exceeds the quartile. The warmer the color, the greater the difference. The heatmap reveals that the AHP profile consistently overestimates shared energy during summer months. The usage of the TMY leads to a similar results, though less pronounced.

Table 6
Yearly estimations of shared energy for all studied locations.

Yearly Shared Energy Estimation			
location	Mean ± SD% (MWh)	AHP profile (MWh)	TMY profile (MWh)
NORTH	52 ± 3 %	57	52
CENTER NORTH	50 ± 3 %	55	50
CENTER SOUTH	36 ± 2 %	42	37
SOUTH	58 ± 3 %	61	57
SICILY	63 ± 3 %	68	67
CAL	52 ± 2 %	56	52
SARD	68 ± 3 %	71	69

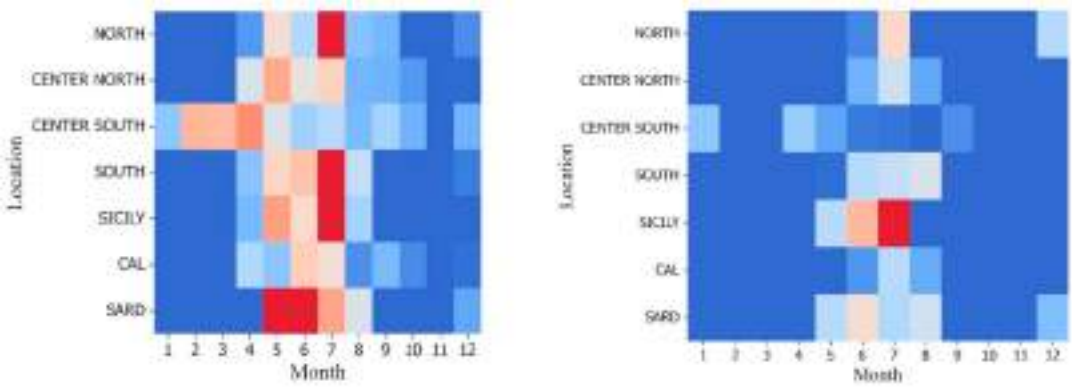


Fig. 19. Difference between the shared energies calculated with the benchmark profiles and the higher quartile of the corresponding probability distributions.

In Table 6, the yearly shared energy estimates are reported for all investigated locations. The table includes the framework's shared energy estimate, which corresponds to the mean of the yearly shared energy probability distribution in MWh associated to its relative uncertainty, identified with the sigma of the distribution divided by the mean itself, as well as the estimates from the benchmark methods. The shared energies calculated with the AHP profiles consistently exceed the upper quartile of the probability distributions by a large margin. In contrast, the shared energy calculated with the TMY profiles generally falls within the quartiles, except for the SICILY and SARD cases. In particular, the SARD case is within 1-sigma from the estimated mean, whereas the SICILY case is within 3-sigma. Fig. 19 suggests that the discrepancy in the SICILY region is due to the behavior of the TMY in June and July, where the TMY significantly exceeds the higher quartile. Comparing the yearly result with the monthly ones, it becomes evident that the TMY is more reliable for calculating the shared energy along longer timespans, while it is less reliable in monthly analyses. Overestimation tends to be systematic during the summer months when energy production is consistently exceeding energy consumption. However, this observation is also influenced by the REC consumption profiles.

The relative width of the distribution with respect to the quartiles, calculated according to equation (14), ranges from 3 % to 4.5 %. With respect to the 0.05 and 0.95 quantiles, it spans from 6.7 % up to 10 %. The shared energy uncertainty generated by a 50 kW peak power PV plant is estimated to be in the MWh order of magnitude. This last result depends on the energy consumption profiles, and cannot be considered general. Values of the modes of the monthly and yearly shared energy probability distributions, the 0.05 and 0.95 quantiles and the quartiles are reported in section 4 of the supplementary materials.

