

Enhancing Probabilistic Solar PV Forecasting: Integrating the NB-DST Method with Deterministic Models

Tawsif Ahmad ¹, Ning Zhou ^{1, *}, Ziang Zhang ¹, and Wenyuan Tang ²

¹ Department of Electrical and Computer Engineering, State University of New York at Binghamton, Binghamton, NY 13902, USA; {tahmad2, ningzhou, ziang.zhang}@binghamton.edu

² Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695, USA; wtang8@ncsu.edu

* Correspondence: ningzhou@binghamton.edu.

Abstract: Accurate quantification of uncertainty in solar photovoltaic (PV) generation forecasts is imperative for the efficient and reliable operation of the power grid. In this paper, a data-driven non-parametric probabilistic method based on the Naïve Bayes (NB) classification algorithm and Dempster-Shafer theory (DST) of evidence is proposed for day-ahead probabilistic PV power forecasting. This NB-DST method extends traditional deterministic solar PV forecasting methods by quantifying the uncertainty of their forecasts by estimating the cumulative distribution functions (CDFs) of their forecast errors and forecast variables. The statistical performance of this method is compared with the analog ensemble method and the persistence ensemble method under three different weather conditions using real-world data. The study results reveal that the proposed NB-DST method coupled with an artificial neural network model outperforms the other methods in that its estimated CDFs have lower spread, higher reliability, and sharper probabilistic forecasts with better accuracy.

Keywords: Continuous rank probability score; Dempster-Shafer theory; naïve Bayes classification; probabilistic solar power forecasting; uncertainty quantification.

1. Introduction

Solar photovoltaic (PV) generation has been penetrating the power grid at an accelerating speed over the past decade as the world moves toward a more sustainable energy grid. Integrating a large number of solar PV power plants into the existing conventional grid presents significant challenges because solar PV generation is not dispatchable and is often treated as a form of negative load in power grid operations. To efficiently maintain the balance between generation and load in the power grid, both load and solar PV generation needs to be forecasted. These forecasts enable the proactive scheduling of dispatchable generation resources, allowing for the management of power ramping and the timely allocation of generation reserves to address power uncertainty [1] [2]. Moreover, solar PV power forecasting plays a significant role in predicting energy market imbalances and optimizing bid scheduling [3].

To support the bidding strategy and operation of a day-ahead electric power market, many data-driven forecasting methods have been developed to forecast solar PV generation. Typically, these day-ahead forecasting methods have forecast horizons between 24 and 48 hours, with a lead time of 24 hours. In other words, the first instance of the forecast is 24 hours ahead in the future relative to the current time [4]. Because of the large lead time, all forecasting methods inherently exhibit some levels of forecast errors that cannot be overlooked. Based on how they handle the uncertainty associated with the forecast errors, the forecasting methods can be broadly categorized into deterministic and

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Revised: date

Accepted: date

Published: date



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

probabilistic approaches. While deterministic methods provide a single-value forecast without quantifying forecast uncertainty, probabilistic methods quantify forecast uncertainty by estimating statistical properties of forecast variables, such as probability distributions and prediction intervals (PIs) [5].

Many deterministic methods have been developed to accurately forecast solar PV generation, gaining widespread acceptance in real-world applications [6, 7, 8, 9, 10]. Well-recognized forecasting methods include artificial neural networks (ANN) [11], recurrent neural networks (RNN) [12], long-short term memory (LSTM) [13], convolutional neural network (CNN) [14], and support vector regression (SVR) [15, 16]. Deterministic forecasting methods excel at estimating the expected solar PV generation and are valuable for managing power ramping. Yet, they cannot be directly used to address the uncertainty associated with solar PV generation.

Many probabilistic methods have been developed to quantify the uncertainty inherent in solar PV forecasts. For instance, the persistence ensemble (PerEn) method assumes that the forecasted solar PV generation follows a Gaussian distribution [17] and subsequently estimates its mean and variance. Meanwhile, the ensemble learning method (ELM) [18] integrates various deterministic forecasting methods, including k-nearest neighbors (kNN), decision trees, gradient boosting, random forest, lasso, and ridge regression, to estimate forecast distributions. For deep learning, an improved deep ensemble method, CNN-BiLSTM, is proposed in [19] for probabilistic wind speed forecasting, where the outputs of the final CNN-BiLSTM layer are fitted to a Gaussian distribution to estimate the mean and variance of the wind speed. Quantile regression (QR), on the other hand, estimates the conditional quantiles of solar irradiance based on numerical weather prediction (NWP) data [20]. In addition, an analog ensemble (AnEn) method is developed by [21] and compared to QR and PerEn methods. The AnEn method requires a frozen meteorological model to identify forecast points in the training dataset most similar to the current forecast. Furthermore, Doubleday et al. propose to combine several deterministic NWP model outputs using the Bayesian model averaging (BMA) technique to estimate forecast distributions [22]. In [23], Gaussian process regression (GPR) with specific kernel functions is applied to obtain prediction intervals of solar power forecast. This method tries to quantify the forecast uncertainty by fitting the forecast errors to a standard normal distribution. A sparse GPR method is proposed in [24] for probabilistic wind-gust forecasting by combining NWP data and on-site measurements. A variant of this method has shown promise for solar power forecasting of geographically sparse distributed solar plants [25]. The uncertainty quantified by probabilistic forecasting methods serves as a valuable guide for operators in booking appropriate generation reserves.

However, these probabilistic forecasting methods have certain limitations and challenges. For example, the AnEn, ELM, CNN-BiLSTM, and BMA rely on the assumption that a predefined family of probabilistic distributions can adequately characterize their forecasts. The GPR-based approaches, however, non-parametrically forecasts the solar power generation but fits the forecast error to a standard normal distribution to quantify the uncertainty. As a result, discrepancies between actual and assumed distributions may lead to significant errors. In addition, both the ELM and BMA methods require forecast outputs from several NWP models, which is time-consuming and may introduce additional uncertainty due to variation across models. As such, these methods have not gained widespread adoption in practical applications.

To overcome the above-mentioned limitations inherent in existing probabilistic forecasting methods and leverage the widespread acceptance of deterministic forecasting methods, this paper proposes a data-driven non-parametric statistical method for day-ahead probabilistic forecasting of PV power [26]. First, the method estimates the probability of the forecast error falling into a certain interval using the Naïve Bayes (NB) classification technique. Then, the estimated evidence is combined using a suitable statistical inference technique. Bayesian inference is a well-known method for combining independent evidence and has been utilized for many probabilistic forecasting applications, such

as wind-gust speed forecasting [27]. However, it requires prior probabilities. To address this limitation, a frequentist approach known as the Dempster-Shafer theory (DST) is utilized in this paper to combine probabilities from various independent sources, specifically from NB's outputs. Hence, the proposed method is named the NB-DST method. As a non-parametric approach, the proposed NB-DST method provides the flexibility needed for real-world applications. Distinguishing from other probabilistic methods such as BMA and ELM, the proposed NB-DST method only requires a single deterministic model output. A similar method is utilized for probabilistic wind-power forecast, which uses sparse Bayesian classification networks that require proper initialization of hyperparameters and a large training dataset [28]. In comparison, the NB classifiers used in this paper are easier to train with a small training dataset and outperform other methods like logistic regression [29].

The proposed NB-DST method is a novel approach in the landscape of probabilistic solar PV power forecasting. Unlike traditional methods that rely on predefined probabilistic distributions or extensive model ensembles, the NB-DST method uniquely combines NB classification with DST to offer a non-parametric, scalable solution for quantifying forecast uncertainty. This method not only addresses the limitations of existing probabilistic approaches by eliminating the need for prior probability distributions but also enhances adaptability to real-world forecasting scenarios. The purpose of this research is to demonstrate the efficacy and practicality of the NB-DST method in improving the accuracy and reliability of day-ahead solar power forecasts, thereby supporting more informed decision-making in grid management and energy market operations.

The structure of the rest of this paper is as follows: Section 2 outlines the proposed NB-DST method. The performance evaluation metrics are introduced in Section 3. In Section 4, the performance of the proposed method is compared with other methods utilizing real-world data from a rooftop solar PV plant. Finally, conclusions are drawn, and future work is discussed in Section 5.

2. NB-DST method for day-ahead solar PV forecast

This section introduces the proposed NB-DST method to estimate the cumulative distribution function (CDF) of day-ahead solar PV forecasts. The application of the proposed NB-DST method consists of the following four steps: (1) collect historical solar PV generation and related weather data; (2) apply a deterministic forecasting method such as SVR, ANN, et al. to forecast solar PV generation and determine the forecast errors; (3) utilize the NB method to determine the probability of the forecast errors falling into probability intervals; (4) employ the DST method to determine the CDF of forecast errors. The procedure of applying the NB-DST method to quantify the day-ahead solar PV forecast errors is detailed as follows.

