

Article

Forecasting the Power Generation of a Solar Power Plant Taking into Account the Statistical Characteristics of Meteorological Conditions

Vitalii Kuznetsov ^{1,*}, Valeriy Kuznetsov ^{2,*}, Zbigniew Ciekanowski ³, Valeriy Druzhinin ^{4,*}, Valerii Tytiuk ⁵, Artur Rojek ², Tomasz Grudniewski ⁶ and Viktor Kovalenko ⁷

- ¹ Department of Electrical Engineering, Faculty of Electromechanics and Electrometallurgy, Dnipro Metallurgical Institute, Ukrainian State University of Science and Technologies, 2 Lazaryana Street, 49000 Dnipro, DR, Ukraine
 - ² Electric Energy Department, Railway Research Institute, 50 Józefa Chłopickiego Street, 04-275 Warsaw, Poland; arojek@ikolej.pl
 - ³ Department of Security Education, War Studies University, av. Chruściela 103, 00-910 Warsaw, Poland; zbigniew@ciekanowski.pl
 - ⁴ Department of Power Engineering, Faculty of Energy, Transport and Management Systems, Non-Profit Joint-Stock Company «Karaganda Industrial University», Republic Ave., 30, Temirtau City 101400, KR, Kazakhstan
 - ⁵ Department of Electromechanics, Electrotechnical Faculty, Kryvyi Rih National University, Vitaly Matusevich, Street, 11, 50027 Kryvyi Rih, DR, Ukraine; tytiuk@knu.edu.ua
 - ⁶ John Paul II Academy in Biała Podlaska, Rector's Office, Sidorska Street 95/97, 21-500 Biała Podlaska, Poland; t.grudniewski@dyd.akademibialska.pl
 - ⁷ Department of Electrical Engineering and Cyber-Physical Systems, Y.M. Potebnia Engineering Educational and Scientific Institute, Zaporizhzhia National University, 66 Universytetska Street, 69600 Zaporizhzhia, ZR, Ukraine; victor.l.kovalenko@znu.edu.ua
- * Correspondence: v.v.kuznetsov@ust.edu.ua (V.K.); vkuznetsov@ikolej.pl (V.K.); v.druzhinin@tttu.edu.kz (V.D.)

Academic Editor: Jesús Polo

Received: 6 September 2025

Revised: 3 October 2025

Accepted: 9 October 2025

Published: 11 October 2025

Citation: Kuznetsov, V.; Kuznetsov, V.; Ciekanowski, Z.; Druzhinin, V.; Tytiuk, V.; Rojek, A.; Grudniewski, T.; Kovalenko, V. Forecasting the Power Generation of a Solar Power Plant Taking into Account the Statistical Characteristics of Meteorological Conditions. *Energies* **2025**, *18*, 5363. <https://doi.org/10.3390/en18205363>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract

The integration of solar generation into national energy balances is associated with a wide range of technical, economic, and organizational challenges, the solution of which requires the adoption of innovative strategies for energy system management. The inherent variability of electricity production, driven by fluctuating climatic conditions, complicates system balancing processes and necessitates the reservation of capacities from conventional energy sources to ensure reliability. Under modern market conditions, the pricing of generated electricity is commonly based on day-ahead forecasts of day energy yield, which significantly affects the economic performance of solar power plants. Consequently, achieving high accuracy in day-ahead electricity production forecasting is a critical and highly relevant task. To address this challenge, a physico-statistical model has been developed, in which the analytical approximation of daily electricity generation is represented as a function of a random variable—cloud cover—modeled by a β -distribution. Analytical expressions were derived for calculating the mathematical expectation and variance of daily electricity generation as functions of the β -distribution parameters of cloudiness. The analytical approximation of daily generation deviates from the exact value, obtained through hourly integration, by an average of 3.9%. The relative forecasting error of electricity production, when using the mathematical expectation of cloudiness compared to the analytical approximation of daily generation, reaches 15.2%. The proposed forecasting method, based on a β -parametric cloudiness model, enhances the accuracy of day-ahead production forecasts, improves the economic efficiency of solar

power plants, and contributes to strengthening the stability and reliability of power systems with a substantial share of solar generation.

Keywords: solar energy integration; photovoltaic power plants; beta distribution; cloudiness modeling; probabilistic energy yield; power system stability

1. Introduction

The increasing share of solar energy in the structure of the energy balance of developed countries generates a number of specific challenges that require a comprehensive approach to their mitigation. One of the key factors is the instability of electricity generation, caused by the dependence of solar installations on climatic conditions and diurnal variations in solar radiation intensity. This necessitates capacity reservation based on conventional energy sources or the wide application of energy storage systems. Another important issue is the uneven distribution of generation over time, which complicates power system balancing processes and requires improvements in dispatch control mechanisms. At the same time, the growing share of solar generation calls for the modernization of grid infrastructure, primarily through the deployment of smart grids capable of flexibly responding to fluctuations in both demand and production. A significant challenge also arises from the reduced efficiency of conventional power plants, which are forced to operate in load-following modes, leading to shortened service life and increased electricity costs. The economic dimension is likewise of major importance, since the intensive financing of renewable energy requires state support and imposes an additional burden on public budgets. In the long term, the excessive concentration of solar power plants may result in local grid overloads and the deterioration of power quality indicators.

Under current market conditions, a pricing practice has been established for the generated electricity, based on day energy yield forecasting, which has a significant impact on the economic efficiency of solar power plants. This further highlights the importance of developing methods that ensure accurate and timely forecasting of day energy yield.

The authors have analyzed modern publications dedicated to the problems of short- and medium-term forecasting of solar energy production. The latest approaches based on artificial neural networks, deep learning, and other machine learning methods, such as Conv-GRU, NARX-GA, ARIMA, Grey Wolf Optimizer, Takagi–Sugeno–Kang neural networks, and Kohonen self-organizing maps, were considered. The use of statistical methods and deep learning models for processing satellite and meteorological data is also discussed, including models such as CNN, LSTM, GRU, CNN-LSTM, Extreme Learning Machine, and various hybrid approaches.

In the paper, the authors proposed an approach based on the use of β -distribution to describe the random nature of cloudiness. It is shown that β -distribution is a universal and convenient tool for modeling cloudiness since it allows effective consideration of regional climatic features. Based on this distribution, a mathematical model was developed that allows for analytical calculation of the main statistical characteristics of day energy yield, such as expected value and variance, taking into account the random nature of cloudiness changes during the day.

The paper provides analytical expressions for calculating day energy yield, taking into account the geographic location of the power plant, solar incidence angle, as well as atmospheric transmittance coefficients depending on cloudiness, humidity, and height above the sea level. A comparative analysis of the exact and approximate methods for

calculating daily electric power generation was carried out, showing satisfactory accuracy of the proposed analytical approach.

The authors considered two main generation forecasting strategies: the first strategy is based on the use of the expected value of cloudiness provided by third-party meteorological services, while the second one takes into account cloudiness as a random variable with β -distribution. The analysis showed that taking into account the statistical nature of cloudiness significantly increases forecasting accuracy, while the relative error of calculations using the traditional method can exceed 15%.

Thus, the proposed approach allows significantly improve the accuracy of forecasting the day energy yield of solar power plants, taking into account regional features of cloudiness and other meteorological conditions. The results obtained can be used in the operational activities of solar power plants and contribute to increasing the efficiency of integrating solar generation into energy systems.

2. A Review of Modern Literature

In contemporary electricity markets, the reliable forecasting of photovoltaic (PV) power generation has become a decisive factor for the efficient operation and financial performance of solar plants, as day-ahead production declarations must be aligned with grid requirements and influence market revenues. The inherent variability of solar irradiance, compounded by the limited accuracy of short- and medium-term meteorological forecasts, continues to pose significant challenges to operators and system planners. In recent years, considerable progress has been made in addressing these challenges through the development of advanced forecasting methodologies that exploit statistical analysis, artificial intelligence, and hybrid learning frameworks. To ensure a systematic and balanced representation of this progress, the present review focuses on peer-reviewed studies published between 2021 and 2025 in high-impact journals that reflect the main directions of methodological innovation. Rather than presenting the literature chronologically, the analysis is organized by classes of models to highlight both methodological distinctions and the incremental improvements introduced by each approach. Statistical baselines, including ARIMA and gray prediction models, continue to serve as important reference methods and are still competitive when enriched with high-quality exogenous variables. Neural and deep learning architectures such as convolutional, recurrent, temporal convolutional, and transformer-based networks have demonstrated the ability to capture nonlinear and spatio-temporal dependencies, significantly reducing forecasting error across multiple horizons. Hybrid approaches that combine fuzzy logic, clustering, ensemble learning, wavelet decomposition, or bias correction modules with machine learning have emerged as particularly effective in leveraging complementary strengths and enhancing robustness under highly variable conditions. At the same time, probabilistic and physics-informed formulations, including generative and diffusion models, have gained increasing attention for their capacity to quantify forecast uncertainty and provide interpretable indicators of daily energy yield, thereby addressing the requirements of system operators and energy markets. Comparative and survey studies underscore that despite these advances, critical gaps remain in terms of interpretability and probabilistic performance, particularly for day-ahead horizons. This structured synthesis of the literature establishes the basis for the subsequent class-by-class discussion of models and situates the present contribution within ongoing efforts to improve both the accuracy and transparency of solar power forecasting.

Statistical models—such as autoregressive formulations, gray prediction methods, and related time-series techniques—constitute the traditional foundation of solar forecasting and remain a necessary reference point for assessing methodological progress.

Their analytical tractability, interpretability, and modest data requirements make them particularly valuable in contexts of limited data availability or when transparency is a priority. Despite the rapid proliferation of advanced machine learning and deep learning architectures, statistical approaches continue to demonstrate competitiveness when supplied with high-quality exogenous meteorological inputs and careful preprocessing. The studies reviewed here illustrate how these methods have been adapted and extended to meet contemporary challenges, including data scarcity, variability of meteorological drivers, and the demand for reliable day-ahead predictions.

As a baseline statistical approach, Das [1] applied an ARIMA model for forecasting solar irradiance and the output of an 89.6 kWp PV plant, showing that even traditional time-series formulations can achieve competitive accuracy when sufficient historical data are available and exogenous variables are properly incorporated. Extending beyond classical autoregressive methods, He et al. [2] introduced a structurally adaptive gray prediction framework optimized with the Grey Wolf Algorithm, enabling the model to assign greater weight to new information and thereby improving adaptability in dynamic renewable energy contexts. This transition from conventional ARIMA to adaptive gray modeling illustrates how statistical paradigms have been modernized through algorithmic optimization. Further advancing this trajectory, Despotovic et al. [3] addressed the persistent challenge of limited historical data—particularly relevant for new PV installations—by combining autoregressive structures with extreme learning methods and implementing transfer learning across meteorological stations. The addition of clustering further enhanced model generalization, demonstrating how statistical approaches can be enriched with concepts from machine learning to mitigate data scarcity. A complementary perspective was offered by Gyeltshen et al. [4], who conducted a statistical evaluation of diversified irradiance repositories in Bhutan. Although their framework incorporated recurrent neural elements, the key contribution lay in underscoring the importance of high-quality and diverse statistical datasets as a foundation for robust forecasting performance. Completing this group of contributions, Benitez et al. [5] carried out a comparative study of SARIMAX, LSTM, and XGBoost for day-ahead photovoltaic output forecasting in the Philippines. Strikingly, their results revealed that SARIMAX, a classical statistical model, outperformed more advanced deep learning and gradient boosting methods when satellite-adjusted irradiance and other exogenous meteorological variables were carefully integrated. Taken together, these studies show that while statistical formulations remain essential benchmarks—and in specific contexts may even surpass more sophisticated methods—their inherent limitations in capturing nonlinear dynamics have ultimately motivated the transition toward neural and deep learning architectures.

Consequently, the next line of research has focused on neural and deep learning approaches. These architectures—including recurrent, convolutional, and more recently transformer-based networks—are specifically designed to capture complex spatio-temporal dependencies and to integrate heterogeneous input features such as irradiance, temperature, cloud cover, and satellite imagery. With the rapid expansion of computational resources and the increasing availability of large-scale datasets, deep learning has become one of the dominant directions of research in solar forecasting, demonstrating substantial improvements in predictive accuracy across a variety of temporal horizons. Against this background, a growing body of literature has explored deep learning techniques for solar forecasting, introducing diverse architectures and methodological refinements aimed at reducing forecast error and enhancing operational reliability.

A number of recent contributions exemplify the growing role of deep learning in photovoltaic forecasting, each highlighting a different methodological strand and

collectively illustrating the evolution from early convolutional–recurrent hybrids to more specialized architectures and comparative evaluations against statistical baselines.