The 20-year shared energy projections are shown in Table 7 for all studied locations. Given the yearly estimations, 20-year results can be calculated by multiplying the yearly estimate for the number of years, while uncertainties propagate according to equation (16). As timespans get larger, the relative error becomes smaller and biases become more evident. The TMY estimations fall above the upper quartile of the distribution for the NORTH, CENTER-SOUTH, SICILY, CAL and SARD regions; specifically, they are not within 1-sigma from the mean, the SARD case is not within 3-sigma and the SICILY case falls out of the 10-sigma region. The Δ% in Table 7 is calculated for the benchmark methods as the distance from the model's estimate divided by the benchmark estimate itself. The AHP model overestimates shared energy by a large margin in all locations, especially in the CENTER SOUTH region where the AHP profile overestimates by 14.4 % with respect to our model. On the other hand, the use of the TMY profile results in a lower bias, of about 1 % except for SICILY, where it accounts for a 6.5 % overestimation. Nevertheless, even if the performance of the TMY usage may be acceptable, the uncertainty is not accounted.

Tables 7
–20 year estimations of shared energy for all studied locations.

20 Year Shared Energy Estimation					
location	Mean ± SD% (GWh)	AHP profile (GWh)	Δ% AHP	TMY profile (GWh)	Δ% TMY
NORTH	1.047 ± 0.7 %	1.131	7.4	1.059	1.1
CENTER NORTH	1.072 ± 0.6 %	1.092	7.8	1.009	0.2
CENTER SOUTH	0.725 ± 0.5 %	0.847	14.4	0.732	1.0
SOUTH	1.150 ± 0.6 %	1.223	6.0	1.139	−1.0
SICILY	1.253 ± 0.6 %	1.370	8.5	1.340	6.5
CAL	1.043 ± 0.5 %	1.110	6.0	1.050	0.7
SARD	1.353 ± 0.6 %	1.420	4.7	1.380	2.0

5. Discussion

The objective of this work was to associate statistical weights to the estimates of shared energy in a REC with photovoltaic energy plants, by obtaining its monthly probability distributions. To achieve this, a novel methodological framework was implemented to generate synthetic hourly production profiles. This framework relies on the existence of dependable historic solar radiation data. We used the SARAH [35] and SARAH-2 [36] databases provided by PVGIS [37], which also provides estimates of photovoltaic production. Seven locations in Italy were studied. Shared energy probability distributions were obtained for each month and location and compared with benchmark methods used typically in feasibility studies, i.e. the TMY and the AHP profiles. The results show that benchmark methods' estimations of shared energy very often fall out of the quartiles of the obtained monthly probability distributions. Additionally, a clear systematic overestimation is present with the AHP profile, as the yearly shared energy estimates consistently fall out of the 0.95 quantile of the yearly distribution. The usage of TMY profiles, while often producing monthly estimates outside the quartiles, generally produces consistent results on a yearly basis. In longer time ranges, the bias becomes more evident, causing the TMY to fall out of at least 1-sigma in most of the studied geographic areas. During months with higher energy production than consumption, the TMY shows a similar, yet less marked, bias compared to the AHP profile. The chosen case studies focus on a PV energy plant connected to the grid, excluding physical self-consumption. While this approach allows for the decoupling of consumption and production stochasticity, it does not account for the variability in individual users' consumption profiles. As a result, this work explicitly accounts for uncertainties deriving exclusively from PV generation, rather than the ones arising from both generation and energy demand. Future work will provide a similar framework for the consumption profiles treatment.

The choice of the form of the fixed-time probability distribution of photovoltaic production, $f(p,t)$ in equation *(9)* is somewhat arbitrary, and all consequent results depend on it. An alternative possibility would have been to use a Gaussian Mixture, as done for the clearness index by Ref. [30], not forcing the model to have three modes at most. Another possibility would have been to use the marginal distribution built by Ref. [44] using Fourier analysis, since it seems to yield better results in terms of Kolmogorov-Smirnov tests. However, the chosen form of $f(p,t)$ is particularly easy to handle and can encapsulate the behavior of radiation throughout the day, including nighttime. The choice of the Dirac Delta distribution is particularly interesting for the CDF treatment, allowing the algorithm to handle also nighttime hours, unlike other models [42,44].

The algorithm avoided the need for day-type classification, sidestepping the problem identified in previous studies [30], since the weather dynamics was completely stored in a set of continuous functions: the propagators of the photovoltaic production CDFs. The primary focus was on the CDFs of production profiles rather than the clearness index or the raw data.