2.1 Data collection

To perform day-ahead solar PV forecasting, day-ahead weather forecast, and solar power generation data are collected. For a particular geographic location, historical weather forecast data (w_h^d) and historical solar power generation data (s_h^d) for the targeted hour (h) and day (d) are collected. After identifying and removing bad data [30], the dataset is divided into three subsets (see Figure 1): (1) training dataset for building deterministic forecasting models (about 75%), i.e., $Tr_1 = \{w_h^d, s_h^d \mid d \in D_1\}$. (2) training dataset for building the NB models (about 15%), i.e., $Tr_2 = \{w_h^d, s_h^d \mid d \in D_2\}$. (3) testing dataset for evaluating the performance of the NB-DST method (about 10%), i.e., $TS = \{w_h^d, s_h^d \mid d \in D_3\}$.

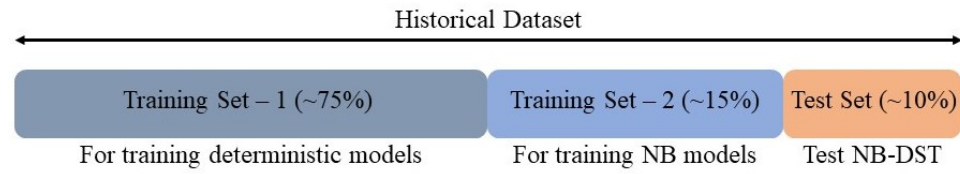


Figure 1. Segmentation of historical dataset for implementing the NB-DST method.

2.2 Deterministic forecasting methods and forecast errors

Using the training dataset Tr_1 as input, deterministic forecasting methods aim to build models for forecasting solar PV generation 24 hours in advance: $\hat{s}_h^d = f_h^{wc}(\mathbf{w}_h^d, s_h^{d-1})$. Here, the output of the deterministic forecasting methods is a forecasting model, denoted as $f_h^{wc}(\cdot)$, which may include models used by SVR, ANN, and persistence methods. To catch hourly variations, separate models are trained for each hour (h) of the day, using historical data for the corresponding hour.

Furthermore, three distinct forecasting models are trained to accommodate different weather conditions (wc): clear, overcast, and partially cloudy. These trained forecasting models generate hourly day-ahead deterministic forecasts ($\hat{s}_h^d$) using dataset Tr_2 . Then, forecast errors (ε_h) can be calculated using (1) as the difference between the observations (s_h^d) and forecasts.

$$\varepsilon_h^d = s_h^d - \hat{s}_h^d \quad (1)$$

To apply the NB method, ε_h^d is cast into equally divided intervals. First, the lower and upper limits of historical forecast errors are determined as $\xi_h^{min} = \min_{d \in D_2} \varepsilon_h^d$ and $\xi_h^{max} = \max_{d \in D_2} \varepsilon_h^d$, respectively. Then, the historical error range $[\xi_h^{min}, \xi_h^{max}]$ is divided into L intervals of equal width. These intervals are defined as $I_h^l = \text{interval}(\text{Low}_h^l, \text{Up}_h^l)$, where $\text{Low}_h^l = \xi_h^{min} + (l-1)\Delta_h$ and $\text{Up}_h^l = \xi_h^{min} + l\Delta_h$ for $l=1, 2, \dots, L$ and $\Delta_h = \frac{\xi_h^{max} - \xi_h^{min}}{L}$.

Finally, an NB binary classifier model is trained for each error interval at hour h . This approach is chosen because the results of each NB model will serve as evidence for the DST. Training a single multi-class NB classifier would result in only one mass function, where estimated probabilities for each class sum to '1,' and individual class probabilities do not form meaningful evidence for the DST. To train the NB classifier for hour h , the day-ahead weather forecast data ($\mathbf{w}_h^d$) and solar power forecast data ($\hat{s}_h^d$) are used as predictors ($\mathbf{x}_h^d = [\mathbf{w}_h^d, \hat{s}_h^d]^T$). The target is represented as a binary class label, $\{0, 1\}$, where '0' represents the event that the forecast error falls outside the corresponding interval, and '1' represents the event that the forecast error falls inside the corresponding interval. Mathematically, the labeling can be written as follows:

$$y_h^l = \mathbb{I}(\varepsilon_h^d \in I_h^l) = \begin{cases} 1 & \text{if } \varepsilon_h^d \in I_h^l \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where y_h^l is the class label of a training sample of the l^{th} NB classifier of hour h , and $\mathbb{I}(\cdot)$ is an indicator function.

NB classifiers are trained using dataset Tr_2 and validated through the k -fold cross-validation technique [31]. For the l^{th} interval at hour h , the NB classifier provides two probabilities: one for the forecast falling into the interval (p_h^l) and another for it falling out of the interval ($1 - p_h^l$), where $l \in \{1, 2, \dots, L\}$. This procedure is illustrated in Figure 2.

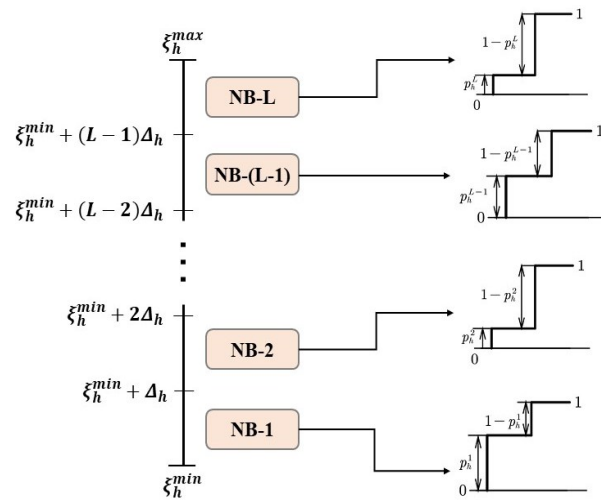


Figure 2. NB classifiers and corresponding probability outputs for forecast intervals at hour, h .

2.3 Probability estimation with the NB method

The NB method, a straightforward probabilistic classification algorithm based on *Bayes' theorem* of probabilistic inference [32], serves as a crucial tool in this study. *Bayes' theorem* connects the conditional and marginal probabilities of two random variables, is often used to calculate posterior probabilities based on given data. In this study, the NB method is used to estimate $P(\varepsilon_h^d \in I_h^l | \mathbf{x}_h^d)$, i.e., the probability of forecast errors ε_h^d falling into interval I_h^l given the predictor feature vector $\mathbf{x}_h^d$.

To be concise, the following study will be focused on interval l at hour h of day d , so that the superscripts l and h can be dropped. Assume that the training data $\{\mathbf{x}^d, y^d\}$ are given for $d = 1, 2, \dots, N_2$. Here, N_2 is the total number of days in Tr_2 . Symbol $\mathbf{x}^d = [x_1^d, x_2^d, \dots, x_v^d, \dots, x_V^d]^T$ is a V -dimensional feature vector, which consists of day-ahead weather forecast $\mathbf{w}^d$ and solar power forecast $\hat{s}^d$. Symbol y^d is the corresponding class label defined in (2). An NB classifier computes the probabilities that an unclassified sample belongs to certain class label(s) conditioned on the feature values using *Bayes' theorem*, as depicted in (3).

$$P(Y = C | \mathbf{X} = \mathbf{x}) = \frac{f_{X|Y=C}(\mathbf{x}) \cdot P(Y = C)}{f_{X|Y=0}(\mathbf{x}) \cdot P(Y = 0) + f_{X|Y=1}(\mathbf{x}) \cdot P(Y = 1)} \quad (3)$$

Here, $C \in \{0,1\}$ is the class label, $f_{X|Y=C}(\mathbf{x})$ is the probability density function (PDF) of feature vector $\mathbf{x}$ given that Y attains class label C , and $P(Y = C)$ is the prior probability of class C . It is assumed that the prior probability follows a Bernoulli distribution.