Abdel-Basset et al. [6] propose PV-Net, a Conv-GRU architecture that integrates convolutional layers to capture local spatial features with gated recurrent units capable of learning temporal dependencies in photovoltaic generation data. The architecture is enhanced through bidirectional recurrence and residual connections, which improve information flow and reduce training instabilities such as vanishing gradients. The model was validated on Australian PV datasets and consistently outperformed classical statistical baselines and shallow machine learning models in short-term forecasting tasks, especially under conditions of high irradiance variability. The significance of this work lies not only in demonstrating the superiority of convolutional–recurrent hybrids over traditional time-series methods but also in setting a structural template that has been adopted in subsequent spatio-temporal forecasting research.

Hassan et al. [7] extend the recurrent modeling paradigm by developing a genetically optimized nonlinear autoregressive recurrent neural network (NARX-GA). In this framework, the genetic algorithm is applied to optimize hyperparameters and network weights, thereby addressing the sensitivity of recurrent networks to parameter initialization and tuning. The study focused on ultra-short-term forecasting, where rapid irradiance fluctuations can cause significant challenges for standard recurrent models. Results demonstrated that the GA-enhanced NARX network achieved more stable and accurate predictions across diverse meteorological conditions, outperforming both unoptimized RNNs and statistical benchmarks. This contribution is important as it illustrates how evolutionary optimization techniques can complement deep learning, enhancing robustness and making recurrent models more reliable in real-world PV operation scenarios.

Arias Velásquez [8] investigates NeuralProphet, a neural extension of the classical Prophet time-series model, applied to short-term PV power forecasting. NeuralProphet combines seasonality and trend decomposition with neural network components, enabling the model to handle both deterministic structures and nonlinear fluctuations in solar generation. The case study demonstrated that the approach can achieve competitive accuracy while offering greater interpretability than purely black-box deep learning models, an aspect highly valued in operational contexts where forecast transparency is required for grid integration. This work underscores the potential of hybrid frameworks that combine the interpretability of statistical models with the adaptability of deep learning, thereby bridging two methodological paradigms in solar forecasting.

Khan et al. [9] present a dual-stream network augmented with an attention mechanism, explicitly designed for photovoltaic forecasting tasks. The architecture processes spatial and temporal information in parallel streams, while the attention layer adaptively assigns weights to the most informative features. This design allows the model to emphasize patterns most relevant to energy generation and to suppress noise or redundant inputs. Experimental validation showed that the dual-stream attention network outperformed standard CNN and RNN baselines, achieving lower forecast errors across several case studies. This contribution is significant in that it reflects the broader trend of incorporating attention mechanisms into PV forecasting, paralleling their transformative impact in natural language processing and computer vision.

Azizi et al. [10] shift the focus from short-term horizons to long-term forecasting of global irradiance and temperature for a 20 MW PV plant in Iran. Their study employed a range of deep learning architectures—including MLP, LSTM, GRU, CNN, and CNN-LSTM—to develop a multivariate, multi-step forecasting framework. By integrating multiple climatic variables and extending the forecasting horizon, the authors demonstrated that deep learning models can maintain predictive accuracy even in long-

term settings where variability and uncertainty are higher. This work is important because it broadens the application of deep learning beyond the traditional short-term domain, showing its viability for long-range planning and operational decision-making in utility-scale PV systems.

Finally, Kim et al. [11] provide a comprehensive comparative study of statistical and deep learning methods using South Korean solar datasets from 2017 to 2021. The analysis included Holt–Winters, ARIMA, SARIMA, and LSTM models. Results indicated that LSTM consistently delivered the lowest forecast errors, particularly under conditions of rapidly fluctuating irradiance, outperforming traditional statistical approaches across multiple test scenarios. The study offers empirical justification for adopting LSTM in operational contexts where sufficient historical data are available, thereby reinforcing the practical advantage of deep learning models over conventional baselines.

Viewed as a whole, these contributions demonstrate the maturation of deep learning in PV forecasting—beginning with convolutional–recurrent hybrids such as PV-Net, progressing through evolutionary-optimized recurrent designs like NARX-GA, and advancing to interpretable frameworks such as NeuralProphet and attention-based dual-stream networks. More recent work extends the scope to multivariate long-term formulations and comparative evaluations that confirm the practical advantage of LSTM architectures over statistical baselines. Collectively, this progression highlights how deep learning has evolved toward solutions that are not only more accurate but also increasingly robust, interpretable, and operationally relevant across different forecasting horizons.

Building on the advances of convolutional–recurrent hybrids, evolutionary optimization, and attention-based mechanisms, more recent research has expanded the scope of deep learning applications in solar forecasting. A distinctive feature of these contributions is the explicit integration of spatial–temporal correlations, the incorporation of satellite imagery and climate variables, and the development of hybrid and physics-informed models that address the limitations of purely data-driven learning. This stream of studies not only extends deep learning to different forecasting horizons, from intra-hour to day-ahead and long-term, but also demonstrates how combining neural architectures with clustering, ensemble methods, or domain knowledge can enhance robustness and interpretability. Within this context, the following works published between 2024 and 2025 [12–18] illustrate the methodological diversification of deep learning and its increasing alignment with practical forecasting challenges in distributed and large-scale PV systems.

Cui et al. [12] exemplify the integration of novel input sources by employing geostationary satellite imagery in deep learning architectures such as DGMR-SO and UNet. Their framework moves beyond traditional ground-based observations by capturing cloud dynamics in near real time, substantially reducing errors in solar radiation nowcasting. The study illustrates how the inclusion of spatial image data complements temporal modeling, setting a precedent for multimodal approaches in ultra-short-term forecasting.

Building on the idea of capturing spatial structure, Lai et al. [13] extend deep learning forecasting to distributed PV networks by introducing a sub-region division method that explicitly accounts for spatio-temporal correlations among installations. Their results show that localized forecasting models outperform centralized approaches, especially in fragmented PV systems with heterogeneous characteristics. This contribution emphasizes the growing importance of geographically adaptive strategies in managing distributed renewable energy resources.

While Lai et al. focus on distributed systems, Xu et al. [14] address the long-term horizon by proposing a complementary fusion of GRU and XGBoost. This hybrid

ensemble leverages the sequential modeling strengths of recurrent networks while harnessing the noise resilience and interpretability of gradient boosting. By demonstrating improved stability and accuracy over extended horizons, their work highlights the potential of combining deep learning with machine learning ensembles to overcome the degradation typically observed in long-term forecasts.

A different innovation is introduced by Li et al. [15], who target ultra-short-term irradiance forecasting using a K-means-ELM model. Their approach applies clustering to identify characteristic weather patterns before using an Extreme Learning Machine for prediction. This preprocessing step improves adaptability to rapid fluctuations and ensures computational efficiency, making the method well suited for operational contexts that require near real-time forecasts. In this way, the study illustrates how the integration of unsupervised learning with fast neural algorithms can yield practical forecasting tools for grid operators.

Further advancing the integration of meteorological information, Ma et al. [16] develop the D-Informer architecture, which combines attention mechanisms and differential transformations to explicitly embed climate variables such as temperature, humidity, and pressure into the forecasting process. By leveraging these atmospheric drivers, the model achieves superior accuracy under weather-sensitive conditions, underscoring the importance of climate-aware neural forecasting frameworks in operational planning.

The shift from short-term and weather-driven designs to regional and market-relevant horizons is represented by Perera et al. [17], who propose hierarchical temporal convolutional networks (HTCNN) for day-ahead regional forecasting. Their approach jointly processes aggregated and site-level data, thereby capturing multi-scale dependencies that flat models fail to represent. Experimental validation demonstrates that hierarchical learning enhances accuracy in large-scale forecasting tasks, reinforcing the operational value of day-ahead predictions for energy markets and system scheduling.

Finally, Han et al. [18] contribute a hybrid approach for intra-hour forecasting that combines topological data analysis with physics-informed deep learning. This framework extracts structural features of irradiance variability while incorporating physical constraints of solar generation, producing forecasts that are not only accurate but also interpretable under highly stochastic conditions. Such physics-aware designs represent a promising direction for bridging the gap between purely data-driven models and the operational transparency required by grid operators.

As the evidence accumulates, a pattern emerges in deep learning-based solar forecasting: the transition moves from the incorporation of new data modalities (satellite imagery, distributed PV correlations) to hybridizations with clustering and boosting methods, and finally toward climate-aware and physics-informed architectures. This trajectory not only delivers consistent improvements in accuracy across different horizons but also enhances robustness, scalability, and interpretability—qualities that are increasingly critical for integrating PV generation into modern electricity markets.

In parallel with earlier advances in spatial-temporal modeling and climate-aware neural designs, the latest body of research increasingly emphasizes multimodal, hybrid, and probabilistic frameworks. These approaches aim to exploit heterogeneous data inputs—such as ground-based cloud imagery, meteorological time series, and contextual climate variables—while simultaneously addressing two critical challenges: ensuring scalability across diverse PV systems and providing explicit quantification of forecast uncertainty. The most recent works published in 2025 [19–23] exemplify this orientation, introducing deep clustering strategies, advanced hybrid learning schemes, multimodal fusion architectures, transformer-based feature enhancements, and probabilistic

forecasting formulations that together represent the current research frontier in photovoltaic power prediction.

Within this line of work, Dou et al. [19] foreground multimodality for day-ahead irradiance forecasting by coupling ground-based cloud imagery with meteorological time-series through a deep clustering framework. The approach first structures heterogeneous inputs via representation learning and cluster assignment, then fuses image- and sequence-derived features for prediction. This design addresses two persistent issues—noisy visual inputs and regime heterogeneity—by letting the model discover weather regimes that mediate how visual and temporal cues should be combined. The study is retained because it establishes a principled route for regime-aware multimodal fusion at a day-ahead horizon, a setting where image information is often underexploited.

Extending the emphasis on complementarity, Song et al. [20] present an advanced hybrid deep-learning pipeline aimed at improving point accuracy for PV power prediction. Rather than relying on a single architecture, the method stages feature extraction, temporal modeling, and fusion/ensembling to capture interactions among meteorological drivers and historical power signals. The value of this contribution lies in demonstrating how carefully engineered hybrid stacks can translate into consistent error reductions across datasets, thereby offering a practical blueprint for utilities seeking accuracy gains without committing to a single “all-purpose” network.

Abad-Alcaraz et al. [21] reinforce the case for multimodal learning in solar radiation forecasting by formalizing feature-level and/or decision-level fusion within a unified deep model. By treating radiation-relevant covariates (e.g., irradiance proxies, meteorology, image or contextual signals) as complementary views, the architecture learns cross-modal dependencies that single-stream models tend to miss. This paper is included because it provides a clear, generalizable template for multimodal fusion, helping explain when and why heterogeneous inputs yield measurable accuracy gains.

On the architectural frontier, Liu et al. [22] refine transformer-based forecasting with targeted feature enhancement for short-term PV power prediction. The model augments attention with learnable feature refinement (e.g., gating/selection or cross-feature interactions), improving the network’s ability to prioritize informative signals under rapidly changing conditions. This work is representative of a broader movement to adapt transformers to energy time-series by controlling feature noise and improving data efficiency—an essential step toward robust deployment beyond image or text domains.

Finally, Song et al. [23] move from pure point prediction to probabilistic ultra-short-term forecasting by combining attention-enhanced neural representations with natural-gradient boosting. The hybrid yields calibrated predictive distributions—rather than single values—thus addressing the operational need to quantify uncertainty for reserve scheduling and risk-aware bidding. We retain this study because it exemplifies how modern DL can be integrated with probabilistic learners to deliver both accuracy and well-behaved uncertainty, closing a gap identified in earlier literature.

Overall, the reviewed contributions trace a coherent methodological arc in deep learning-based solar forecasting—beginning with multimodal fusion strategies that integrate diverse data sources such as cloud imagery and meteorological time series, advancing through hybrid pipelines and generalized multimodal architectures that strengthen cross-modal representation and accuracy, and culminating in transformer-based refinements and probabilistic formulations that deliver calibrated uncertainty estimates. While these advances have significantly improved predictive accuracy, scalability, and robustness, purely deep learning-based approaches continue to face persistent challenges, particularly in terms of interpretability, adaptability across heterogeneous PV systems, and reliance on large-scale training datasets. To address these gaps, a complementary research stream has emerged around hybrid models that integrate

neural architectures with fuzzy logic, statistical formulations, ensemble methods, predictive control, or physics-based corrections. By combining the nonlinear representation capacity of machine and deep learning with the robustness, transparency, and domain-awareness of traditional approaches, hybrid frameworks aim to deliver forecasts that are not only accurate but also interpretable, adaptive, and operationally reliable across diverse climatic and market contexts.

The trajectory of this research begins with efforts to explicitly model uncertainty and interpretability. Li and Liu [24] introduce interval type-2 Takagi–Sugeno–Kang fuzzy systems for short-term PV power prediction, where fuzzy logic captures uncertainty while neural components approximate nonlinear patterns. This design provides interval forecasts rather than point estimates, offering decision-makers confidence bounds better suited for grid operations under variable irradiance. Building on the goal of interpretability, Sehwat et al. [25] combine digital twin technology with machine learning to forecast solar irradiance. By creating virtual replicas of PV systems that are continuously updated with real data, the framework ensures forecasts remain system-specific and adaptive, representing a shift toward hybrid digital environments where physical system knowledge is directly embedded in predictive workflows.