Using the CDF to encapsulate the weather behavior in a single variable is well-suited for long term planning. However, it may be less effective for short-term forecasting, where a multivariate regression model like the one proposed in Ref. [24] is more appropriate. For monthly estimates, our model proved to be a better choice than the TMY, offering the advantage of accounting for potential variations in outcomes over longer timespans. Moreover, in RECs, where weather conditions may be assumed to remain consistent across a small geographical area, the same CDF time series can be used to generate production profiles of different PV plants by exploiting their hourly marginal distributions. This approach effectively models perfectly correlated production across all plants within the REC. This assumption of correlated production offers an approximation that simplifies the analysis without compromising its applicability to real-world scenarios. Future work could explore cases with partially decorrelated PV plants to examine the

impact of spatial variability in weather conditions.

The methodological approach, which is based on the path-integral and uses the Fokker-Planck formalism to solve the integral numerically using Monte Carlo, is not yet widely used. However, it offers a generalizable method applicable to other stochastic time series problems and different renewable energy technologies, i.e. wind turbines. This could be valuable for studying the effects stochasticity in real-world scenarios.

The primary drawback of this approach lies in the computational effort required to estimate the propagators' parameters during the training phase. However, this issue is mitigated by the fact that the parameters only need to be calculated once and can then be reused. Future research could address this limitation by exploring the geographical dependency of these parameters, potentially enabling more efficient estimation methods. For RECs, which are typically localized within small geographical areas, a single set of propagators may suffice for all energy plants in the region. This would eliminate the need to estimate parameters individually for each plant, thereby reducing computational demands while maintaining accuracy. Moreover, the chosen form of the propagator in equation (11) and the initial potential shape might not be optimal, especially for extreme initial conditions. A more refined potential in the Fokker-Planck equation could significantly improve the results. Additionally, imposing an absorbing boundary condition at zero might enhance the model's precision, particularly given the usual presence of two peaks at the low end of the transition matrix. Potentially, dealing with a more elegant analytical form of the propagator may enhance performance significantly. Further research can build upon the findings discussed in this paper, both from a mathematical perspective and in terms of practical applications. For instance, testing the hypothesis that the diffusion coefficient in the Fokker-Planck equation *(8)* is time-independent could be explored using smaller timestep profiles. This would require regressing the parameters of the marginal distributions of fixed-time productions to select the appropriate energy production marginal distribution for a given fixed-time. Given the regular evolution of the distribution observed in Fig. 9, this approach appears feasible and could be tested with an appropriate dataset.

The generality of the framework and the versatility of the output make it suitable for use in optimization problems that consider uncertainties using a Monte Carlo approach, such as the ones in the works of [17–20], [21–23]. This enables the generation of more accurate and statistically robust results. Despite these potential improvements, the model effectively captures the diffusive behavior of the CDF, demonstrating the promise of this approach.

6. Conclusions

In this paper, we introduced a new simulation framework methodology to generate synthetic hourly photovoltaic production profiles over year timespan to understand the behavior of the probability distributions of the monthly shared energy of a solar REC. Seven locations in Italy were studied suing PVGIS photovoltaic production profiles as training and testing datasets, with the first dataset comprising hourly data from 2005 to 2016, and the second from 2017 to 2020. The proposed framework relies on the path-integral and Fokker-Planck formalisms from a mathematical standpoint and employs a Monte Carlo approach for the numerical implementation.

Validations proved the effectiveness and precision of the framework in all the seven selected chosen locations, accurately mimicking the variability of the energy production and the monthly and yearly producibility of the PV plants.

For each of the chosen locations, 10,000 synthetic photovoltaic production profiles were generated, simulating a 50 kW peak power plant connected to the grid and supplying a REC of 100 residential consumers. The RECs consumption profiles were assumed to be fixed and were generated using official location-dependent data made available by official public sources. Shared energy was calculated in the RECs for each synthetically generated weather scenario, resulting in monthly probability distributions of production and shared energy. The yearly shared energy probability distribution was calculated, exhibiting uncertainties ranging from 2 % to 3 %. Future work should focus on incorporating the stochasticity of consumption profiles and simulating the decorrelation of geographically distant PV energy plants. Moreover, mathematical advancements in the propagator treatment could significantly improve the performance of its parameter estimation algorithm.