The NB classifier assumes conditional independence of the features, meaning that the value of a particular feature is considered independent of other features for a given class, as expressed in (4).

$$f_{X|Y=C}(\mathbf{x}) = \prod_{v=1}^V f_{X_v|Y=C}(x_v) \quad (4)$$

Although this assumption may not always hold in some practical applications, NB classifiers, when coupled with *kernel density estimation* (KDE), can still achieve accurate estimations in such cases [32]. KDE is a non-parametric estimation technique for estimating PDF, which can be described by (5.a).

$$\hat{f}_X(\mathbf{x}) = \frac{1}{N_2 \lambda} \sum_{d=1}^{N_2} K\left(\frac{\mathbf{x} - \mathbf{x}^d}{\lambda}\right) \quad (5.a)$$

Here, $\hat{f}_X(\mathbf{x})$ represents the KDE of the PDF of a random variable $\mathbf{X}$ given N_2 realizations of $\mathbf{x}$, $[x_1, x_2, \dots, x_{N_2}] \sim f_X(\mathbf{x})$. The smoothing parameter λ is also known as

bandwidth. Additionally, $K(\cdot)$ denotes the kernel function. A commonly used choice for this function is the Gaussian kernel, which is defined in (5.b).

$$K\left(\frac{\mathbf{x} - \mathbf{x}^d}{\lambda}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left\|\frac{\mathbf{x} - \mathbf{x}^d}{\lambda}\right\|^2} \quad (5.b)$$

The KDE of the conditional PDF of the feature vector $\mathbf{X}$ given $Y=C$ can be obtained through (6). Here, N_C is the number of training samples belonging to class C , and λ_v^C is the bandwidth for the v^{th} feature for class C and selected to optimize the Gaussian kernel in (5.b).

$$\hat{f}_{X|Y=C}(\mathbf{x}) = \prod_{v=1}^V \left\{ \frac{1}{N_C \lambda_v^C} \sum_{n_C=1}^{N_C} K\left(\frac{x_v^{n_C} - x_v}{\lambda_v^C}\right) \right\} \quad (6)$$

An NB classifier for a specific error interval is trained by simultaneously simulating (3) and (6) for each training sample of the corresponding interval of a particular hour in dataset Tr_2 . The output probabilities of an NB classifier for the l^{th} interval of hour h can be written as (7).

$$p_h^l = P(Y = 1 | \mathbf{X} = \mathbf{x}) \quad (7.a)$$

$$1 - p_h^l = P(Y = 0 | \mathbf{X} = \mathbf{x}) \quad (7.b)$$

2.4 Review of the DST

The DST is a mathematical framework employed to combine evidence from multiple independent sources in the presence of uncertainty [26, 33]. The DST operates by collecting information from individual sources represented by a *mass function*, also known as (a.k.a.) the *basic probability assignment* (BPA). The BPAs are then merged using *Dempster's rule of combination* to determine the degrees of belief in a set of events of interest.

When formulating a problem for DST application, a set of all possible solutions, a.k.a. the frame of discernment, is considered. It is assumed that the events of interest are subsets of this frame. Let Ω represent frame of discernment and E represent an event of interest, where $E \subseteq \Omega$. Then, $\bar{E} \cong \Omega - E$ represents the events in Ω that are distinct from event E . The BPA for Ω is a function $m: 2^\Omega \rightarrow [0,1]$ that has the following two properties:

- 1) $m(\phi) = 0$, where ϕ denotes the empty set.
- 2) $\sum_{E \in 2^\Omega} m(E) = 1$ where $m(E)$ is the degree of belief assigned to event E .

Let $m_1, m_2, \dots, m_z$ represent the BPAs for independent evidence $F_1, F_2, \dots, F_z$ in support of event E from Ω . *Dempster's rule of combination* can be exploited to construct a new *mass function* for combined evidence in support of E using (8).

$$m(E) = (m_1 \oplus m_2 \oplus \dots \oplus m_z)(E) = \frac{1}{U} \sum_{F_1 \cap F_2 \cap \dots \cap F_z = E} m_1(F_1) \cdot m_2(F_2) \cdot \dots \cdot m_z(F_z) \quad (8)$$

Here, $\oplus$ denotes the operator for *Dempster's rule of combination*, which performs orthogonal sum of *mass functions* supporting event E . And U is a normalization constant representing the degree of conflict between individual BPAs, as described by (9).

$$U = \sum_{F_1 \cap F_2 \cap \dots \cap F_z \neq \phi} m_1(F_1) \cdot m_2(F_2) \cdot \dots \cdot m_z(F_z) \quad (9)$$

2.5 Incorporating evidence from NB classifiers using the DST

The NB classifiers provide individual probabilities indicating the likelihood of forecast errors falling within a specific interval. To accurately determine the final probability of an error falling into a specified interval, DST is used to combine the evidence that the error falls into that interval with the evidence that error does not fall into any other intervals. Because there are L possible intervals for the error at hour h , there are L possible events, which are denoted as $E_1^h, E_2^h, \dots, E_l^h, \dots, E_L^h$. Here, E_l^h represents the event that the error falls into interval I_h^l , where $\{l \in \mathbb{Z}^+: 1 \leq l \leq L\}$. The frame of discernment at hour h , denoted by Ω_h , can be written as (10).

$$\Omega_h = E_1^h \cup E_2^h \cup \dots \cup E_l^h \cup \dots \cup E_L^h \quad (10)$$

The mass functions of these individual events, a.k.a. BPAs, are obtained from the NB classifiers. The mass function $m_l(E_l^h)$ of the event E_l^h is defined in (11) for all values of l . Here, $\overline{E_l^h}$ denotes the event that the error does not fall into the l^{th} interval.

$$m_l(E_l^h) = p_h^l \quad (11.a)$$

$$m_l(\overline{E_l^h}) = 1 - p_h^l \quad (11.b)$$

Because solar generation is inherently non-negative and cannot exceed the generation capacity of the solar power plant, the constraint described by (12) is imposed on solar power generation. Here, s_{cap} represents the maximum generation capacity of the solar power plant.

$$0 \leq s_h \leq s_{cap} \quad (12)$$

The range of the forecast error is then established using (13). The bounds on the index of error interval are set through (14). Here, $\lceil \cdot \rceil$ represents the ceiling operation, l_{lower} and l_{upper} are the indices of the lowermost and uppermost error intervals such that $1 \leq l_{lower} < \dots < l_{upper} \leq L$.

$$0 \leq \hat{s}_h + \varepsilon_h \leq s_{cap} \quad (13.a)$$

$$-\hat{s}_h \leq \varepsilon_h \leq s_{cap} - \hat{s}_h \quad (13.b)$$

$$l_{lower} = \left\lceil \frac{-\hat{s}_h - \xi_h^{min}}{\Delta_h} \right\rceil \quad (14.a)$$

$$l_{upper} = L - \left\lceil \frac{\xi_h^{max} - (s_{cap} - \hat{s}_h)}{\Delta_h} \right\rceil \quad (14.b)$$

The frame of discernment is updated according to the lowermost and uppermost error intervals, as shown in (15).

$$\Omega_h = E_{l_{lower}}^h \cup \dots \cup E_{l_{upper}}^h \quad (15)$$

Here, instead of combining the evidence from the whole set of error intervals, the frame of discernment is shrunk to a subset of M intervals, where $M = l_{upper} - l_{lower} + 1$ and $M \leq L$. Next, integrate these pieces of evidence using (8) and (9) to determine the probabilities of the forecast errors that fall into each of these M intervals, which constitute the mass function $m(E_i^h) \forall i: l_{lower} \leq i \leq l_{upper}$, as illustrated in (16).

$$m(E_i^h) = (m_{l_{lower}} \oplus m_{l_{lower}+1} \oplus \dots \oplus m_{l_{upper}})(E_i^h) \quad (16)$$

The CDF of the forecast error is derived by accumulating the probabilities. The set of error values $\{\varepsilon_h^1, \varepsilon_h^2, \dots, \varepsilon_h^k, \dots, \varepsilon_h^M\}$ and their corresponding probabilities $\{\hat{p}_h^1, \hat{p}_h^2, \dots, \hat{p}_h^k, \dots, \hat{p}_h^M\}$ for the CDF are calculated using the interval bounds and the width, as outlined in (17).

$$\varepsilon_h^1 = \xi_h^{min} + (l_{lower} - 1) \Delta_h \quad (17.a)$$

$$\varepsilon_h^k = \varepsilon_h^1 + (k - 1) \Delta_h \text{ for } k = 2, 3, \dots, M \quad (17.b)$$

$$\hat{p}_h^1 = m(E_{l_{lower}}^h) \quad (17.c)$$

$$\hat{p}_h^k = \hat{p}_h^{k-1} + m(E_{l_{lower}+k-1}^h) \text{ for } k = 2, 3, \dots, M \quad (17.d)$$

Finally, the CDF of the solar generation at hour h is calculated by adding the day-ahead deterministic forecast to the error values for hour h , as shown in (18). Here, $\hat{S}_h^k$ represents the NB-DST forecast whose CDF value is $\hat{p}_h^k$.

$$\hat{S}_h^k = \varepsilon_h^k + \hat{s}_h \text{ for } 1 \leq k \leq M \quad (18)$$

The framework of the NB-DST method is summarized in Figure 3. In this study, the NB-DST method is coupled with three deterministic forecasting methods: SVR, ANN, and QR (median only), resulting in SVR-NB-DST, QR-NB-DST, and ANN-NB-DST models, respectively. To mitigate the risk of overfitting the NB-DST models to the training data, a separate dataset, TS , is reserved for generating performance metrics to evaluate the models' effectiveness.