Moving from interpretability to ensemble stabilization, Abumohsen et al. [26] propose a CNN–LSTM–RF model, in which convolutional and recurrent networks extract spatio-temporal patterns, while a Random Forest ensemble improves generalization and reduces overfitting. Their study demonstrates how blending deep and tree-based learners can achieve more robust predictions than either family of methods alone. A different hybrid pathway is illustrated by Mbungu et al. [27], who frame solar forecasting within a predictive control paradigm. Here, forecasts are continuously refined in response to deviations between expected and actual outputs, highlighting the operational role of hybrid forecasting not only in accuracy improvement but also in real-time responsiveness—a critical attribute for grid management.

The importance of weather regime adaptation is emphasized by Dai et al. [28]. Their method combines credibility prediction for weather types with a dynamic ensemble of forecasting models, assigning higher weights to those models best suited for current meteorological conditions. This approach reflects an evolution toward adaptive hybridization, where models are not fixed but context-sensitive to the prevailing weather regime. Extending to signal decomposition and enriched feature design, Bai et al. [29] introduce a hybrid model that incorporates wavelet packet decomposition and an improved similar-day method before feeding data into an LSTM predictor. By disentangling multi-scale frequency components and enhancing input selection, their model achieves superior day-ahead forecasts, particularly under fluctuating irradiance conditions.

At a larger scale, Dou et al. [30] demonstrate how numerical weather prediction (NWP) data can be effectively hybridized with machine learning. Their framework disentangles seasonal and trend components of NWP outputs and corrects them with a mixture-of-experts (MoE) model, bridging the gap between meteorological physics and data-driven refinement. Finally, Pereira et al. [31] present one of the most comprehensive physics-informed hybrid frameworks, combining deterministic solar radiation models with data-driven predictors. By embedding physical laws into the learning process, their approach prevents unrealistic outputs, reduces error propagation, and strengthens interpretability—an attribute increasingly demanded by system operators.

These contributions outline a clear evolutionary path in hybrid solar forecasting research. The trajectory begins with fuzzy and digital twin frameworks designed to enhance interpretability, progresses through ensemble- and control-based hybrids that improve robustness and adaptability, and culminates in physics-informed and NWP-

enhanced architectures that integrate domain knowledge with machine intelligence. This progression highlights the pivotal role of hybrid models as a bridge between purely data-driven deep learning and operationally transparent forecasting systems, providing solutions that are not only accurate but also robust, interpretable, and practical for deployment in modern electricity markets.

Alongside the advances in hybrid and deep learning frameworks, a further strand of research has increasingly focused on physico-statistical and probabilistic models, reflecting the growing importance of uncertainty quantification and interpretability in PV forecasting. Unlike purely deterministic approaches, these models explicitly characterize the stochastic nature of solar irradiance and power generation, often combining statistical formulations with physical insights or probabilistic inference. Their value lies not only in providing point forecasts but also in generating confidence intervals, probability distributions, or scenario sets that can be directly incorporated into grid operation, reserve scheduling, and market bidding strategies.

This research stream has gained momentum in recent years, particularly as system operators and market regulators place greater emphasis on risk-aware decision-making and the reliable integration of high shares of variable renewable energy. The selected contributions [32–37] illustrate the breadth of methodological innovation in this area, spanning diffusion-based generative models, weather-informed probabilistic forecasting, hidden Markov formulations, copula-based temporal decomposition, and ensemble-based Gaussian mixture networks. Together, these studies exemplify how probabilistic and physico-statistical approaches complement machine and deep learning by enhancing transparency, capturing uncertainty, and aligning forecasting outcomes with the operational requirements of modern electricity markets.

Huang et al. [32] advance the state of probabilistic forecasting by introducing an enhanced conditional diffusion model tailored for net load prediction in grids with high renewable penetration. Diffusion-based generative modeling, widely adopted in computer vision, is here adapted to the energy domain, enabling the model to capture complex probability distributions of load and PV generation under uncertainty. Unlike deterministic predictors, the diffusion framework produces diverse, calibrated scenarios that can inform reserve planning and reliability assessment. This contribution is significant as it marks one of the first attempts to apply conditional diffusion techniques to renewable-heavy power systems, showing how generative models can be leveraged to improve both accuracy and probabilistic calibration.

Zhang et al. [33] complement this direction with a weather-informed probabilistic framework that integrates scenario generation into day-ahead system operation. By coupling meteorological forecasts with probabilistic learning, their approach not only predicts PV output but also generates scenario sets consistent with weather uncertainty, directly usable in stochastic optimization for unit commitment and market bidding. This work is notable because it bridges the gap between forecasting and operational decision-making, highlighting how scenario-based probabilistic outputs can enhance system flexibility and resilience in high-VRE environments.

Ahmad et al. [34] propose a hybrid strategy that combines deterministic forecasts with the NB-DST probabilistic enhancement method. Their framework integrates Neural Bayesian (NB) inference with a Deterministic–Stochastic Transformation (DST), yielding both accurate point forecasts and well-calibrated probability distributions. The results demonstrate that augmenting deterministic predictors with probabilistic layers substantially improves reliability, especially under volatile irradiance. This study illustrates how deterministic deep learning models can be extended into the probabilistic domain without compromising accuracy, offering a pragmatic pathway for operators accustomed to conventional forecasting tools.

Zhang and Shang [35] take a different route, focusing on interpretability through a multi-observation non-homogeneous Hidden Markov Model (HMM). By modeling solar power generation as a sequence of hidden states influenced by exogenous meteorological variables, their approach delivers fast and interpretable probabilistic forecasts. Unlike black-box neural models, the HMM structure allows operators to directly associate hidden states with physical regimes (e.g., clear sky, partial cloud, overcast), thereby enhancing transparency while maintaining competitive predictive performance. This contribution is particularly relevant in regulatory and operational contexts where explainability is as critical as accuracy.

Wang et al. [36] present a copula-based approach, combining temporal decomposition with vine copula functions to model dependency structures in solar power time series. By first decomposing the PV output into trend and fluctuation components, and then capturing nonlinear dependencies via vine copulas, the method generates probabilistic forecasts that respect temporal correlations across different horizons. Their results show improved calibration and sharpness compared to traditional Gaussian-based approaches, underscoring the importance of advanced statistical tools for capturing joint variability in renewable energy time series.

Doelle et al. [37] extend the probabilistic forecasting paradigm by leveraging ensembles of deep Gaussian mixture density networks (GMDNs) for intraday PV prediction. Unlike classical probabilistic regressors, GMDNs directly estimate conditional probability densities, allowing the generation of full predictive distributions rather than single-point estimates. The ensemble design further improves robustness by reducing variance and mitigating overfitting, while the mixture density formulation captures multimodal uncertainty inherent in rapidly changing irradiance conditions. Their findings demonstrate that deep probabilistic ensembles can achieve high calibration quality and sharpness, making them particularly suitable for intraday horizons where uncertainty quantification is most critical for grid balancing and reserve allocation.

The evidence from [32–37] underscores the distinctive role of probabilistic and physics–statistical approaches in advancing photovoltaic forecasting. Diffusion-based generative models enable calibrated scenario generation, weather-informed and Bayesian–deterministic hybrids enhance operational usability, hidden Markov structures provide transparent links between statistical states and physical regimes, copula-based formulations capture nonlinear temporal dependencies, and deep Gaussian mixture ensembles demonstrate how probabilistic learning can be embedded into neural architectures. What unites these methods is their explicit capacity to quantify uncertainty and produce scenario-based outputs that are both interpretable and operationally relevant. In this way, probabilistic and physics–statistical frameworks complement neural and hybrid deep learning models by addressing the challenges of calibration, robustness, and transparency, thereby offering tools that are increasingly indispensable for reliable system operation in renewable-dominated power grids.

A distinct stream of research is represented by survey and comparative studies, which provide meta-analyses of methodological progress and consolidate lessons from the diverse body of work reviewed above. Rather than focusing on single architectures, these contributions synthesize statistical, machine learning, hybrid, and probabilistic approaches, while also mapping their operational implications for modern energy systems.

Di Leo et al. [38] present a comprehensive review of advancements and challenges in PV forecasting, offering a structured analysis of statistical, machine learning, deep learning, and hybrid methodologies. Their findings emphasize that forecasting performance depends not only on algorithmic sophistication but also on the quality of meteorological inputs, preprocessing strategies, and horizon-specific model adaptation.

Importantly, the authors argue that future improvements will require integrated frameworks that couple forecasting accuracy with practical constraints of grid operation, market bidding, and reserve scheduling.

Yu et al. [39] focus specifically on deep learning models, providing one of the most detailed reviews of convolutional, recurrent, attention-based, and hybrid neural networks for PV power forecasting. Their analysis demonstrates that deep learning has rapidly become the dominant paradigm due to its superior capacity for capturing spatio-temporal dependencies and handling heterogeneous data. However, they also stress unresolved challenges, including limited interpretability, difficulties in generalizing across heterogeneous PV systems, and the high computational cost of training advanced architectures.

Blazakis et al. [40] extend the discussion by conducting an empirical evaluation of one-day-ahead solar irradiation and wind speed forecasting using state-of-the-art deep learning techniques. Their results confirm the competitive advantage of advanced neural models over traditional statistical methods but also reveal vulnerability to performance degradation under extreme weather variability. This underscores the importance of integrating deep learning with hybrid and probabilistic refinements to achieve consistent reliability across diverse operating conditions.

Delgado et al. [41] examine the integration of Hidden Markov Models (HMM) with Long Short-Term Memory (LSTM) networks under both single-input and multiple-input configurations. Their study illustrates how combining interpretable statistical state-space representations with recurrent neural architectures can improve both robustness and transparency. This line of work points to the potential of hybrid architectures that preserve interpretability while retaining the nonlinear representation capacity of deep learning.

Beyond PV-specific applications, Lim et al. [42] provide a broader review of deep learning in power system decision-making. Their study situates PV forecasting within the larger ecosystem of energy management, unit commitment, and reliability assessment, stressing that forecasting models must ultimately be evaluated by their capacity to inform operational and market-level decisions. The authors highlight the growing need for explainable artificial intelligence (XAI) and risk-aware frameworks, which can bridge the gap between black-box predictors and decision-making requirements in regulated energy markets.

Kousounadis-Knousen et al. [43] complement this perspective by focusing on scenario generation methods for solar forecasting, with particular attention to weather classifications, temporal horizons, and the application of deep generative models. Their review highlights how scenario-based forecasting enables stochastic optimization and robust decision-making under uncertainty, offering practical pathways for integrating probabilistic forecasts into energy market operations and grid reliability assessments.

Viewed collectively, the survey and comparative contributions [38–43] show that while methodological innovations—ranging from statistical baselines to advanced neural and hybrid models—have substantially improved forecasting accuracy, persistent gaps remain at the interface between algorithms and operational practice. These reviews converge on several key points: the necessity of high-quality and diverse input data; the value of hybridization to balance accuracy, interpretability, and robustness; and the importance of probabilistic and scenario-based outputs for risk-aware system management. In this way, meta-analyses provide not only a synthesis of methodological evolution but also a roadmap for future research directions, linking algorithmic advances to the practical demands of modern electricity markets.

As highlighted by the reviewed literature, reliable forecasting of solar power generation remains especially critical for short-term resource planning, power

dispatching, and ensuring the operational safety of energy systems. While most research efforts concentrate on improving short-term prediction methods—primarily through weather-informed deep learning architectures—existing approaches still fall short in effectively addressing the problem of forecasting the average daily energy yield of PV plants. This gap is of particular importance for operational interaction with the power grid and for maximizing plant revenues under market conditions.

To address this challenge, the present research aims to develop methods for estimating statistical indicators of daily energy yield based on the probabilistic characterization of meteorological conditions. Specifically, the objectives include: developing a mathematical model of hourly PV generation that accounts for geographical location and key meteorological drivers; constructing a model of daily energy yield derived from hourly generation; justifying the appropriate probability density law for meteorological inputs as random variables; and deriving analytical expressions for the main statistical indicators of daily yield as functions of the probability density parameters.

3. Research Materials

3.1. Calculation of the Day Energy Yield of a Solar Panel

The expression for calculating the daily solar panel generation, considering the geographical location and meteorological conditions of the atmosphere, is described by the well-known expression [44]:

$$E = S \cdot \int_{t_{sunrise}}^{t_{sunset}} I_0(t) \cdot f_{atm}(H, h, c) \cdot \cos(\theta(t, LAT, LON)) dt, \quad (1)$$

where

E is a day energy yield (W·h);

S is a solar panel area (m²);

$t_{sunrise}, t_{sunset}$ are moments of sunrise and sunset (hours);

$I_0(t)$ is the solar constant (~1361 W/m²), adjusted for the time of day and day of the year;

$f_{atm}(H, h, c)$ is an atmospheric transmittance coefficient, depending on the altitude H , humidity h and cloudiness c ;

$\theta(t, LAT, LON)$ is the angle of incidence of sun rays on the horizontal surface, which depends on the time of the day, latitude, and longitude of the location.