The results were compared with the shared energy estimations obtained from the production time series built by benchmark profiles, i.e. the AHP and the TMY. The comparison revealed that benchmark methods introduce a systematic overestimation in shared energy during months when energy production exceeds consumption, and that very often their estimates do not fall within 1-sigma with respect to estimated means in 20-year timespans.

This study demonstrates the value of the proposed framework in providing a more accurate and reliable estimation of shared energy in photovoltaic RECs by considering the inherent stochasticity of solar production.

CRediT authorship contribution statement

F. De Bettin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Data curation, Conceptualization. **F.D. Minuto:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Investigation. **D.S. Schiera:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation. **A. Lanzini:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

A. Lanzini, F. De Bettin received funding from DBA PRO. S.p.A.

F.D. Minuto carried out this study within Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014–2020.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2025.122604.

References

[1] P. Bórawski, L. Holden, A. Bełdycka-Bórawska, Perspectives of photovoltaic energy market development in the European union, Energy 270 (2023) 126804, https://doi.org/10.1016/j.energy.2023.126804.

[2] M. Haji Bashi, L. De Tommasi, A. Le Cam, L.S. Relaño, P. Lyons, J. Mundó, I. Pandelieva-Dimova, H. Schapp, K. Loth-Babut, C. Egger, M. Camps, B. Cassidy, G. Angelov, C.E. Stancioff, A review and mapping exercise of Energy Community Regulatory Challenges in European member states based on a survey of collective energy actors, Renew. Sustain. Energy Rev. 172 (2023) 113055, https://doi.org/10.1016/j.rser.2022.113055.

[3] R. Trevisan, E. Ghiani, F. Pilo, Renewable energy communities in positive energy districts: a governance and realisation framework in compliance with the Italian regulation, Smart Cities 6 (1) (2023) 563–585, https://doi.org/10.3390/smartcities6010026.

[4] V.Z. Gjorgievski, S. Cundeva, G.E. Georghiou, Social arrangements, technical designs and impacts of energy communities: a review, Renew. Energy 169 (2021) 1138–1156, https://doi.org/10.1016/j.renene.2021.01.078.

[5] J. Sousa, J. Lagarto, C. Camus, C. Viveiros, F. Barata, P. Silva, R. Alegria, O. Paraíba, Renewable Energy Communities Optimal design supported by an optimization model for investment in PV/wind capacity and renewable electricity