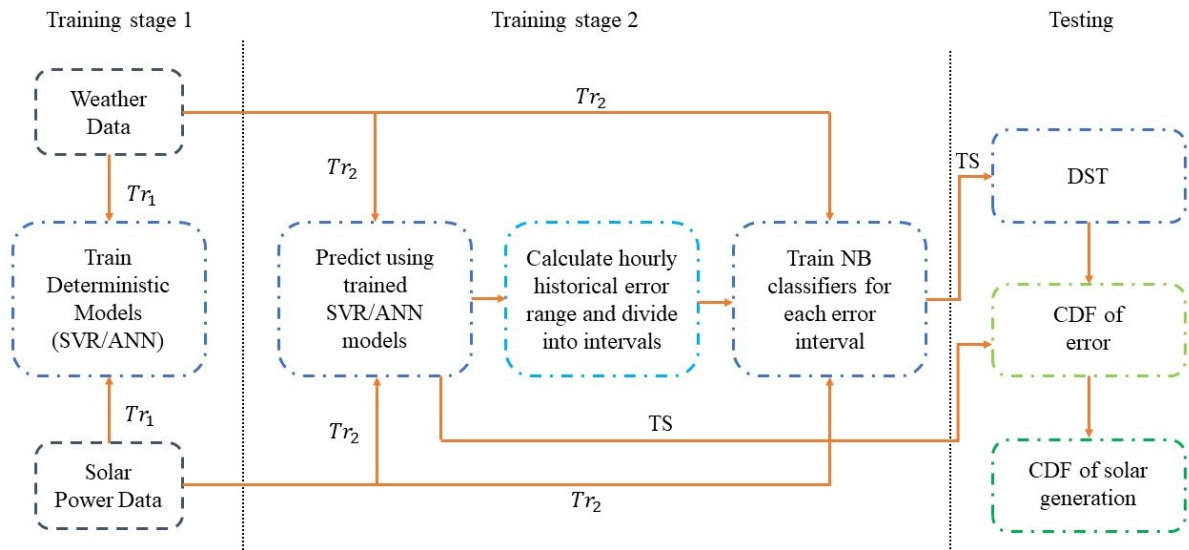


Figure 3. Framework of the proposed NB-DST method.

3. Performance metrics for evaluating probabilistic forecasts

Assessing probabilistic forecasts is more challenging than deterministic ones because probabilistic forecasts have various statistical properties. As can be found in the literature, many state-of-the-art probabilistic forecasting methods are only assessed using certain sets of evaluation metrics. But the evaluation process needs to balance several critical aspects, such as forecast accuracy, reliability, sharpness, and uncertainty. This paper aims at addressing this challenge by evaluating the probabilistic forecasting methods using a handful of metrics that assess distinct aspects of the forecasts, providing the forecasters with a framework for critical evaluation. In this section, these evaluation metrics are defined to evaluate the performance of the probabilistic forecasts.

3.1 Continuous rank probability score

The continuous rank probability score (CRPS) is an evaluation metric widely used in various forecasting domains, such as weather, solar and wind power, and load forecasting [5, 34, 35]. It effectively measures forecast accuracy, reliability, and sharpness. CRPS is defined in (19).

$$CRPS = \frac{1}{N} \sum_{k=1}^N \int_{-\infty}^{+\infty} [\hat{p}_{fcst}^k(x) - o_{s_h}^k(x)]^2 dx \quad (19)$$

Here, $\hat{p}_{fcst}^k(x)$ represents the CDF of the forecast at instant k . Symbol $o_{s_h}^k(x)$ denotes the CDF of the observation at instance k , which is defined as $o_{s_h}^k(x) = \mathbb{I}(x \geq s_h)$. Here, $\mathbb{I}(\cdot)$ is an indicator function, s_h denotes the observed value of solar generation. N stands for the total number of forecast instances in dataset TS .

A lower CRPS value is desirable, as it indicates that the estimated CDF has a lower spread around the observation. It is evident from (19) that the CRPS penalizes any

discrepancies between the forecasted CDF and the actual observations. Moreover, for a deterministic forecast, the CRPS transforms into the *mean absolute error* (MAE). This property makes the CRPS a preferred metric for comparing probabilistic and deterministic forecasts. CRPS can be further decomposed into three components as outlined in (20) [36]. The methods for calculating CRPS, along with its components - REL, UNC, and RES – are detailed in [36] and used in this paper.

$$CRPS = Reliability (REL) + Uncertainty(UNC) - Resolution(RES) \quad (20)$$

3.2 Brier score

The Brier score (BS) is a scoring function widely used to measure the accuracy of probabilistic forecasts [37]. The BS for a probabilistic forecast at hour h is defined in (21).

$$BS_h = \frac{1}{M} \sum_{j=1}^M [\hat{p}_h^j - y_h^j]^2 \quad (21)$$

Here, $\hat{p}_h^j$ represents the forecasted probability of event j happening at hour h . Term y_h^j is defined as $y_h^j = \mathbb{I}(s_h \geq \hat{S}_h^j)$. M denotes the total number of events considered in the probability forecast at hour h . To compute the BS for an entire day's forecast, one can average BS_h across all hours, from $h = 1$ to N_h , where N_h represents the total number of hours of the day.

3.3 Prediction interval coverage probability

After the CDF of the forecast is obtained from a probabilistic forecasting method, a prediction interval (PI) of the forecast can be calculated. For a confidence level of $(1 - \alpha) \times 100\%$, the PI is determined using (22). Here, $\hat{S}_{h,\tau}$ represents the τ^{th} quantile of the forecasted CDF, meaning $P(S_h \leq \hat{S}_{h,\tau}) = \tau$.

$$\widehat{PI}_{h,(1-\alpha) \times 100\%} = Interval\left(\hat{S}_{h,\tau=\frac{\alpha}{2}}, \hat{S}_{h,\tau=1-\frac{\alpha}{2}}\right) \quad (22)$$

The prediction interval coverage probability (PICP) is a measure of the *reliability* of a probabilistic forecasting method. PICP can be calculated from the $(1 - \alpha) \times 100\%$ PI as outlined in (23) [38].

$$PICP(\alpha) = \frac{1}{N_h} \sum_{h=1}^{N_h} \mathbb{I}(\hat{S}_h \in \widehat{PI}_{h,(1-\alpha) \times 100\%}) \quad (23)$$

Here, $\hat{S}_h$ denotes the forecast at hour h , N_h denotes the total number of forecasting hours in a day, and $\mathbb{I}(\cdot)$ is an indicator function. A forecasting method is considered more reliable if its $PICP(\alpha)$ is closer to the nominal coverage for the confidence intervals of the same size.

3.4 Prediction interval normalized average width

The prediction interval normalized average width (PINAW) is a metric for measuring the *sharpness* of a probabilistic forecast method [38]. Mathematically, PINAW is defined in (24), where s_h denotes the observed solar generation at hour h on the forecast day. In this formula, the numerator computes the average width of the $(1 - \alpha) \times 100\%$ PI of the probabilistic forecast, while the denominator computes the average hourly solar power generation. A lower PINAW value is desirable, as it indicates a shaper forecast.

$$PINAW(\alpha) = \frac{\frac{1}{N_h} \sum_{h=1}^{N_h} \left(\hat{S}_{h,\tau=1-\frac{\alpha}{2}} - \hat{S}_{h,\tau=\frac{\alpha}{2}} \right)}{\frac{1}{N_h} \sum_{h=1}^{N_h} s_h} \quad (24)$$

3.5 Coverage width criterion

The coverage width criterion (CWC) is defined in (25) [39] and can effectively address the limitations of both the PICP and PINAW. While PICP values closer to the nominal coverage indicate a more reliable forecast, they do not assess the sharpness of the forecast. Conversely, a lower PINAW signifies a sharper forecast but does not measure its

reliability. The CWC integrates both PICP and PINAW, providing a comprehensive metric to evaluate the reliability and sharpness of a forecast simultaneously.

$$CWC = \begin{cases} \beta \cdot PINAW & \text{for } PICP \geq PINC \\ (\kappa + \beta \cdot PINAW) \cdot \left(1 + \exp\left(-\eta \cdot \frac{PICP - PINC}{100}\right)\right) & \text{for } PICP < PINC \end{cases} \quad (25)$$

Here, PINC represents the prediction interval nominal coverage. Symbols κ , β , and η represent hyperparameters, which are chosen as suggested in [39]. A lower CWC value is desirable, as it indicates that a probabilistic forecasting method is more reliable and sharper.