Let us give formulas for calculating the components of Equation (1).

3.1.1. Atmospheric Transmittance Coefficient

Atmospheric conditions including humidity, cloudiness and solar panel altitude above sea level are approximately taken into account as follows [44,45]:

$$f_{atm}(H, h, c) = (1 - 0.75 \cdot c^{3.4}) \cdot (1 - 0.1 \cdot h) \cdot (1 + 0.0001 \cdot H) \quad (2)$$

At low altitude of solar panel installation above sea level H (up to 500 m) the last factor in Equation (2) can be neglected. We also neglect the effect of humidity h .

$$f_{atm}(h, c) = (1 - 0.75 \cdot c^{3.4}) \cdot (1 - 0.1 \cdot h) \approx (1 - 0.75 \cdot c^{3.4}) \quad (3)$$

As can be seen from Figure 1, the value of the atmospheric transmittance coefficient is decisively influenced by the “Cloudiness” parameter. Later, Humidity’s influence can be neglected.

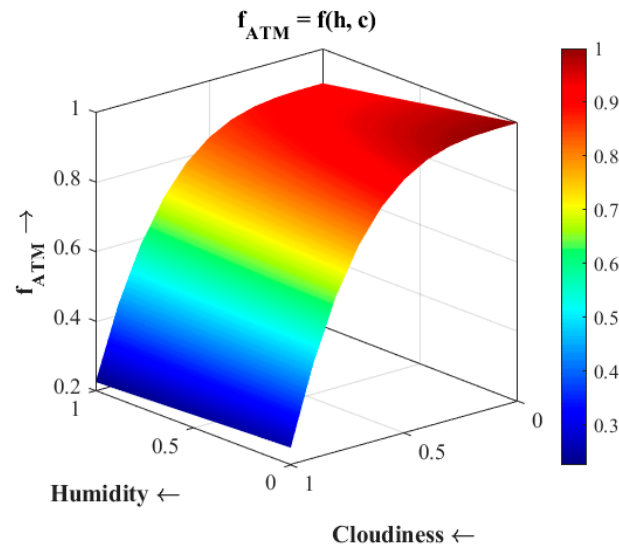


Figure 1. Dependence of the atmospheric transmittance coefficient on Humidity and Cloudiness.

3.1.2. Angle of Incidence of Solar Rays (θ)

$$\cos(\theta) = \sin(\delta) \cdot \sin(\varphi) + \cos(\delta) \cdot \cos(\varphi) \cdot \cos(\omega), \quad (4)$$

where

δ is declination of the Sun (calculated depending on the date);

φ is latitude of the location (LAT);

ω is an hour angle (depends on the time of day and longitude LON).

3.1.3. Solar Constant I_0

The value of the solar constant $I_{sc} \approx 1361 \text{ W/m}^2$ is used with a correction for the Earth-Sun distance, depending on the day of the year:

$$I_0 = I_{sc} \left(1 + 0.034 \cdot \cos \left(\left(\frac{2\pi}{365} \cdot N \right) \right) \right) \quad (5)$$

where N is the day number of the year.

This expression takes into account the main factors that influence the daily generation of solar panels, allowing for accurate calculations taking into account climatic and geographical features.

3.1.4. Sunrise and Sunset Times ($t_{sunrise}, t_{sunset}$)

The hourly sunrise/sunset angle (ω_s) is determined by the formula:

$$\cos(\omega_s) = -\tan(\varphi) \cdot \tan(\delta) \quad (6)$$

Then

- $t_{sunrise} = 12 - \omega_s/15$, hourly time of sunrise;
- $t_{sunset} = 12 + \omega_s/15$, hourly time of sunset.

Here φ is the latitude, δ is the declination of the Sun.

The declination of the Sun (δ) is calculated as follows:

$$\delta = 23.45^\circ \sin \left(\frac{360}{365} \cdot (284 + N) \right) \quad (7)$$

Let us present the results of calculations of hourly power generation by a solar panel throughout the day, under the following conditions: area of the panel— 1 m^2 ; humidity

$h = 0.1$; cloudiness $c = 0.1$; panel tilt angle 30° ; panel efficiency 18%. The results of the performed calculations are illustrated in Figure 2, which shows the hourly distribution of solar panel electricity production for two representative days of the year.

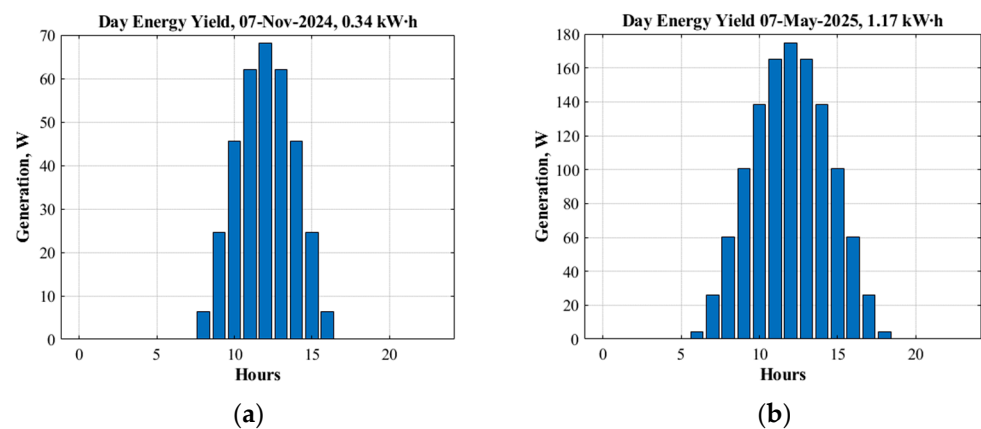


Figure 2. Result of the calculation of the solar panel’s hourly electricity production over a day: (a) 7 November 2024; (b) 7 May 2025.

These results were obtained under constant cloud cover and atmospheric humidity.

Below are the results of experimental studies of solar radiation carried out using the meteorological station instruments at the solar power plant in Saran, Kazakhstan [46]. The solar power plant occupies an area of 160 ha in the northeastern part of the city of Saran, at geographic coordinates (49.8138, 72.8256). The site elevation is 491 m above sea level, and the official time zone is UTC+5. The panel tilt angle is 30° . Measurements were performed with a Kipp & Zonen SMP10-V spectrally flat Class A pyranometer (Figure 3), configured for GHI (Global Horizontal Irradiance). The data logging interval is 1 hour.



Figure 3. Pyranometer Kipp & Zonen SMP10-V.

Main technical characteristics of Kipp & Zonen SMP10-V pyranometer: analog outputs—0–1 V; directional response— $<15^\circ$ W/m²; irradiance saturation—4000 W/m²; operating temperature range— -40 to $+80^\circ\text{C}$; spectral accuracy—285 to 2800 nm.

As the initial data, the company’s management provided access only to information on solar radiation registration.

For the approximate pipeline $\text{GHI} \rightarrow \text{DNI/DHI} \rightarrow \text{POA} \rightarrow \text{DC} \rightarrow \text{AC}$, based on the hourly GHI profile, we make the following assumptions: the Sun’s position is calculated according to standard astronomy (e.g., NREL SPA) [47]; for the decomposition $\text{GHI} \rightarrow \text{DNI/DHI}$, the Erbs model is used, [48]; for transposition onto the module plane (POA), the isotropic sky model [49] is chosen; albedo is used (0.15 for soil/vegetation); the DC module model is assumed to be linear with respect to irradiance and temperature; cable/connector losses are considered as constant DC losses; and the inverter efficiency and DC/AC ratio are assumed constant.

The complete MATLAB 2025b program for calculating the actual daily energy of a solar panel, implementing the specified calculations, is given in Appendix A.

Figure 4 shows the results of the experimental study of hourly solar radiation during the day, as well as the results of calculating the actual daily energy of a solar panel.

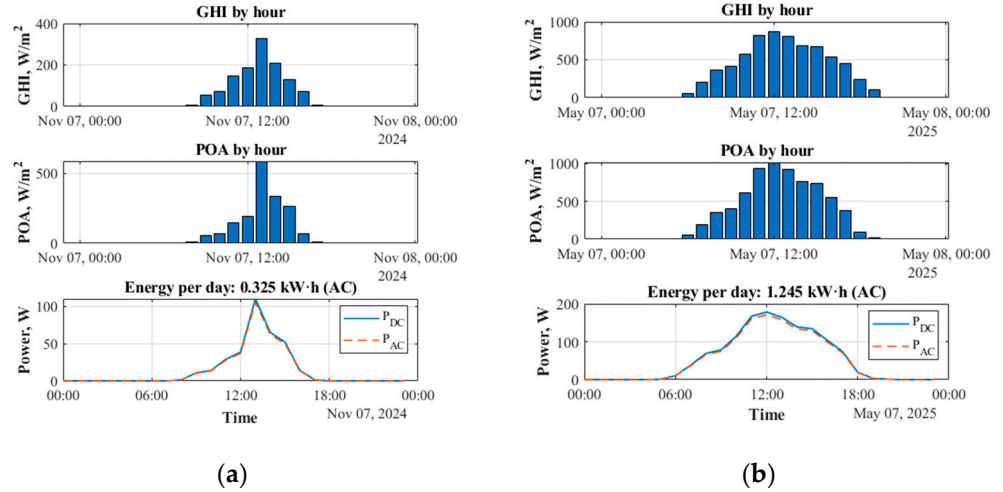


Figure 4. Experimental studies of hourly solar radiation during the day: (a) 7 November 2024; (b) 7 May 2025.

Attention is drawn to the asymmetry of the GHI solar radiation histograms, Figure 4. In addition, the analytical results of the hourly solar panel output calculation, Figure 2, do not coincide with the results of the experimental study, Figure 4. This can be explained by changes in cloudiness during the day.

The following discussion is devoted to the development of methods for accounting for cloudiness as a probabilistic process in the calculations of solar panel power generation.

3.2. Analytical Determination of Daily Energy Yield

Analytical integration of the given expression (1) is possible if we accept some simplifications:

- The solar constant I_0 by (5) is considered constant throughout the day (for a specific day N).
- To consider the symmetry of the sun's position relative to solar noon.
- The atmospheric transmittance coefficient $f_{atm}(c)$ can be considered constant throughout the day.

Last assumption can be reasonably substantiated as follows. In the present study, our primary concern is the value of the definite integral over a prescribed time interval. The parameters of the β -distribution may likewise be interpreted as integral descriptors of cloudiness. Accordingly, it is permissible to approximate the atmospheric transmittance coefficient not as a time-dependent function, but rather as a constant defined by the β -distribution parameters corresponding to the specified calculation date.

Let us transform the Formula (1) under the assumption of a constant atmospheric transmittance coefficient and taking into account (3):

$$E = S \cdot I_0(t) \cdot f_{atm}(c) \cdot \int_{t_{sunrise}}^{t_{sunset}} \cos(\theta(t, LAT, LON)) dt \quad (8)$$

Since the function $\cos(\theta)$ changes sinusoidally during the day, the integral can be simplified to the following analytical expression.

The final formula after integration of (8):

$$E = \alpha \cdot S \cdot I_0(t) \cdot f_{atm}(c) \cdot \frac{24}{\pi} [\sin(\delta) \cdot \sin(\varphi) \cdot \omega_s + \cos(\delta) \cdot \cos(\varphi) \cdot \sin(\omega_s)] \quad (9)$$

where

S is the area of the panel, m^2 ;

I_0 is an extraterrestrial solar constant for day N ;

$f_{atm}(h, c)$ is an average daily atmospheric transmittance;

φ is the latitude of the location;

δ is the declination of the Sun;

ω_s is an angle of sunrise (sunset) in radians.

α is the coefficient accounting for panel soiling and degradation.

Thus, with the indicated simplifications, it is possible to make analytical integration and obtain the final formula that can be used for an approximate calculation of the daily generation of a solar panel.

Let us compare the results of the exact (1) and approximate (9) formulas for calculating daily solar generation.

From the analysis of the graphs in Figure 5, we can conclude that the approximate model for calculating solar generation has satisfactory accuracy. The mean relational error value in the calculation using the approximate model (9) was 3.9%.