sharing, Energy 283 (2023) 128464, https://doi.org/10.1016/j.energy.2023.128464.
[6] A. Jafari, T. Khalili, H.G. Ganjehlou, A. Bidram, Optimal integration of renewable energy sources, diesel generators, and demand response program from pollution, financial, and reliability viewpoints: a multi-objective approach, J. Clean. Prod. 247 (2020) 119100, https://doi.org/10.1016/j.jclepro.2019.119100.
[7] A. Cosic, M. Stadler, M. Mansoor, M. Zellinger, Mixed-integer linear programming based optimization strategies for Renewable Energy Communities, Energy 237 (2021) 121559, https://doi.org/10.1016/j.energy.2021.121559.
[8] G. Barone, A. Buonomano, C. Forzano, G.F. Giuzio, A. Palombo, Increasing renewable energy penetration and energy independence of Island Communities: a novel dynamic simulation approach for energy, economic, and environmental analysis, and Optimization, J. Clean. Prod. 311 (2021) 127558, https://doi.org/10.1016/j.jclepro.2021.127558.
[9] A. Cielo, P. Margiaria, P. Lazzeroni, I. Mariuzzo, M. Repetto, Renewable energy communities business models under the 2020 Italian regulation, J. Clean. Prod. 316 (2021) 128217, https://doi.org/10.1016/j.jclepro.2021.128217.
[10] S.-H. Park, Y.-S. Jang, E.-J. Kim, Multi-objective optimization for sizing multi-source renewable energy systems in the community center of a residential apartment complex, Energy Convers. Manag. 244 (2021) 114446, https://doi.org/10.1016/j.enconman.2021.114446.
[11] T. Weckesser, D.F. Dominković, E.M.V. Blomgren, A. Schledorn, H. Madsen, Renewable energy communities: optimal Sizing and distribution grid impact of photo-voltaics and Battery Storage, Appl. Energy 301 (2021) 117408, https://doi.org/10.1016/j.apenergy.2021.117408.
[12] E. Gomes, L. Pereira, A. Esteves, H. Morais, PyECOM: a python tool for analyzing and simulating energy communities, SoftwareX 24 (2023) 101580, https://doi.org/10.1016/j.softx.2023.101580.
[13] S. Alsadi, T. Khatib, Photovoltaic Power Systems Optimization Research Status: a review of criteria, constrains, models, techniques, and software tools, Appl. Sci. 8 (10) (2018) 1761, https://doi.org/10.3390/app8101761.
[14] T. Huld, E. Paietta, P. Zangheri, I. Pinedo Pascua, Assembling typical meteorological year data sets for building energy performance using reanalysis and satellite-based data, Atmosphere 9 (2) (2018) 53, https://doi.org/10.3390/atmos9020053.
[15] L. Giannuzzo, F.D. Minuto, D.S. Schiera, A. Lanzini, Reconstructing hourly residential electrical load profiles for renewable energy communities using non-intrusive machine learning techniques, Energy and AI 15 (2024) 100329, https://doi.org/10.1016/j.egyai.2023.100329.
[16] Z. Zhang, R. Jing, J. Lin, X. Wang, K.H. van Dam, M. Wang, C. Meng, S. Xie, Y. Zhao, Combining agent-based residential demand modeling with design optimization for Integrated Energy Systems Planning and Operation, Appl. Energy 263 (2020) 114623, https://doi.org/10.1016/j.apenergy.2020.114623.
[17] M. Bashir, J. Sadeh, Optimal sizing of hybrid wind/photovoltaic/battery considering the uncertainty of wind and photovoltaic power using Monte Carlo. 2012 11th International Conference on Environment and Electrical Engineering, 2012, pp. 1081–1086, https://doi.org/10.1109/eeeic.2012.6221541.
[18] S. Kaplanis, E. Kaplani, A model to predict expected mean and stochastic hourly global solar radiation I(H;N) values, Renew. Energy 32 (8) (2007) 1414–1425, https://doi.org/10.1016/j.renene.2006.06.014.
[19] Z. Zhou, J. Zhang, P. Liu, Z. Li, M.C. Georgiadis, E.N. Pistikopoulos, A two-stage stochastic programming model for the optimal design of Distributed Energy Systems, Appl. Energy 103 (2013) 135–144, https://doi.org/10.1016/j.apenergy.2012.09.019.
[20] A. Rathore, N.P. Patidar, Reliability constrained socio-economic analysis of Renewable Generation based standalone hybrid power system with storage for off-grid communities, IET Renew. Power Gener. 14 (12) (2020) 2142–2153, https://doi.org/10.1049/iet-rpg.2019.0906.
[21] A. Rathore, N.P. Patidar, Optimal sizing and allocation of renewable based distribution generation with gravity energy storage considering stochastic nature using particle swarm optimization in radial distribution network, J. Energy Storage 35 (2021) 102282, https://doi.org/10.1016/j.est.2021.102282.
[22] M. Sharafi, T.Y. ElMekkawy, Stochastic optimization of hybrid renewable energy systems using sampling average method, Renew. Sustain. Energy Rev. 52 (2015) 1668–1679, https://doi.org/10.1016/j.rser.2015.08.010.
[23] G. Mavromatidis, K. Orehounig, J. Carmeliet, Design of Distributed Energy Systems under uncertainty: a two-stage stochastic programming approach, Appl. Energy 222 (2018) 932–950, https://doi.org/10.1016/j.apenergy.2018.04.019.
[24] J. Zheng, Y. Kou, M. Li, Q. Wu, Stochastic optimization of cost-risk for integrated energy system considering wind and solar power correlated, Journal of Modern Power Systems and Clean Energy 7 (6) (2019) 1472–1483, https://doi.org/10.1007/s40565-019-0519-4.
[25] R.H.B. Exell, The fluctuation of solar radiation in Thailand, Sol. Energy 18 (6) (1976) 549–554, https://doi.org/10.1016/0038-092x(76)90074-8.