4. Case study

In this section, day-ahead probabilistic forecasting of solar PV generation is performed using the proposed NB-DST method. In addition to the SVR-NB-DST, QR-NB-DST, and ANN-NB-DST methods, the study also implements the AnEn and PerEn methods [40] as benchmarks. Although there are many probabilistic methods in the literature, they have not been widely adopted in practice. For example, the PerEn method is still preferred for benchmarking within both industry settings and research environments, such as those conducted by the US Department of Energy [41]. The effectiveness of these methods is evaluated and compared using the performance metrics outlined in Section 3 using real-world data.

4.1 Data collection and selection

The solar generation data used in this study are collected from a 450-kW solar PV plant located on the rooftop of the Target® store in Vestal, NY, USA (latitude 42°05'37.0"N, longitude 76°00'06.0"W). This plant was established in 2016 with support from the New York State Energy Research and Development Authority (NYSERDA). Hourly solar generation data of this plant between 2016 and 2022 are downloaded from NYSERDA's website [42].

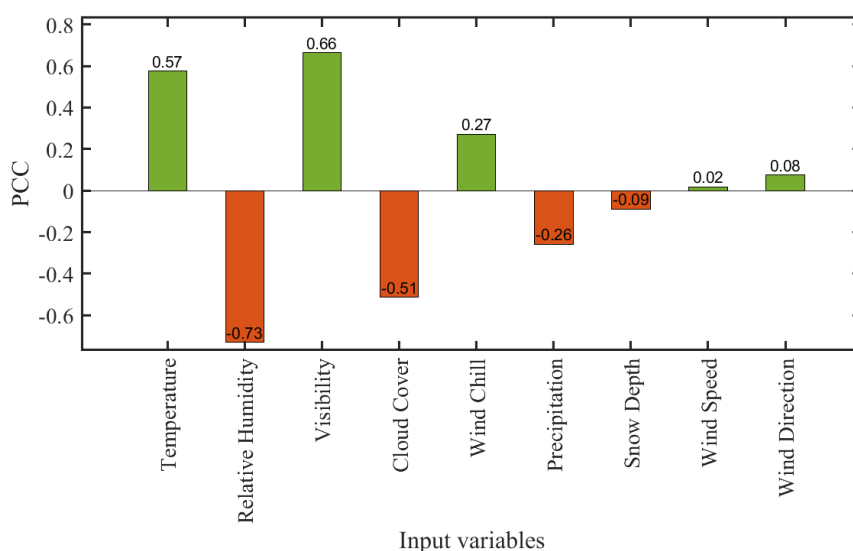


Figure 4. PCC of weather variables with the hourly solar generation data.

Additionally, this study incorporates various weather variables that significantly influence solar irradiance [43]. Historical weather data for the study site from 2016 to 2021, along with historical forecast data for 2022, are collected on an hourly basis from the website of Visual Crossing weather data service [44]. The weather variables available in this dataset are listed in Table 1. To select the predictor variables for the forecasting algorithms, the correlation between solar generation and each numerical weather variable is

measured using Pearson's correlation coefficient (PCC) [45]. After pre-processing and removing outliers from the data, the PCC for each numerical weather predictor variable with solar generation data is calculated. The results are presented in Figure 4. It is observed that *temperature* and *visibility* have a strong positive correlation with solar power generation, while *relative humidity* and *cloud cover* have a strong negative correlation. The correlations of other predictor variables are weak, as shown in Figure 4, and thus, these variables are excluded from the model. *Solar generation* data from the previous day's corresponding hours are also included in the model to account for the auto-correlation of the solar power generation data.

The datasets used in this study are divided as follows: the data from 2016 to 2019 are designated as the Tr_1 dataset, while data recorded from 2020 to 2021 are used as the Tr_2 dataset. The data recorded in 2022 are used as the TS dataset. Because weather conditions

Table 1. Weather variables for the study site available from Visual Crossing weather data service (2016–2022)

Weather variable	Description	Unit
Temperature	Actual temperature	°C
Dew point	Dew point	°C
Wind chill	Wind chill factor	°C
Heat index	Heat index	°C
Feels like	Combination of temperature, wind chill, and heat index	°C
Humidity	Relative humidity	%
Precipitation	Amount of liquid equivalent precipitation	mm
Snow depth	Average amount of snow on the ground	cm
Wind speed	Wind speed at 10m above ground	kmph
Wind direction	Direction of wind in reference to the north direction	degrees
Pressure	Sea level pressure	mbars
Cloud cover	Amount of sky covered with cloud	%
Visibility	Distance visible in daylight	km
Conditions	Weather condition reported by the weather station	none

are complex and have varying impact on solar PV generation, separate forecasting models are built and evaluated under different cloud patterns. Accordingly, the datasets are further divided into three categories on a daily basis according to the weather *conditions* using the approach described in [46]: “clear”, “overcast”, and “partially cloudy”. The daily solar generation profile on a *clear* day typically exhibits a smooth bell curve, with the power generation peaking near the generation capacity during noon. On *overcast* days, the total daily solar generation is significantly lower. On *partially cloudy* days, the hourly solar generation shows considerable fluctuation, characterized by multiple peaks and troughs throughout the day.

Subsequently, the forecasting models are first trained using Tr_1 and Tr_2 datasets for each hour of the day under these distinct weather conditions. Then, the forecasting models are evaluated using the performance metrics computed through the TS dataset. Note that out of 365 days in the TS dataset, there are 124 days classified as “clear”, 129 days classified as “overcast”, and 112 days classified as “partially cloudy”. The effectiveness of the forecasting models is evaluated using a 5-fold cross-validation approach.

Table 2. Performance metrics of all the forecasting methods evaluated over the 124 “Clear” days.

Method	CRPS (%)	CRPS Decomposition	BS (%)	PICP (%)	PINAW	CWC
--------	----------	--------------------	--------	----------	-------	-----

		REL	RES	UNC				
PerEn	11.14	1.20	48.38	58.32	6.87	91.62	19.87	119.32
AnEn	15.38	1.51	44.45	58.32	7.44	84.71	14.67	88.12
QR-NB-DST	7.86	1.58	52.04	58.32	5.73	87.18	9.33	56.08
SVR-NB-DST	7.97	1.41	51.76	58.32	5.41	86.35	9.81	58.96
ANN-NB-DST	6.28	1.38	53.42	58.32	4.62	88.42	7.25	43.60

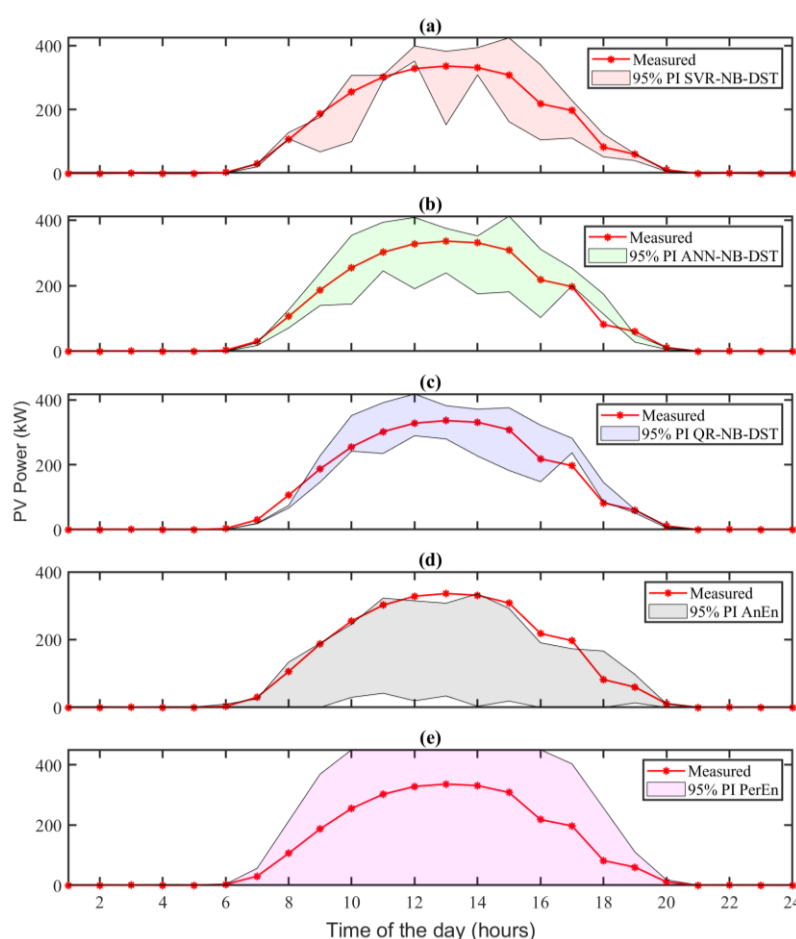


Figure 5. 95% PIs obtained from all the probabilistic solar forecasting methods on a particular “clear” day (19 May 2022) (a) SVR-NB-DST, (b) ANN-NB-DST, (c) QR-NB-DST, (d) AnEn, (e) PerEn. The red dotted lines denote the actual hourly observations of solar power; the shaded areas denote the 95% PIs of the forecasted solar generation obtained from the corresponding methods.