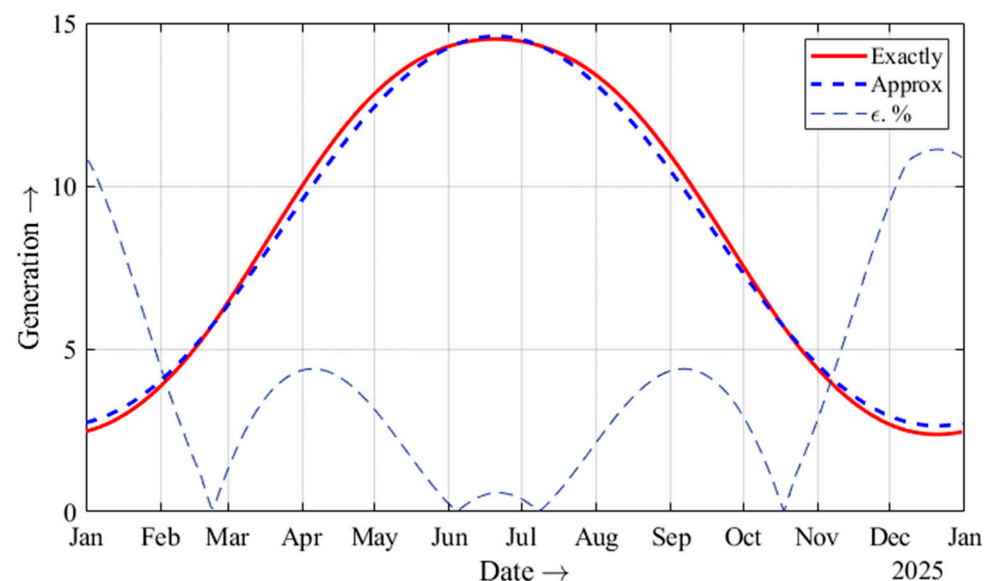


Figure 5. Comparison of the exact (1) and approximate (9) methods for calculating daily solar generation.

The main reason for the calculated data deviation according to Equation (9) is that the calculations according to (1) consider the daily movement of the Sun and the change in the solar panel's illumination level more accurately.

3.3. Physical and Statistical Nature of the Variable "Cloudiness"

Cloudiness (c) is a continuous random variable taking values in the interval $[0, 1]$, where

$c = 0$ —absolutely clear;

$c = 1$ —fully cloudy.

This makes it a typical candidate for modeling using random variable distributions on the interval $[0, 1]$, where the β -distribution is the most universal [50].

β -distribution of a random variable X is given by the probability density f_X , which has the following form:

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (10)$$

where α and β are distribution parameters.

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx - \text{beta-function.}$$

In this case, the random variable X has a β -distribution. Formally, this fact is written by the expression $X \sim B(\alpha, \beta)$.

The beta-distribution depends on the values of the distribution parameters α and β , and can take various forms:

- Unimodal (bell-shaped),
- U-shaped (often clear or often cloudy), Figure 6a,
- Uniform (when $\alpha = \beta = 1$),
- Shifted to the left/right (if the weather is often clear or cloudy), Figure 6b.

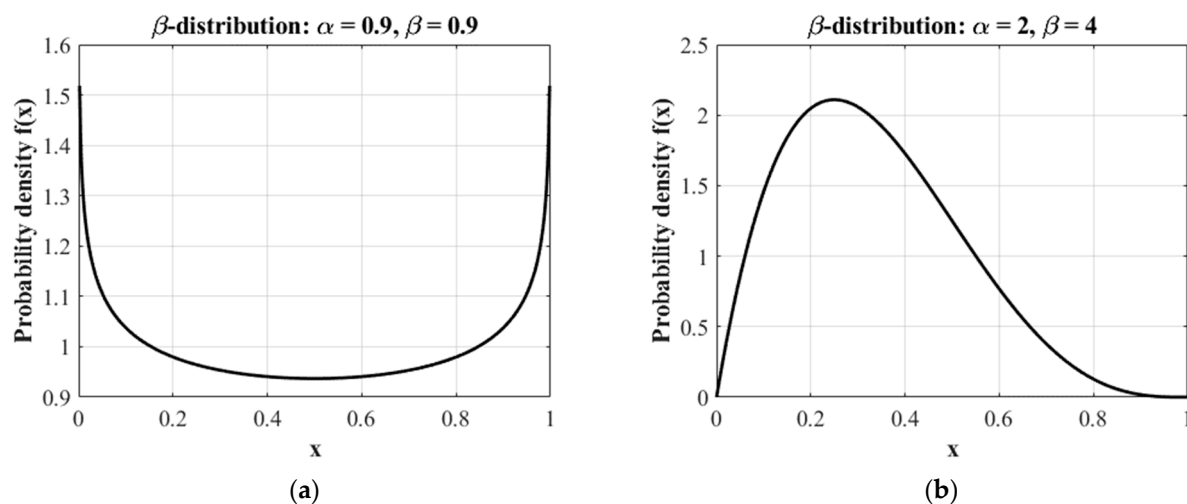


Figure 6. Examples of β -distribution for different values of distribution parameters: (a)—U-shaped distribution; (b)—shifted to the left (often clear weather).

This makes it easy to adjust the distribution to the climatic features of the region.

As an alternative to the β -distribution, the probability density function of the Kumaraswamy distribution can be applied to model a random variable taking values in the interval $[0, 1]$. It is an analog of the β -distribution but is simpler for analytical work due to the closed-form expression of its distribution function.

If a random variable X follows the Kumaraswamy distribution with parameters $a > 0, b > 0$, its probability density function (PDF) is given by:

$$f(x; a, b) = abx^{a-1}(1-x^a)^{b-1}, 0 < x < 1$$

where

a — shape parameter;

b — shape parameter.

Similarly to the β -distribution, it flexibly describes various density shapes on $[0, 1]$ (U-shaped, increasing, decreasing, bell-shaped).

The main advantage is the simple closed form of the cumulative distribution function (CDF), which is convenient for generating random numbers using the inverse transform method.

In the case of bimodal weather with sharply variable cloudiness during the day, the β -distribution may be insufficient to describe multi-peaked probability density functions.

In such cases, a Beta Mixture Distribution (BMD) can be used. The main areas of application of BMD include:

- Bayesian statistics (approximation of complex posterior distributions on $[0, 1]$);
- models for fractions, probabilities, and proportions (for example, the share of renewable energy in the energy balance, distribution of humidity, etc.).

In the further discussion, we will use the β -distribution, since integration with the β -function is relatively straightforward. The importance of this property will become evident when deriving the formulas for the statistical characteristics of the daily energy yield.

3.4. The Statistical Nature of Solar Panel Generation

The statistical nature of the variable “Cloudiness” which is a part of the expressions for the atmospheric transmittance coefficient, Equations (2) and (3), leads to the fact that the value of the atmospheric transmittance coefficient will also be a random variable. Since the variable “Cloudiness” is a part of expressions (2) and (3) in the form of a power dependence, the statistical characteristics of the atmospheric transmittance coefficient as a random variable will differ significantly from the characteristics of the β -distribution of the variable “Cloudiness”. The same applies to the daily power generation of the solar panel, calculated according to (9).

Determining the statistical characteristics of daily generation, which is formally a function of the random argument “Cloudiness”, is a non-trivial mathematical problem, the solution to which we will try to obtain below.

Let us consider the process of electric power $E(t)$ generation as a random function, that is, a function of its argument whose value for any value of the argument t is a random variable. If the argument of the random function t takes any values in a given interval, then the random function will also be a random process.

For conducting a statistical analysis of the daily energy yield of a solar power plant, it is necessary to address two interrelated research tasks. The first task is associated with determining the actual probability distribution law of the random variable “Cloudiness,” which has a decisive influence on the transmission of solar radiation through the atmosphere and, consequently, on the operation of photovoltaic modules. To achieve this goal, it is essential to establish the form of the probability density function describing cloudiness, identify the distribution parameters, and calculate its statistical characteristics, including the mean, variance, skewness, and kurtosis.

The second task concerns the investigation of the statistical distribution of the daily energy yield, considered as a random function of the parameter “Cloudiness.” Since cloudiness varies within the interval $[0, 1]$ and is described by a specific probability distribution, the resulting value of the daily energy yield also acquires a stochastic nature. In this regard, it becomes necessary to determine the probability distribution function and statistical characteristics of this random variable, including the mean value, variance, and the possible range of fluctuations.

Solving these tasks makes it possible to form a substantiated understanding of the statistical nature of electricity generation by solar power plants. This approach provides an opportunity for a more accurate description of uncertainties associated with the variability of meteorological conditions and establishes a basis for developing reliable probabilistic models for forecasting the daily energy yield of photovoltaic systems.

To address this problem, methods of mathematical statistics can be applied to determine the numerical characteristics of functions of random variables [50]. In the present case, the random variable $E(c)$ is a function of a single random variable.

Therefore, the task reduces to estimating the statistical characteristics of this random variable, provided that the distribution law of the random argument is known.

The general method for solving the problem is to find the distribution law of the function of a random argument when the distribution law of the argument is known.

The problem of finding the distribution law of the function $E(c)$ often turns out to be quite complex. However, for practical purposes it is quite sufficient to know only the numerical characteristics of this distribution, which significantly simplifies the solution of the problem.

Let us consider the problem of determining the numerical characteristics of a function of one random argument with a known law of its distribution.

There is a random variable γ with a known distribution law $\varphi(\gamma)$; another random variable, e , is related to γ by a functional dependence $e = \varphi(\gamma)$. Then, according to [50], the expected value of a function of one random argument can be determined by the formula:

$$M[e] = \int_{-\infty}^{+\infty} \varphi(\gamma) \cdot f(\gamma) d\gamma, \quad (11)$$

The variance of a function of one random argument can be determined by the formula:

$$D[e] = \int_{-\infty}^{+\infty} (\varphi(\gamma) - M[e])^2 f(\gamma) d\gamma, \quad (12)$$

3.5. Statistical Characteristics of Daily Energy Yield by a Solar Panel

In the considered case, the random argument is the “Cloudiness” parameter, denoted by the symbol c .

As it was justified above, the random variable c has a β -distribution with a probability density $f(c)$ of the form:

$$f(c) = \frac{1}{B(\alpha, \beta)} c^{\alpha-1} (1-c)^{\beta-1}. \quad (13)$$

Let us write down the functional dependence of the daily power generation by a solar panel on the “Cloudiness” parameter in expanded form.

Substituting the simplified expression for the atmospheric transmittance coefficient (3) into the expression for calculating the daily power generation by a solar panel (9), we obtain:

$$E = A \cdot f_{atm}(c), \quad (14)$$

where $A = \alpha \cdot S \cdot I_0(t) \cdot \frac{24}{\pi} [\sin(\delta) \cdot \sin(\varphi) \cdot \omega_s + \cos(\delta) \cdot \cos(\varphi) \cdot \sin(\omega_s)]$ is a constant depending on the day number of the year for the date of forecasting the daily generation.

Substituting Equation (3) into (14), we obtain

$$E = A \cdot (1 - 0.75 \cdot c^{3.4}). \quad (15)$$

Then, according to Equations (3) and (11), the expected value of daily generation can be determined by the formula:

$$M = A \cdot \frac{1}{B(\alpha, \beta)} \cdot \int_0^1 (1 - 0.75 \cdot c^{3.4}) \cdot c^{\alpha-1} (1-c)^{\beta-1} dc. \quad (16)$$

When deriving Formula (16), it was taken into account that the “Cloudiness” parameter varies within the range from 0 to 1.