[26] B.Y.H. Liu, R.C. Jordan, The Interrelationship and characteristic distribution of direct, diffuse and total solar radiation, Sol. Energy 4 (3) (1960) 1–19, https://doi.org/10.1016/0038-092x(60)90062-1.
[27] B. Bartoli, B. Coluzzi, V. Cuomo, M. Francesca, C. Serio, Autocorrelation of daily global solar radiation, Il Nuovo Cimento C 4 (2) (1981) 113–122, https://doi.org/10.1007/bf02507395.
[28] R.J. Aguiar, M. Collares-Pereira, J.P. Conde, Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices, Sol. Energy 40 (3) (1988) 269–279, https://doi.org/10.1016/0038-092x(88)90049-7.
[29] V.A. Graham, K.G.T. Hollands, A method to generate synthetic hourly solar radiation globally, Sol. Energy 44 (6) (1990) 333–341, https://doi.org/10.1016/0038-092x(90)90137-2.
[30] Â. Frimane, T. Soubdhan, J.M. Bright, M. Aggour, Nonparametric Bayesian-based recognition of solar irradiance conditions: application to the generation of high temporal resolution synthetic solar irradiance data, Sol. Energy 182 (2019) 462–479, https://doi.org/10.1016/j.solener.2019.02.052.
[31] M. Pérez-Ortiz, S. Jiménez-Fernández, P. Gutiérrez, E. Alexandre, C. Hervás-Martínez, S. Salcedo-Sanz, A review of classification problems and algorithms in renewable energy applications, Energies 9 (8) (2016) 607, https://doi.org/10.3390/en9080607.
[32] M.M. Rienecker, M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, J. Woollen, Merra: NASA's modern-era retrospective analysis for research and applications, J. Clim. 24 (14) (2011) 3624–3648, https://doi.org/10.1175/jcli-d-11-00015.1.
[33] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, F. Vitart, The era-interim reanalysis: configuration and performance of the Data Assimilation System, Q. J. R. Meteorol. Soc. 137 (656) (2011) 553–597, https://doi.org/10.1002/qj.828.
[34] S. Kobayashi, Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, K. Takahashi, The JRA-55 reanalysis: general specifications and basic characteristics, Journal of the Meteorological Society of Japan. Ser. II 93 (1) (2015) 5–48, https://doi.org/10.2151/jmsj.2015-001.
[35] Richard Müller, Uwe Pfeifroth, Christine Träger-Chatterjee, Roswitha Cremer, Jörg Trentmann, Rainer Hollmann, Surface solar radiation data set - heliosat (SARAH) - edition 1, satellite application facility on climate monitoring, 10.5676/EUM_SAF_CM/SARAH/V001, https://doi.org/10.5676/EUM_SAF_CM/SARAH/V001, 2015.
[36] Uwe Pfeifroth, Steffen Kothe, Jörg Trentmann, Rainer Hollmann, Petra Fuchs, Johannes Kaiser, Martin Werscheck, Surface radiation data set - heliosat (SARAH) - edition 2.1, satellite application facility on climate monitoring, 10.5676/EUM_SAF_CM/SARAH/V002_01, https://doi.org/10.5676/EUM_SAF_CM/SARAH/V002_01, 2019.
[37] M. Šúri, T.A. Huld, E.D. Dunlop, H.A. Ossenbrink, Potential of solar electricity generation in the European Union member states and candidate countries, Sol. Energy 81 (10) (2007) 1295–1305, https://doi.org/10.1016/j.solener.2006.12.007.
[38] S. Pfenninger, I. Staffell, Long-term patterns of European PV output using 30 years of validated hourly reanalysis and Satellite Data, Energy 114 (2016) 1251–1265, https://doi.org/10.1016/j.energy.2016.08.060.
[39] J. Zhang, L. Zhao, S. Deng, W. Xu, Y. Zhang, A critical review of the models used to estimate solar radiation, Renew. Sustain. Energy Rev. 70 (2017) 314–329, https://doi.org/10.1016/j.rser.2016.11.124.
[40] M. Guermoui, F. Melgani, K. Gairaa, M.L. Mekhalfi, A comprehensive review of hybrid models for solar radiation forecasting, J. Clean. Prod. 258 (2020) 120357, https://doi.org/10.1016/j.jclepro.2020.120357.
[41] I.K. Bazionis, M.A. Kousounadis-Knousen, P.S. Georgilakis, E. Shirazi, D. Soudris, F. Catthoor, A taxonomy of short-term Solar power forecasting: classifications focused on climatic conditions and input data, IET Renew. Power Gener. 17 (9) (2023) 2411–2432, https://doi.org/10.1049/rpg2.12736.
[42] J. Munkhammar, J. Widén, An autocorrelation-based copula model for generating realistic clear-sky index time-series, Sol. Energy 158 (2017) 9–19, https://doi.org/10.1016/j.solener.2017.09.028.
[43] R.B. Nelsen, An Introduction to Copulas, Springer, 2011.
[44] A.F. Ramírez, C.F. Valencia, S. Cabrales, C.G. Ramírez, Simulation of photo-voltaic power generation using copula autoregressive models for solar irradiance and air temperature time series, Renew. Energy 175 (2021) 44–67, https://doi.org/10.1016/j.renene.2021.04.115.
[45] M. Chaichian, A.P. Demichev, Path Integrals in Physics. Volume 1, Stochastic Processes and Quantum Mechanics, Institute of Physics Pub, 2001.
[46] H. Risken, The Fokker-Planck Equation, Springer Series in Synergetics, 1996, https://doi.org/10.1007/978-3-642-61544-3.
[47] F. Cohen Tenoudji, The Dirac distribution, Modern Acoustics and Signal Processing (2016) 59–75, https://doi.org/10.1007/978-3-319-42382-1_4.
[48] F.D. Minuto, P. Lazzeroni, R. Borchiellini, S. Olivero, L. Bottaccioli, A. Lanzini, Modeling technology retrofit scenarios for the conversion of condominium into an energy community: an Italian case study, J. Clean. Prod. 282 (2021) 124536, https://doi.org/10.1016/j.jclepro.2020.124536.