4.2 Performance evaluation under “clear” weather conditions

This subsection focuses on assessing the probabilistic forecasting methods during “clear” weather conditions. The performance metrics for the 124 “clear” days in the *TS* dataset are calculated for all the methods and summarized in Table 2. It can be observed from Table 2 that the three NB-DST methods outperform the AnEn and PerEn methods in terms of the CRPS. Also, the BS values for the NB-DST methods are considerably lower than those for the other methods, which indicates NB-DST’s higher accuracy. These three NB-DST methods show satisfactory *reliability* and coverage with *resolution* metrics notably surpassing those of the AnEn and PerEn methods. Also, a slightly lower value of the PINAW for the ANN-NB-DST gives it an advantage over the other two NB-DST methods. While the PICP values for the three NB-DST methods are comparable to those for the AnEn and PerEn methods, the NB-DST methods, particularly ANN-NB-DST, exhibit superior overall performance as indicated by their lower CWC values.

To further illustrate the performance, Figure 5 shows the estimated 95% PIs from all the forecasting methods for a particular “clear” day (19 May 2022). The 95% PIs are obtained using (22) by setting $\alpha=0.05$. It can be observed that the 95% PIs of the three NB-DST methods are narrower than those of the two benchmark methods. Figure 6(a) shows the CDF estimate for PV power at 10:00 AM on this “clear” day (19 May 2022). It can be observed that the CDFs estimated using the ANN-NB-DST and the QR-NB-DST methods exhibit the lowest deviation from the actual observation CDF. Although the CDF estimated by the SVR-NB-DST method shows a deviation slightly higher than other NB-DST methods, it still performs commendably when compared to the benchmark methods.

4.3 Performance evaluation under “overcast” weather conditions

In this subsection, the performance metrics are calculated over the 129 “overcast” days in the TS dataset for all the probabilistic forecasting methods and are summarized in Table 3. Among these methods, the ANN-NB-DST stands out by achieving the lowest CRPS. It also exhibits superior *reliability* and *resolution*, as indicated by its REL and RES values, compared to the other methods. Notably, the *resolution* of all the NB-DST methods is superior to the two benchmark methods as their PINAW values are much lower. While the PICP values for the NB-DST methods are further from the ideal 95% nominal coverage than the benchmark methods, their lower CWC values reflect better overall performance considering both forecast *reliability* and *sharpness*. Additionally, the lower BS values of the NB-DST methods suggest that the NB-DST methods are more accurate than the other methods. Among the three NB-DST methods, the ANN-NB-DST method exhibits the most impressive performance.

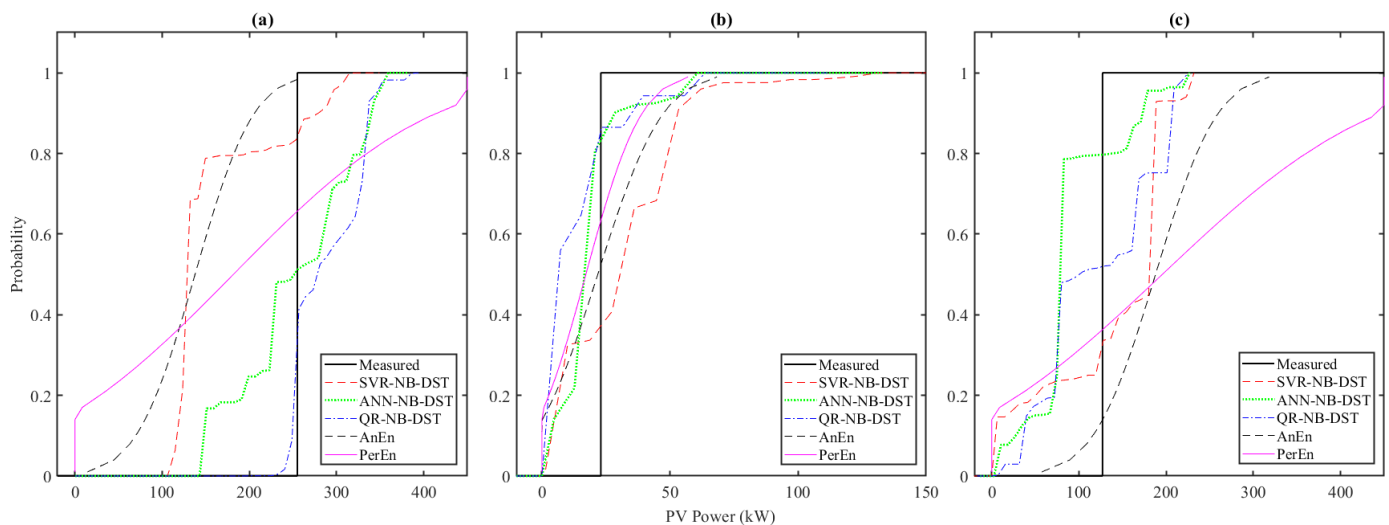


Figure 6. Estimated CDFs of the PV power using all the probabilistic forecasting methods at 10:00 AM on (a) a “clear” day (19 May 2022), (b) an “overcast” day (18 December 2022), (c) a “partially cloudy” day (09 July 2022). The solid black line represents the CDF of the actual observation.

Table 3. Performance metrics of all the forecasting methods evaluated over the 129 “Overcast” days.

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW	CWC
		REL	RES	UNC				
PerEn	5.89	0.51	7.33	12.71	4.73	91.68	1316.20	7897.3
AnEn	6.21	0.54	7.04	12.71	5.61	89.54	1353.90	8123.5
QR-NB-DST	5.74	0.48	7.45	12.71	3.68	86.61	632.80	3796.9
SVR-NB-DST	6.78	0.39	6.32	12.71	3.87	85.47	670.94	4025.7
ANN-NB-DST	5.37	0.41	7.75	12.71	3.13	88.25	581.42	3488.6

To further illustrate the performance, Figure 7 shows the 95% PIs obtained from the forecasting methods on a particular “*overcast*” day (28 December 2022). It can be observed that the 95% PIs for the AnEn and PerEn methods are notably broader than those for the NB-DST methods. Additionally, the SVR-NB-DST method meets the targeted coverage rate of 95% PIs, whereas the coverage rates of the ANN-NB-DST and QR-NB-DST methods are lower than the targeted 95%. This variation highlights the challenges of forecasting under *overcast* conditions. The estimated CDFs of the hourly solar generation estimated by the forecasting methods at 10:00 AM on this “*overcast*” day (28 December 2022) are illustrated in Figure 6(b). It can be observed from Figure 6(b) that the CDFs estimated by the NB-DST methods align more closely to the actual observation CDF than those estimated by the two benchmark methods.

4.4 Performance evaluation under “*partially cloudy*” weather conditions

In this subsection, the performance metrics are calculated over the 112 “*partially cloudy*” days in the TS dataset for all the probabilistic forecasting methods and are summarized in Table 4. It can be observed that the ANN-NB-DST method achieves the lowest CRPS value, which indicates the best overall performance. Additionally, the *reliability* and *resolution* of the ANN-NB-DST method are also superior to those of other NB-DST and benchmark methods, as indicated by its lower value of REL and higher value of RES. Moreover, the lower PINAW values of the NB-DST methods suggest that they have better *resolution* than the AnEn and the PerEn methods. While the PICP values for the NB-DST methods are not as good as those for the benchmark methods, the CWC values, which consider both PICP and PINAW, suggest a better overall performance for the NB-DST methods. Notably, the ANN-NB-DST method stands out with the lowest BS, indicating the highest accuracy among the five methods.

Table 4. Performance metrics of all the forecasting methods evaluated over the 112 “*partially cloudy*” days.