Let us calculate the definite integral (16). We expand the integrand as follows:

$$(1 - 0.75 \cdot c^{3.4}) \times \\ \times c^{\alpha-1}(1-c)^{\beta-1} = c^{\alpha-1}(1-c)^{\beta-1} - 0.75 \cdot c^{3.4} \cdot c^{\alpha-1}(1-c)^{\beta-1}. \quad (17)$$

Let us integrate each term of (17):

$$\int_0^1 c^{\alpha-1}(1-c)^{\beta-1} dc = B(\alpha, \beta), \\ \int_0^1 0.75 \cdot c^{3.4} \cdot c^{\alpha-1}(1-c)^{\beta-1} = 0.75 \cdot B(\alpha + 3.4, \beta). \quad (18)$$

Combining the results, we obtain

$$M = A \cdot \frac{1}{B(\alpha, \beta)} \cdot (B(\alpha, \beta) - 0.75 \cdot B(\alpha + 3.4, \beta)). \quad (19)$$

or in an alternative form

$$M = A \cdot \left(1 - 0.75 \cdot \frac{B(\alpha + 3.4, \beta)}{B(\alpha, \beta)} \right). \quad (20)$$

According to Equations (3) and (12), the variance of daily generation can be determined by the formula:

$$D = \frac{1}{B(\alpha, \beta)} \cdot \int_0^1 (A \cdot (1 - 0.75 \cdot c^{3.4}) - M)^2 \cdot c^{\alpha-1}(1-c)^{\beta-1} dc. \quad (21)$$

By expanding the square in the integrand, we get

$$(A \cdot (1 - 0.75 \cdot c^{3.4}) - M)^2 = A^2(1 - 0.75 \cdot c^{3.4})^2 - 2 \cdot A \cdot M(1 - 0.75 \cdot c^{3.4}) + M^2. \quad (22)$$

and

$$(1 - 0.75 \cdot c^{3.4})^2 = 1 - 1.5 \cdot c^{3.4} + 0.5625 \cdot c^{6.8}. \quad (23)$$

Let us represent the integral (21) as the sum of three integrals:

$$I = \int_0^1 (A \cdot (1 - 0.75 \cdot c^{3.4}) - M)^2 \cdot c^{\alpha-1}(1-c)^{\beta-1} dc = \\ = A^2 \int_0^1 (1 - 1.5 \cdot c^{3.4} + 0.5625 \cdot c^{6.8}) \cdot c^{\alpha-1}(1-c)^{\beta-1} dc \\ - 2AM \int_0^1 (1 - 0.75 \cdot c^{3.4}) \cdot c^{\alpha-1}(1-c)^{\beta-1} dc + \\ + M^2 \int_0^1 c^{\alpha-1}(1-c)^{\beta-1} dc \quad (24)$$

We use the following equality. For any $k \geq 0$

$$\int_0^1 c^{\alpha+k-1}(1-c)^{\beta-1} dc = B(\alpha + k, \beta) \quad (25)$$

Then from (24) we obtain the following:

$$\begin{aligned}
\int_0^1 c^{\alpha-1}(1-c)^{\beta-1}dc &= B(\alpha, \beta) \\
\int_0^1 c^{\alpha+3.4-1}(1-c)^{\beta-1}dc &= B(\alpha + 3.4, \beta) \\
\int_0^1 c^{\alpha+6.8-1}(1-c)^{\beta-1}dc &= B(\alpha + 6.8, \beta)
\end{aligned} \tag{26}$$

By combining (21), (24) and (26) we finally obtain the following:

$$\begin{aligned}
D &= \frac{1}{B(\alpha, \beta)} \cdot \\
&\cdot ((A - M)^2 B(\alpha, \beta) - 1.5 \cdot A \cdot (A - M) B(\alpha + 3.4, \beta) + 0.5625 A^2 B(\alpha + 6.8, \beta))
\end{aligned} \tag{27}$$

or in an alternative form

$$D = (A - M)^2 - 1.5 \cdot A \cdot (A - M) \frac{B(\alpha + 3.4, \beta)}{B(\alpha, \beta)} + 0.5625 A^2 \frac{B(\alpha + 6.8, \beta)}{B(\alpha, \beta)} \tag{28}$$

Expressions (20) and (28) allow us to calculate the main statistical characteristics of the daily power generation by a solar panel based on the given parameters of the β -distribution of cloudiness. The proposed approach considers the statistical nature of cloudiness and allows us to more accurately predict power generation by a solar panel.

Currently, in order to forecast the power generation by a solar panel, the personnel of a solar power plant is forced to rely on forecast values of meteorological conditions provided by third-party meteorological services. Thus, the expected value of the cloudiness parameter is used to forecast generation, and the statistical nature of the change in cloudiness throughout the day is not taken into account.

Let us consider two main strategies for forecasting solar panel power generation:

- Strategy “Baseline A”. The influence of the “Cloudiness” parameter is taken into account in the form of a constant, which is equal to the mathematical expectation of the β -distribution of cloudiness. This value is provided by third-party meteorological services. The forecast of power generation is calculated according to expression (9) for the specified constant value of the cloudiness parameter;
- Strategy “Probabilistic B”. The influence of the “Cloudiness” parameter is taken into account as a random variable with β -distribution. The forecast of power generation is calculated using expression (20) for the given values of the parameters of the β -distribution of cloudiness.

Below, in Figure 7, the dependence of the relative error of power generation forecasts is given from the parameters of the β -distribution of cloudiness, calculated using the forecasting strategies “Baseline A” and “Probabilistic B”.

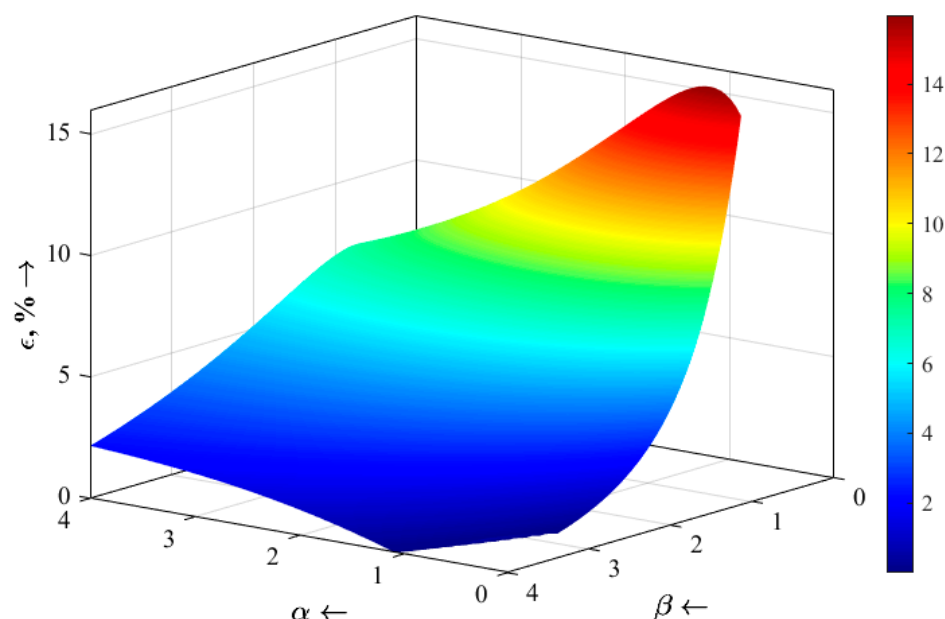


Figure 7. Dependence of the relative error between the values of daily power generation forecast calculated by Baseline A and Probabilistic B forecasting strategies on the parameters of the β -distribution of cloudiness.

Below, Figure 7 shows the dependence of the relative error between the values of daily forecast of power generation calculated by Baseline A and Probabilistic B forecasting strategies on the parameters of β -distribution of cloudiness.

As the analysis of the obtained results shows, in a fairly wide area of determining the parameters of the β -distribution of cloudiness, the relative error in calculating the forecast of solar generation exceeds 5%, and the maximum value of this error reaches 15.2%. The forecast of solar generation calculated taking into account the statistical nature of cloudiness according to strategy “Probabilistic B” turns out to be less than the forecast calculated based on the average value of cloudiness according to strategy “Baseline A”.

4. Discussion

The study shows that the β -distribution can be applied for the description of cloudiness and for the estimation of its impact on the daily energy yield of photovoltaic (PV) systems. The analytical model gave an average error of less than 4%, which confirms the possibility of using this approach in practice. This result also shows that simple probabilistic models can be more effective than methods based only on average meteorological parameters.

The comparison of analytical and experimental data indicates that the model is able to reproduce the main shape of the GHI histograms and also their asymmetry, which is connected with the random nature of cloudiness. Similar conclusions are given in previous studies, where probabilistic models of cloud cover were used to improve the accuracy of PV energy yield forecasts. This confirms that probabilistic methods are more suitable than deterministic ones when day-ahead forecasts are needed under variable weather conditions.

Some restrictions of the model must also be mentioned. The influence of humidity is not taken into account, which may limit the scope of application of the results to areas with arid or specific climatic conditions. The paper also considers only β -distribution does not allow describing multimodal or extreme cases of cloudiness. This can reduce the accuracy in regions with strong seasonal changes or with unstable weather. Future

research should also include additional meteorological factors, such as humidity, to improve the applicability of the model. Another direction for future research is to utilize Beta Mixture Model as a more flexible and powerful cloudiness distribution law.

The use of hybrid models can be one more way to increase accuracy. The combination of the proposed probabilistic framework with machine learning can give both interpretability and predictive performance. Such development can make the method more universal for different climate zones and improve the reliability of PV energy yield forecasting for practical tasks in power systems.

From a practical perspective, prediction of instantaneous values of solar power generation using machine learning and ANN-based methods can be important for ensuring the stability of the power grid. However, in practice, data on instantaneous values of power generation are often redundant. Much more valuable is information about the predicted value of daily solar generation. Based on these data, the economic relationship between the solar power plant and the state power system is formed. For example, in the Republic of Kazakhstan, insufficient power generation is the basis for penalties, and excessive power generation is paid for at reduced tariffs, which also reduces the technical and economic performance of the solar power plant. Therefore, accurate probabilistic forecasting of daily PV yield is of particular importance not only for technical stability but also for the economic sustainability of solar energy projects.

A quantitative comparison of the proposed β -distribution model with methods based on the application of machine learning and ANN is not performed within the framework of this study due to the limited scope of the publication. It should be noted that the application of machine learning methods is actually an alternative method to account for the nonstationarity of meteorological conditions. Estimation of the statistical characteristics of the β -distribution of cloudiness on the basis of analytical methods for calculating meteorological conditions is, in our opinion, a more accurate approach.

5. Conclusions

This research solves the current scientific and practical problem of forecasting daily power generation by solar power plants based on statistical characteristics of meteorological conditions, in particular, random variation in cloudiness. The importance of the topic considered is due to the need to ensure the high accuracy of generation forecasts for coordinating production plans of power plants with dispatch control of energy systems, as well as optimizing the financial performance of power plants.

The analysis conducted of the modern scientific literature has shown that the problem of forecasting power generation by solar installations is actively studied using various methods of artificial intelligence, deep learning, genetic algorithms and statistical models. The authors of the research have proposed a methodological approach based on the use of β -distribution to describe the statistical characteristics of cloudiness as a random variable. This favorable approach differs in terms of universality and the ability to adapt to regional climatic conditions, which allows to significantly increase the accuracy of power generation forecasts.

An analytical model has been developed during research. This model considers main factors affecting daily generation, such as the power plant's geographic location, the angle of incidence of solar rays, and atmospheric characteristics. The resulting analytical expressions for calculating daily power generation made it possible to estimate statistical indicators, including the expected value and variance of power generation, taking into account random changes in cloudiness throughout the day.

As a result of comparing the exact and approximate analytical calculation of daily power generation, it was found that the proposed approach provides satisfactory forecasting accuracy, with the average calculation error not exceeding 4%. It was also

demonstrated that taking into account the statistical characteristics of cloudiness, expressed through the β -distribution, allows reducing the maximum relative forecasting error to a level of less than 5%, which is significantly better than the traditional forecasting method based on the average cloudiness value.

Despite the high level of elaboration of the proposed approach, the research revealed certain drawbacks that limit the application of the obtained results in practice.

The first drawback is the simplified consideration of the influence of humidity on the atmospheric transmittance coefficient. During the research, it was assumed that at low altitudes of solar panels placement, the influence of humidity can be ignored. However, in real conditions, humidity can significantly affect the transmission of solar radiation, especially in regions with high variability of atmospheric humidity during the year. The underestimation of this factor can lead to inaccuracies in the calculation of energy production.

The second drawback is related to the limitation of the research to one type of distribution (β -distribution) for cloudiness modeling. Although the β -distribution is quite universal, it cannot always accurately reflect the statistical features of cloudiness in different climatic zones, especially in the presence of extreme meteorological conditions and rapid changes in weather conditions throughout the day.

The authors see the following potential areas for further research:

- Comprehensive research of the influence of humidity and other meteorological parameters on solar power generation: It is advisable to develop and test an extended mathematical model that takes into account the variability of humidity and its statistical distribution along with cloudiness, which will improve the accuracy of power generation forecasting.
- Development and research of hybrid forecasting models that combine different types of distributions (for example, combinations of β -distribution with other statistical models) to more accurately account for regional and seasonal features of cloudiness changes: It is also possible to use machine learning and deep learning methods for adaptive adjustment of models based on operational weather data, which will further improve the quality of short-term forecasts of solar power generation.

Thus, the implementation of the proposed directions will significantly improve the accuracy and efficiency of solar power plants power generation forecasting, which is an important task for the further development and integration of renewable energy sources into energy systems.