[49] F. James, M. Roos, Minuit - a system for function minimization and analysis of the parameter errors and correlations, Comput. Phys. Commun. 10 (6) (1975) 343–367, https://doi.org/10.1016/0010-4655(75)90039-9.
[50] N.L. Johnson, A.W. Kemp, S. Kotz, Univariate Discrete Distributions, Wiley, 2005.
[51] M. Gitterman, Mean first passage time for Anomalous Diffusion, Phys. Rev. 62 (5) (2000) 6065–6070, https://doi.org/10.1103/physreve.62.6065.
[52] B.E. Rapp, Microfluidics: Modelling, Mechanics and Mathematics, Elsevier, 2017.
[53] B.G. da Costa, I.S. Gomez, E.P. Borges, Deformed Fokker-Planck equation: inhomogeneous medium with a position-dependent mass, Phys. Rev. 102 (6) (2020), https://doi.org/10.1103/physreve.102.062105.
[54] N. Salazar-Peña, A. Tabares, A. González-Mancera, Sequential stochastic and bootstrap methods to generate synthetic solar irradiance time series of high temporal resolution based on historical observations, Sol. Energy 264 (2023) 112030, https://doi.org/10.1016/j.solener.2023.112030.
[55] Autorità di Regolazione per Energia, Reti e Ambiente, Analisi dei Consumi dei Clienti domestici, Arera, 2024. https://www.arera.it/dati-e-statistiche/dettaglio/analisi-dei-consumi-dei-clienti-domestici.
[56] Gestore dei Servizi Energetici, Comunità di energia rinnovabile e gruppi di autoconsumatori, PUBBLICATI i profili standard per prevlievo e immissione 2023. https://www.gse.it/servizi-per-te/news/comunita-di-energia-rinnovabile-e-gruppi-di-autoconsumatori-pubblicati-i-profili-standard-per-prevlievo-e-immissione-2023, 2024.