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW	CWC
		REL	RES	UNC				
PerEn	9.30	0.55	33.93	42.68	5.24	90.86	27.67	166.12
AnEn	9.98	0.64	33.34	42.68	6.77	86.33	28.60	171.70
QR-NB-DST	8.15	0.65	35.18	42.68	5.72	88.28	16.92	101.62
SVR-NB-DST	8.55	0.78	34.91	42.68	5.28	84.91	17.37	104.32
ANN-NB-DST	6.21	0.49	36.96	42.68	3.94	87.89	13.24	79.54

Table 5. Performance metrics of all the forecasting methods evaluated over the 1-year testing dataset

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW	CWC
		REL	RES	UNC				
PerEn	8.72	0.76	29.44	37.40	5.61	91.41	480.42	2882.61
AnEn	10.48	0.90	27.82	37.40	6.59	86.91	492.26	2953.67
QR-NB-DST	7.20	0.91	31.11	37.40	5.00	87.32	232.01	1392.15
SVR-NB-DST	7.73	0.86	30.53	37.40	4.83	85.60	245.79	1474.82
ANN-NB-DST	5.94	0.76	32.23	37.40	3.88	88.20	212.01	1272.18

To further illustrate the performance, Figure 8 shows the 95% PIs obtained from all the forecasting methods on a particular “*partially cloudy*” day (09 July 2022). It can be observed that the 95% PIs obtained from the benchmark methods are wider than those from the NB-DST methods. Also, in Figure 8, some measured points fall outside the 95% PIs obtained from the CDFs estimated by the NB-DST methods. In contrast, such occurrences

are not observed with the AnEn and PerEn methods. The observation highlights a strategic trade-off made by the NB-DST methods, prioritizing improved *resolution* over *coverage probability*. Figure 6(c) shows the CDFs estimated by all the methods at 10:00 AM on this “partially cloudy” day (09 July 2022). The CDFs estimated by the NB-DST methods show smaller deviations from the actual observed solar power CDF than those estimated by the AnEn and PerEn methods.

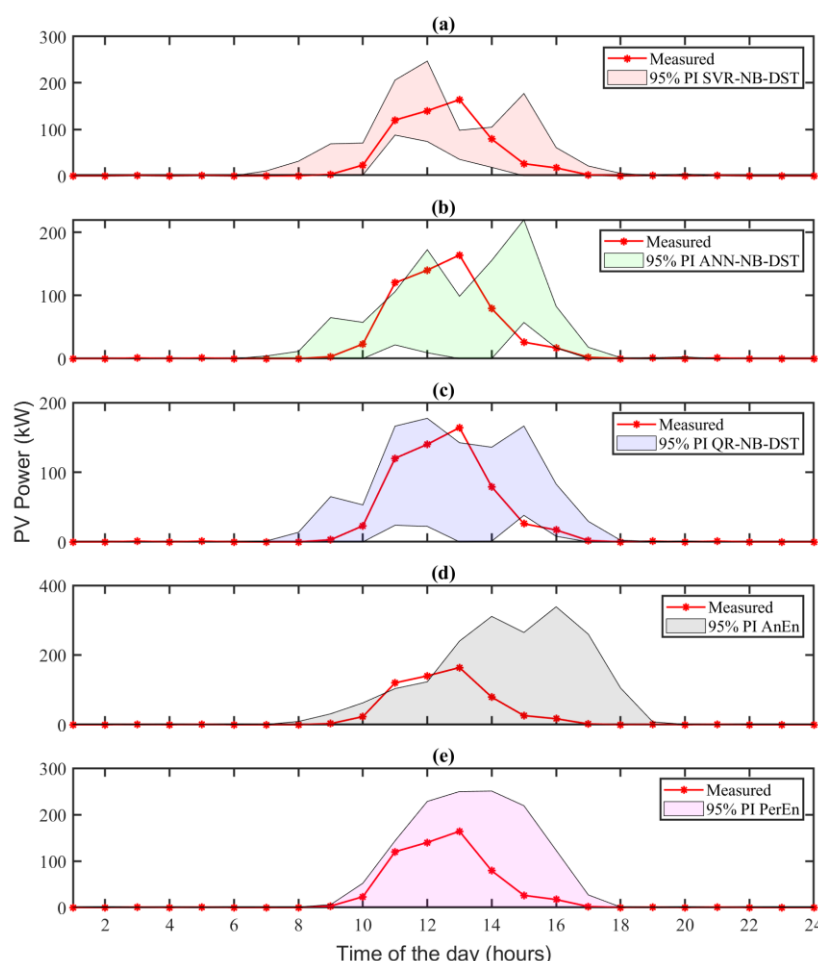


Figure 7. 95% PIs obtained from all the probabilistic solar forecasting methods on a particular “overcast” day (28 December 2022) (a) SVR-NB-DST, (b) ANN-NB-DST, (c) QR-NB-DST, (d) AnEn, and (e) PerEn.

4.5 Overall comparative analysis

The results from Table 2 – Table 4 are synthesized to evaluate the overall performance of the proposed method over the 1-year testing dataset, *TS*. The findings are presented in Table 5. Notably, the NB-DST methods outperform the benchmark methods on nearly all performance metrics except for the PICP. This exception is understandable because the PI estimates of the PerEn and AnEn methods are considerably wider than those of the NB-DST methods. Even with the exception, the significantly better PINAW and CWC values suggest the overall superiority of the NB-DST methods. According to Table 5, the ANN-NB-DST emerges as the best-performing method among the NB-DST variants.

Figure 9 further illustrates the percentage improvements in CRPS and CWC of the ANN-NB-DST method over the PerEn method across all weather conditions. As explained before, CRPS measures the overall *deviation* of the estimated CDF from the actual measurements, indicating the forecast accuracy, while CWC measures the *reliability* and *sharpness* of the forecasted CDF. Notably, the ANN-NB-DST method achieves over 33% improvement in CRPS under *clear* and *partially cloudy* weather conditions. This result

indicates that the CDF estimated by the ANN-NB-DST method deviates less from the actual solar power measurements than the PerEn method. Also, the improvement in CRPS is less pronounced ($< 10\%$) under *overcast* weather conditions, reflecting the challenges in predictive accuracy under such weather. Despite this, results from Figure 7 and Table 3 suggest the overall superiority of the ANN-NB-DST method in *accuracy* and *sharpness* with a slight reduction in *reliability*. Moreover, the CWC improvement remains consistent (more than 52%) across all the weather conditions, indicating the higher *reliability* and better *sharpness* of the ANN-NB-DST method than the benchmark method.

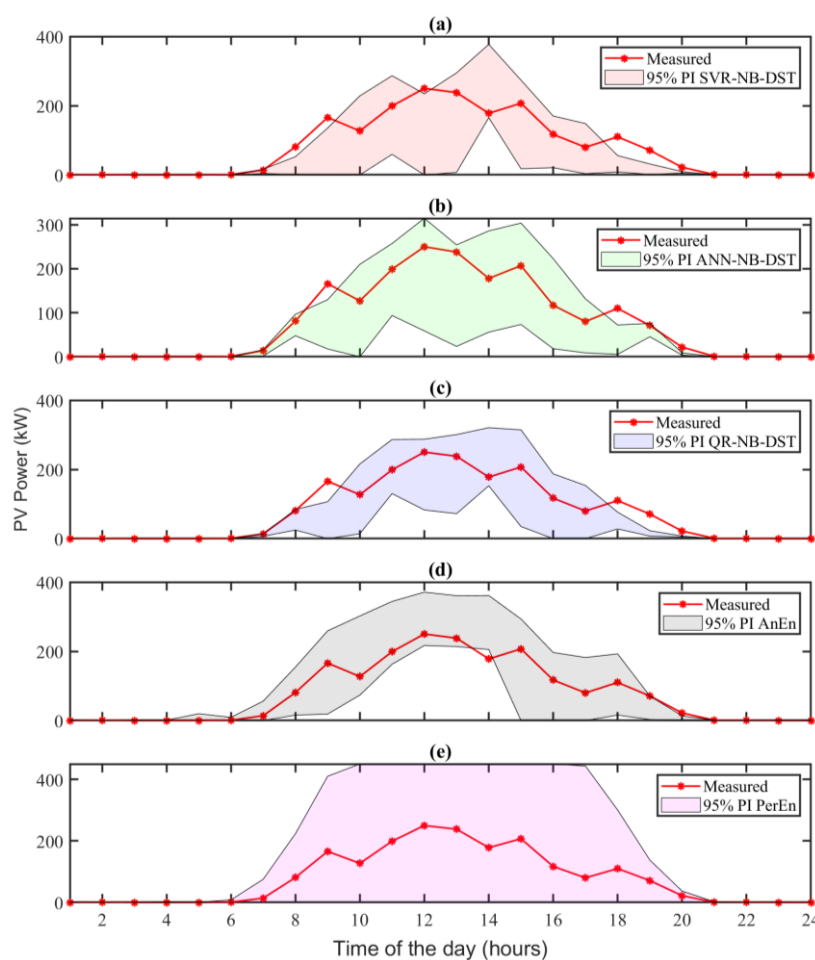


Figure 8. 95% PIs obtained from all the probabilistic solar forecasting methods on a particular “partially cloudy” day (09 July 2022) (a) SVR-NB-DST, (b) ANN-NB-DST, (c) QR-NB-DST, (d) AnEn, and (e) PerEn.