Author Contributions: Conceptualization, V.K. (Vitalii Kuznetsov) and V.K. (Valeriy Kuznetsov); methodology, V.K. (Vitalii Kuznetsov) and V.K. (Valeriy Kuznetsov); software, V.K. (Vitalii Kuznetsov) and V.K. (Valeriy Kuznetsov); validation, Z.C. and V.D.; formal analysis, V.D. and V.T.; investigation, V.K. (Vitalii Kuznetsov) and V.K. (Valeriy Kuznetsov); resources, A.R. and T.G.; data curation, T.G. and V.K. (Viktor Kovalenko); writing—original draft preparation, V.K. (Vitalii Kuznetsov) and Z.C.; writing—review and editing, V.D.; visualization, A.R.; supervision, T.G.; project administration, V.K. (Vitalii Kuznetsov); funding acquisition, V.K. (Valeriy Kuznetsov). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GHI	Global Horizontal Irradiance	
DNI	Direct Normal Irradiance	W/m ²
DHI	Diffuse Horizontal Irradiance	W/m ²
POA	Plane of Array Irradiance	
NWP	Numerical Weather Prediction	
Conv-GRU	Convolutional Gated Recurrent Unit	
NARX-GA	Nonlinear AutoRegressive model with eXogenous inputs optimized by Genetic Algorithm	
ARIMA	AutoRegressive Integrated Moving Average	
CNN	Convolutional Neural Network	
LSTM	Long Short-Term Memory	
GRU	Gated Recurrent Unit	
CNN-LSTM	Convolutional Long Short-Term Memory	
I_0	extraterrestrial solar irradiance on a plane normal to the radiation direction	W/m ²
I_{sc}	solar constant	~1361 W/m ²
E_0	eccentricity correction factor of the Earth's orbit (dimensionless)	°
φ	site latitude	°
δ	solar declination	°
ω	hour angle	°

Appendix A

Program Listing Implementing the GHI → DNI/DHI → POA → DC → AC Pipeline Using the Hourly GHI Profile

```
%% pv_daily_energy_from_GHI.m
% Calculation of daily AC-energy PV (1 m2) from hourly GHI.

clear; clc;

%% ----- INPUT DATA -----
% 1) Hourly GHI (W/m2), local standard time (no summer transitions)
%   There must be N=24 values for a day.
GHI = [... % 2025/05/07
        0 0 0 0 5.917 55.833 205.083 362.25 409.167 580.167 816.5 874.75 ...
        804.417 684 671.25 542.25 448.75 246.083 101.583 6.75 0 0 0];

GHI = [... % 2024/11/07
        0 0 0 0 0 0 0.083 8.583 55.833 72.917 147.667 186.5 ...
        326.75 210.25 131.083 74.667 8.417 0 0 0 0 0 0];

% 2) Date and sampling
year = 2025; month = 5; day = 7; % <- enter your 24 hours
t = datetime(year,month,day,0:23,0,0); % 24 points per hour, local time
(without DST)

% 3) Location and time zone (offset relative to UTC, hours)
%   Example: Saran, Kazakhstan ≈ 49.80°N, 72.86°E; from 2024 the country is in
UTC+5
lat = 49.80; % [deg] latitude (+N)
lon = 72.86; % [deg] longitude (+E)
```

```

tz = +5; % [h] time zone is relative UTC

% 4) Panel installation
tilt_deg = 30; % [deg] slope  $\beta$  from the horizontal (0= flat, 90= vertical)
azm_deg = 180; % [deg] panel azimuth  $\gamma_p$  (0= North, 90= East, 180= South,
270= West)
albedo = 0.15; % [-] ground reflectivity (grass/ground) 0.15–0.25

% 5) Meteo for thermal model (can be scalars or vector for 24 hours)
Tamb_C = 20*ones(size(GHI)); % [°C] air temperature

% 6) "Passport" module parameters (typical mono-Si)
eta_STC = 0.20; % [-] Efficiency at STC(1000 W/m2, 25°C)
gamma_P = -0.0045; % [1/°C] temperature power factor (approx -0.4...-0.5
%/°C)
NOCT_C = 45; % [°C] NOCT (800 W/m2, 20 °C, 1 m/s)
Area_m2 = 1.0; % [m2] panel area = 1 m2

% 7) Inverter parameters (simplified efficiency + power limitation)
Pinv_AC_rated_W = eta_STC*1000*Area_m2; % [W] rated AC ~ peak STC DC (~200
W/m2*m2)
inv_eff_nom = 0.96;

%% ----- CALCULATION BLOCKS -----
% Solar position (zenith/azimuth) and extraterrestrial radiation
[theta_z, gamma_s, I0n, G0h] = sunpos_and_extrat(t, lat, lon, tz);

% GHI -> (DHI, DNI) decomposition by Erbs + overnight/low angle protection
[DHI, DNI] = erbs_decomp(GHI, G0h, theta_z);

% Transpose to array plane: POA = beam + sky(diffuse) + ground
POA = transpose_to_POA(DNI, DHI, GHI, theta_z, gamma_s, tilt_deg, azm_deg,
albedo);

% NOCT cell temperature
Tcell_C = tcell_NOCT(POA, Tamb_C, NOCT_C);

% DC-module capacity (at 1 m2)
Pdc_W = max(0, POA * eta_STC .* (1 + gamma_P * (Tcell_C - 25))) * Area_m2;

% AC power of the inverter (simplified)
Pac_W = min(inv_eff_nom * Pdc_W, Pinv_AC_rated_W);
Pac_W(Pac_W < 1) = 0; % we cut the "creeping" power below 1 W

% Time integration (kWh)
dt_h = hours(diff(t)); dt_h(end+1) = dt_h(end); % last interval = previous interval
Eac_kWh = sum(Pac_W./1000 .* dt_h');

% Total
fprintf('Daily AC energy (1 m2): %.3f kWh\n', Eac_kWh);

```

```
%% ----- PLOTS -----
```

```
figure;
subplot(3,1,1); bar(t, GHI); grid on;
ylabel('GHI, W/m2'); title('GHI by hour');

subplot(3,1,2);
%plot(t, POA, 'LineWidth', 1.2); grid on;
bar(t, POA); grid on;
ylabel('POA, W/m2'); title('POA by hour');

subplot(3,1,3);
plot(t, Pdc_W, 'LineWidth', 1.2); hold on;
plot(t, Pac_W, 'o-', 'LineWidth', 1.2); grid on;
ylabel('Power, W'); legend('P_{DC}', 'P_{AC}');
title(sprintf('Energy per day: %.3f kW·h (AC)', Eac_kWh(1)));
xlabel('Time');
```

```
%% ===== LOCAL FUNCTIONS
```

```
function [theta_z, gamma_s, I0n, G0h] = sunpos_and_extrat(t, lat, lon, tz)
% Approximate astronomy: zenith  $\theta_z$ , azimuth  $\gamma_s$  (0 = C, 90 = B, 180 = S, 270 = Z).
% Extraterrestrial normal I0n and horizontal G0h radiation.
% All angles in radians / degrees as specified, output theta_z, gamma_s [rad].
deg2rad = pi/180;
N = numel(t);
theta_z = zeros(N,1); gamma_s = zeros(N,1);

% Day of the year
n = day(t, 'dayofyear');
% Declination (simple)
delta = 23.45 * deg2rad * sin(deg2rad*(360*(284 + n)/365));
% Time equation (min), approximate
B = 2 * pi * (n - 81)/364;
EoT_min = 9.87 * sin(2 * B) - 7.53 * cos(B) - 1.5 * sin(B);
% Longitude/hour zone correction: true solar time
Lst = tz * 15; % reference meridian of time zone, degrees
% Local time in hours
hour_local = hour(t) + minute(t)/60 + second(t)/3600;
TC_min = 4 * (Lst - lon) + EoT_min; % min
solar_time_h = hour_local + TC_min/60;
% Clock angle
H_deg = 15 * (solar_time_h - 12);
H = H_deg * deg2rad;
% Altitude and zenith
phi = lat * deg2rad;
cos_tz = sin(phi) * sin(delta) + cos(phi) * cos(delta) * cos(H);
cos_tz = max(cos_tz, 0); % -> 0
cos_tz(cos_tz > 1.5) = 0;
theta_z = acos(cos_tz);
sin_tz = sqrt(max(0, 1 - cos_tz.^2));
% Azimuth of the sun (0 = C, clockwise)
```

```

num = cos(delta). * sin(H);
den = cos(phi). * sin(delta) - sin(phi). * cos(delta). * cos(H);
gamma_s = atan2(num, den); % [-pi..pi], 0= North, + East
gamma_s(gamma_s<0) = gamma_s(gamma_s<0) + 2 * pi; % [0..2π)

% Extraterrestrial normal radiation I0n
% Eccentricity factor (simplified)
E0 = 1 + 0.033 * cos(2 * pi * n/365);
I_sc = 1367; % W/m² (solar constant, classical value)
I0n = I_sc . * E0;
G0h = I0n . * cos_tz;
end

function [DHI, DNI] = erbs_decomp(GHI, G0h, theta_z)
% Erbs: kd = DHI/GHI as a function of Kt = GHI/G0h
% Overnight protection and low cos(theta_z)
eps = 1×10−6;
cos_tz = cos(theta_z);
N = numel(GHI);
DHI = zeros(N,1); DNI = zeros(N,1);
for i = 1:N
    if GHI(i) <= 1 || G0h(i) <= eps || cos_tz(i) <= eps
        DHI(i) = GHI(i);
        DNI(i) = 0;
    else
        Kt = max(0, min(1.2, GHI(i)/G0h(i)));
        if Kt <= 0.22
            kd = 1 - 0.09 * Kt;
        elseif Kt <= 0.80
            kd = 0.9511 - 0.1604 * Kt + 4.388 * Kt^2 - 16.638 * Kt^3 + 12.336 * Kt^4;
        else
            kd = 0.165;
        end
        kd = max(0, min(1, kd));
        DHI(i) = kd * GHI(i);
        DNI(i) = max(0, (GHI(i) - DHI(i))/cos_tz(i));
    end
end
end

function POA = transpose_to_POA(DNI, DHI, GHI, theta_z, gamma_s, tilt_deg,
azm_deg, albedo)
% Simple transposition: direct beam + isotropic diffusion + ground reflection
% Azimuths: 0 = C, 90 = B, 180 = S, 270 = Z. Angle θz in radians.
deg2rad = pi/180;
beta = tilt_deg * deg2rad;
gamma_p = azm_deg * deg2rad;
cos_tz = cos(theta_z);
sin_tz = sin(theta_z);
% Cosine of the angle of incidence on the plane
cos_i = cos_tz. * cos(beta) + sin_tz. * sin(beta). * cos(gamma_s - gamma_p);

```

```

cos_i = max(cos_i, 0);
% Constituents
I_beam = DNI . * cos_i';
I_sky = DHI . * (1 + cos(beta))/2; % isotropic sky model
I_grnd = albedo . * GHI . * (1 - cos(beta))/2;
POA = max(0, I_beam + I_sky + I_grnd);
end

function Tcell = tcell_NOCT(POA, Tamb, NOCT_C)
% NOCT-model: Tcell = Tamb + (POA/800) * (NOCT-20)
Tcell = Tamb + (POA/800) . * (NOCT_C - 20);
end

```

References

1. Das, S. Short term forecasting of solar radiation and power output of 89.6kWp solar PV power plant. *Mater. Today Proc.* **2021**, *39*, 1959–1969. <https://doi.org/10.1016/j.matpr.2020.08.449>.
2. He, X.; Wang, Y.; Zhang, Y.; Ma, X.; Wu, W.; Zhang, L. A novel structure adaptive new information priority discrete grey prediction model and its application in renewable energy generation forecasting. *Appl. Energy* **2022**, *325*, 119854. <https://doi.org/10.1016/j.apenergy.2022.119854>.
3. Despotovic, M.; Voyant, C.; Garcia-Gutierrez, L.; Almorox, J.; Notton, G. Solar irradiance time series forecasting using auto-regressive and extreme learning methods: Influence of transfer learning and clustering. *Appl. Energy* **2024**, *365*, 123215. <https://doi.org/10.1016/j.apenergy.2024.123215>.
4. Gyeltshen, S.; Hayashi, K.; Tao, L.; Dem, P. Statistical evaluation of a diversified surface solar irradiation data repository and forecasting using a recurrent neural network-hybrid model: A case study in Bhutan. *Renew. Energy* **2025**, *245*, 122706. <https://doi.org/10.1016/j.renene.2025.122706>.
5. Benitez, I.B.; Ibañez, J.A.; Lumabad, C.I.D.; Cañete, J.M.; Principe, J.A. Day-Ahead Hourly Solar Photovoltaic Output Forecasting Using SARIMAX, Long Short-Term Memory, and Extreme Gradient Boosting: Case of the Philippines. *Energies* **2023**, *16*, 7823. <https://doi.org/10.3390/en16237823>.
6. Abdel-Basset, M.; Hawash, H.; Chakraborty, R.K.; Ryan, M. PV-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. *J. Clean. Prod.* **2021**, *303*, 127037. <https://doi.org/10.1016/j.jclepro.2021.127037>.
7. Hassan, M.A.; Bailek, N.; Bouchouicha, K.; Nwokolo, S.C. Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. *Renew. Energy* **2021**, *171*, 191–209. <https://doi.org/10.1016/j.renene.2021.02.103>.
8. Velásquez, R.M.A. A case study of NeuralProphet and nonlinear evaluation for high accuracy prediction in short-term forecasting in PV solar plant. *Heliyon* **2022**, *8*, e10639. <https://doi.org/10.1016/j.heliyon.2022.e10639>.
9. Khan, Z.A.; Hussain, T.; Baik, S.W. Dual stream network with attention mechanism for photovoltaic power forecasting. *Appl. Energy* **2023**, *338*, 120916. <https://doi.org/10.1016/j.apenergy.2023.120916>.
10. Azizi, N.; Yaghoubirad, M.; Farajollahi, M.; Ahmadi, A. Deep learning based long-term global solar irradiance and temperature forecasting using time series with multi-step multivariate output. *Renew. Energy* **2023**, *206*, 135–147. <https://doi.org/10.1016/j.renene.2023.01.102>.
11. Kim, E.; Akhtar, M.S.; Yang, O.-B. Designing solar power generation output forecasting methods using time series algorithms. *Electr. Power Syst. Res.* **2023**, *216*, 109073. <https://doi.org/10.1016/j.epsr.2022.109073>.
12. Cui, Y.; Wang, P.; Meirink, J.F.; Ntantis, N.; Wijnands, J.S. Solar radiation nowcasting based on geostationary satellite images and deep learning models. *Sol. Energy* **2024**, *282*, 112866. <https://doi.org/10.1016/j.solener.2024.112866>.
13. Lai, W.; Zhen, Z.; Wang, F.; Fu, W.; Wang, J.; Zhang, X.; Ren, H. Sub-region division based short-term regional distributed PV power forecasting method considering spatio-temporal correlations. *Energy* **2024**, *288*, 129716. <https://doi.org/10.1016/j.energy.2023.129716>.
14. Xu, Y.; Zheng, S.; Zhu, Q.; Wong, K.-C.; Wang, X.; Lin, Q. A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting. *Expert Syst. Appl.* **2024**, *254*, 124286. <https://doi.org/10.1016/j.eswa.2024.124286>.

15. Li, M.; Li, Y.; Diao, Y. A precise and efficient K-means-ELM model to improve ultra-short-term solar irradiance forecasting. *Renew. Energy Focus* **2024**, *51*, 100645. <https://doi.org/10.1016/j.ref.2024.100645>.
16. Ma, C.; Han, R.; An, Z.; Hu, T.; Jin, M. Weather-Driven Solar Power Forecasting Using D-Informer: Enhancing Predictions with Climate Variables. *Energy Eng.* **2024**, *121*, 1245–1261. <https://doi.org/10.32604/ee.2024.046644>.
17. Perera, M.; De Hoog, J.; Bandara, K.; Senanayake, D.; Halgamuge, S. Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data. *Appl. Energy* **2024**, *361*, 122971. <https://doi.org/10.1016/j.apenergy.2024.122971>.
18. Han, T.; Li, R.; Wang, X.; Wang, Y.; Chen, K.; Peng, H.; Gao, Z.; Wang, N.; Peng, Q. Intra-hour solar irradiance forecasting using topology data analysis and physics-driven deep learning. *Renew. Energy* **2024**, *224*, 120138. <https://doi.org/10.1016/j.renene.2024.120138>.
19. Dou, W.; Wang, K.; Shan, S.; Chen, M.; Zhang, K.; Wei, H.; Sreeram, V. A multi-modal deep clustering method for day-ahead solar irradiance forecasting using ground-based cloud imagery and time series data. *Energy* **2025**, *321*, 135285. <https://doi.org/10.1016/j.energy.2025.135285>.
20. Song, D.; Rehman, M.S.U.; Deng, X.; Xiao, Z.; Noor, J.; Yang, J.; Dong, M. Accurate solar power prediction with advanced hybrid deep learning approach. *Eng. Appl. Artif. Intell.* **2025**, *148*, 110367. <https://doi.org/10.1016/j.engappai.2025.110367>.
21. Abad-Alcaraz, V.; Castilla, M.; Carballo, J.A.; Bonilla, J.; Álvarez, J.D. Multimodal deep learning for solar radiation forecasting. *Appl. Energy* **2025**, *393*, 126061. <https://doi.org/10.1016/j.apenergy.2025.126061>.
22. Liu, M.; Rao, S.; Huang, M.; Deng, S. Short-term photovoltaic power forecasting based on improved transformer with feature enhancement. *Sustain. Energy Grids Netw.* **2025**, *43*, 101759. <https://doi.org/10.1016/j.segan.2025.101759>.
23. Song, Z.; Xiao, F.; Chen, Z.; Madsen, H. Probabilistic ultra-short-term solar photovoltaic power forecasting using natural gradient boosting with attention-enhanced neural networks. *Energy AI* **2025**, *20*, 100496. <https://doi.org/10.1016/j.egyai.2025.100496>.
24. Li, J.; Liu, Q. Forecasting of short-term photovoltaic power generation using combined interval type-2 Takagi-Sugeno-Kang fuzzy systems. *Int. J. Electr. Power Energy Syst.* **2022**, *140*, 108002. <https://doi.org/10.1016/j.ijepes.2022.108002>.
25. Sehrawat, N.; Vashisht, S.; Singh, A. Solar irradiance forecasting models using machine learning techniques and digital twin: A case study with comparison. *Int. J. Intell. Netw.* **2023**, *4*, 90–102. <https://doi.org/10.1016/j.ijin.2023.04.001>.
26. Abumohsen, M.; Owda, A.Y.; Owda, M.; Abumihsan, A. Hybrid machine learning model combining of CNN-LSTM-RF for time series forecasting of Solar Power Generation. *e-Prime Adv. Electr. Eng. Electron. Energy* **2024**, *9*, 100636. <https://doi.org/10.1016/j.prime.2024.100636>.
27. Mbungu, N.T.; Bashir, S.B.; Michael, N.E.; Farag, M.M.; Hamid, A.-K.; Ismail, A.A.A.; Bansal, R.C.; Abo-Khalil, A.G.; Elnady, A.; Hussein, M. Predictive control technique for solar photovoltaic power forecasting. *Energy Convers. Manag. X* **2024**, *24*, 100768. <https://doi.org/10.1016/j.ecmx.2024.100768>.
28. Dai, H.; Zhen, Z.; Wang, F.; Lin, Y.; Xu, F.; Duić, N. A short-term PV power forecasting method based on weather type credibility prediction and multi-model dynamic combination. *Energy Convers. Manag.* **2025**, *326*, 119501. <https://doi.org/10.1016/j.enconman.2025.119501>.
29. Bai, R.; Li, J.; Liu, J.; Shi, Y.; He, S.; Wei, W. Day-ahead photovoltaic power generation forecasting with the HWGC-WPD-LSTM hybrid model assisted by wavelet packet decomposition and improved similar day method. *Eng. Sci. Technol. Int. J.* **2025**, *61*, 101889. <https://doi.org/10.1016/j.jestch.2024.101889>.
30. Dou, W.; Wang, K.; Shan, S.; Zhang, K.; Wei, H.; Sreeram, V. A hybrid correction framework using disentangled seasonal-trend representations and MoE for NWP solar irradiance forecast. *Appl. Energy* **2025**, *397*, 126295. <https://doi.org/10.1016/j.apenergy.2025.126295>.
31. Pereira, S.; Canhoto, P.; Oozeki, T.; Salgado, R. Comprehensive approach to photovoltaic power forecasting using numerical weather prediction data and physics-based models and data-driven techniques. *Renew. Energy* **2025**, *251*, 123495. <https://doi.org/10.1016/j.renene.2025.123495>.
32. Huang, Y.; Pei, J.; Chen, L.; Chen, J.; Du, Z.; Peng, Z. Probabilistic Net Load Forecasting for High-Penetration RES Grids Utilizing Enhanced Conditional Diffusion Model. *arXiv* **2025**, arXiv:2503.17770v2. Available online: <https://arxiv.org/abs/2503.17770> (accessed on 16 September 2025).
33. Zhang, H.; Zandehshahvar, R.; Tanneau, M.; Van Hentenryck, P. Weather-informed probabilistic forecasting and scenario generation in power systems. *Appl. Energy* **2025**, *384*, 125369. ISSN 0306-2619. <https://doi.org/10.1016/j.apenergy.2025.125369>.
34. Ahmad, T.; Zhou, N.; Zhang, Z.; Tang, W. Enhancing Probabilistic Solar PV Forecasting: Integrating the NB-DST Method with Deterministic Models. *Energies* **2024**, *17*, 2392. <https://doi.org/10.3390/en17102392>.

35. Zhang, J.; Shang, S. Fast and Interpretable Probabilistic Solar Power Forecasting via a Multi-Observation Non-Homogeneous Hidden Markov Model. *Energies* **2025**, *18*, 2602. <https://doi.org/10.3390/en18102602>.
36. Wang, X.; Li, Z.; Fu, C.; Liu, X.; Yang, W.; Huang, X.; Yang, L.; Wu, J.; Zhao, Z. Short-Term Photovoltaic Power Probabilistic Forecasting Based on Temporal Decomposition and Vine Copula. *Sustainability* **2024**, *16*, 8542. <https://doi.org/10.3390/su16198542>.
37. Doelle, O.; Klinkenberg, N.; Amthor, A.; Ament, C. Probabilistic Intraday PV Power Forecast Using Ensembles of Deep Gaussian Mixture Density Networks. *Energies* **2023**, *16*, 646. <https://doi.org/10.3390/en16020646>.
38. Di Leo, P.; Ciocia, A.; Malgaroli, G.; Spertino, F. Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review. *Energies* **2025**, *18*, 2108. <https://doi.org/10.3390/en18082108>.
39. Yu, J.; Li, X.; Yang, L.; Li, L.; Huang, Z.; Shen, K.; Yang, X.; Yang, X.; Xu, Z.; Zhang, D.; et al. Deep Learning Models for PV Power Forecasting: Review. *Energies* **2024**, *17*, 3973. <https://doi.org/10.3390/en17163973>.
40. Blazakis, K.; Katsigiannis, Y.; Stavrakakis, G. One-Day-Ahead Solar Irradiation and Windspeed Forecasting with Advanced Deep Learning Techniques. *Energies* **2022**, *15*, 4361. <https://doi.org/10.3390/en15124361>.
41. Delgado, C.J.; Alfaro-Mejía, E.; Manian, V.; O'Neill-Carrillo, E.; Andrade, F. Photovoltaic Power Generation Forecasting with Hidden Markov Model and Long Short-Term Memory in MISO and SISO Configurations. *Energies* **2024**, *17*, 668. <https://doi.org/10.3390/en17030668>.
42. Lim, Y.; Son, M.; Park, K.; Kim, M.; Song, K.; Lee, H.; Kim, H. Power System Decision Making in the Age of Deep Learning: A Comprehensive Review. *Energies* **2025**, *18*, 4867. <https://doi.org/10.3390/en18184867>.
43. Kousounadis-Knousen, M.A.; Bazionis, I.K.; Georgilaki, A.P.; Catthoor, F.; Georgilakis, P.S. A Review of Solar Power Scenario Generation Methods with Focus on Weather Classifications, Temporal Horizons, and Deep Generative Models. *Energies* **2023**, *16*, 5600. <https://doi.org/10.3390/en16155600>.
44. Kasten, F.; Czeplak, G. Solar and terrestrial radiation dependent on the amount and type of clouds. *Sol. Energy* **1980**, *24*, 177–179. [https://doi.org/10.1016/0038-092X\(80\)90391-6](https://doi.org/10.1016/0038-092X(80)90391-6).
45. Sengupta, M.; Habte, A.; Wilbert, S.; Gueymard, C.; Remund, J. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition*; NREL: Golden, CO, USA, 2021. <https://doi.org/10.2172/1778700>.
46. Gelmanova, Z.; Batyrbek, A.; Sivyakova, G.; Fathi, M.S. Feasibility by modeling a photoelectric system in a solar power substation in Saran, Karaganda. In Proceedings of the 2024 International Conference on Electrical, Computer and Energy Technologies (ICECET), Sydney, Australia, 25–27 July 2024; pp. 1–8. <https://doi.org/10.1109/ICECET61485.2024.10698468>.
47. Reda, I.; Andreas, A. Solar position algorithm for solar radiation applications. *Sol. Energy* **2004**, *75*, 577–589. <https://doi.org/10.1016/j.solener.2003.12.003>.
48. Erbs, D.G.; Klein, S.A.; Duffie, J. A. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Sol. Energy* **1982**, *28*, 293–302. [https://doi.org/10.1016/0038-092X\(82\)90302-4](https://doi.org/10.1016/0038-092X(82)90302-4).
49. Chow, T.T.; Fong, K.F.; He, W. Perspectives on the origin, derivation, meaning, and significance of the isotropic sky model. *Sol. Energy* **2020**, *202*, 387–399. <https://doi.org/10.1016/j.solener.2020.03.089>.
50. Bobkov, S.G.; Chistyakov, G.P. On concentration functions of random variables. *J. Theor. Probab.* **2013**, *28*, 976–988. <https://doi.org/10.1007/s10959-013-0504-1>.

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