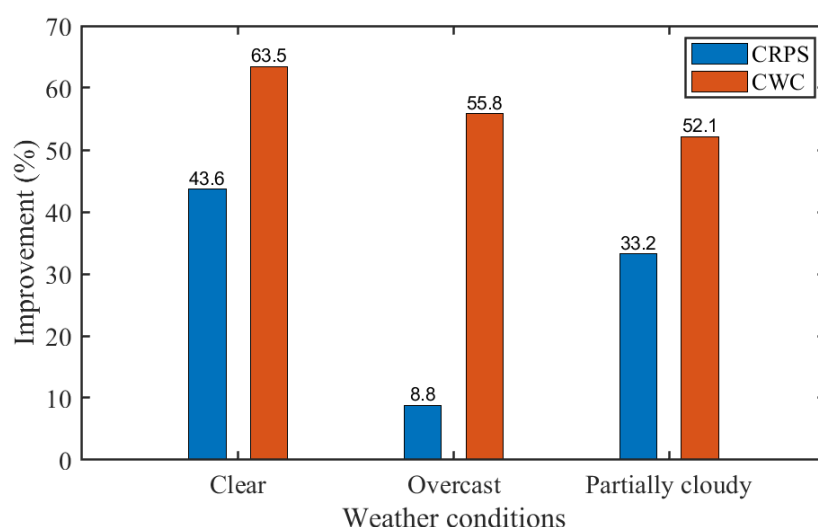


Figure 9. Comparative improvement of the ANN-NB-DST method over the widely used benchmark, the PerEn method.

5. Conclusions and future work

This paper proposes the NB-DST method, a novel approach for quantifying forecast errors of a deterministic forecasting method by estimating the CDF using a non-parametric technique. The NB-DST method has been integrated with deterministic forecasting methods such as SVR, ANN, and QR to generate probabilistic forecasts under different weather conditions. Comparisons with established benchmark methods (AnEn and PerEn) using real-world data reveal that the NB-DST methods consistently outperform the benchmarks in terms of lower CRPS, BS, and CWC values. Among the tested methods, the ANN-NB-DST method emerges as the top performer, consistently producing reliable and high-resolution CDF estimates across various weather conditions, establishing it as the superior option for probabilistic photovoltaic (PV) forecasting among the methods evaluated.

Future studies will focus on exploring a wider range of deterministic forecasting methods in conjunction with the NB-DST method to enhance forecast robustness and consistency. Additionally, this study considered only a limited number of meteorological variables due to data availability limitations. Thus, an important direction for future investigation is to incorporate a more comprehensive set of meteorological variables into the forecasting models. This expansion will enable a deeper exploration of their impact on forecasting accuracy and provide more comprehensive insights into the dynamics of probabilistic PV forecasting.

Author Contributions:

Tawsif Ahmad: Methodology, Data curation, Software, Visualization, Investigation, Validation, Writing – original draft. **Ning Zhou:** Conceptualization, Supervision, Writing – review & editing. **Ziang Zhang:** Writing – review & editing. **Wenyuan Tang:** Writing – review & editing.

Funding:

This paper is based on work supported by the U.S. National Science Foundation under grant #1845523 and by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0009341. The views, findings, conclusions, and recommendations presented in this paper are solely those of the authors and do not necessarily reflect those of the funding agencies.

Data Availability Statement:

The historical solar power generation data used in this study can be found at (<https://der.nyserda.ny.gov/reports/view/performance/?project=318>) and the related weather data can be downloaded from (<https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition>) [42] [44].

Conflicts of Interest:

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

References

1. B. Li and J. Zhang, "A review on the integration of probabilistic solar forecasting in power systems," *Solar Energy*, vol. 210, pp. 68–86, 2020.
2. J. M. M. González, A. J. Conejo, H. Madsen, P. Pinson and M. Zugno, "Integrating renewables in electricity markets: Operational problems," Springer Science & Business Media, vol. 205, p. 429, 2014.
3. A. Kaur, L. Nonnenmacher, H. T. C. Pedro and C. F. M. Coimbra, "Benefits of solar forecasting for energy imbalance markets," *Renewable Energy*, vol. 86, pp. 819–830, 2016.
4. G. Notton, M.-L. Nivet, C. Voyant, C. Paoli, C. Darras, F. Motte and A. Fouilloy, "Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting," *Renewable and Sustainable Energy Reviews*, vol. 87(C), pp. 96–105, 2018.
5. T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151, 2014.
6. M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing and Applications*, vol. 31, pp. 2727–2740, 2019.
7. H. Aprillia, H.-T. Yang and C.-M. Huang, "Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm," *Energies*, vol. 13, no. 8, p. 1879, 2020.
8. E. Michael, M. M. Neethu, S. Hasan and A. Al-Durra, "Short-term solar power predicting model based on multi-step CNN stacked LSTM technique," *Energies*, vol. 15, no. 6, p. 2150, 2022.
9. X. Huang, Q. Li, Y. Tai, Z. Chen, J. Zhang, J. Shi, B. Gao and W. Liu, "Hybrid deep neural model for hourly solar irradiance forecasting," *Renewable Energy*, vol. 171, pp. 1041–1060, 2021.
10. H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renewable Energy*, vol. 160, pp. 26–41, 2020.
11. A. R. Pazikadin, D. Rifai, K. Ali, M. Z. Malik, A. N. Abdalla and M. A. Faraj, "Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend," *Science of The Total Environment*, vol. 715, 2020.
12. A. P. Yadav, A. Kumar and L. Behera, "RNN based solar radiation forecasting using adaptive learning rate," in *Swarm, Evolutionary, and Memetic Computing: 4th International Conference, SEMCCO 2013*, Chennai, India, 2013.
13. C.H. Liu, J.C. Gu and M.T. Yang, "A simplified LSTM neural networks for one day-ahead solar power forecasting," *IEEE Access*, vol. 9, pp. 17174–17195, 2021.
14. C. Feng, Y. Liu and J. Zhang, "A taxonomical review on recent artificial intelligence applications to PV integration into power grids," *International Journal of Electrical Power & Energy Systems*, vol. 132, p. 107176, 2021.
15. M. Abuella and B. Chowdhury, "Solar power forecasting using support vector regression," in *International Annual Conference of the American Society for Engineering Management*, 2016.
16. A. Fentis, L. Bahatti, M. Mestari, M. Tabaa, A. Jarrou and B. Chouri, "Short-term PV power forecasting using support vector regression and local monitoring data," in *2016 International Renewable and Sustainable Energy Conference (IRSEC)*, 2016.
17. M. Abuella and B. Chowdhury, "Hourly probabilistic forecasting of solar power," in *2017 North American Power Symposium (NAPS)*, 2017.
18. A. Ahmed Mohammed and Z. Aung, "Ensemble learning approach for probabilistic forecasting of solar power generation," *Energies*, vol. 9, no. 12, p. 1017, 2016.
19. Y.-M. Zhang and H. Wang, "Multi-head attention-based probabilistic CNN-BiLSTM for day-ahead wind speed forecasting," *Energy*, vol. 278, no. A, 2023.
20. P. Lauret, M. David and H. T. Pedro, "Probabilistic solar forecasting using quantile regression models," *Energies*, vol. 10, no. 10, p. 1591, 2017.
21. S. Alessandrini, L. D. Monache, S. Sperati and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied energy*, vol. 157, pp. 95–110, 2015.
22. K. Doubleday, S. Jascourt, W. Kleiber and B.-M. Hodge, "Probabilistic solar power forecasting using bayesian model averaging," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 352–337, 2020.
23. F. Najibi, D. Apostolopoulou and E. Alonso, "Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast," *International Journal of Electrical Power & Energy Systems*, vol. 130, 2021.

24. H. Wang, Y.-M. Zhang and J.-X. Mao, "Sparse Gaussian process regression for multi-step ahead forecasting of wind gusts combining numerical weather predictions and on-site measurements," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 220, 2022. 620-622
25. A. Dahl and E. Bonilla, "Scalable Gaussian process models for solar power forecasting," in *Data Analytics for Renewable Energy Integration: Informing the Generation and Distribution of Renewable Energy: 5th ECML PKDD Workshop, DARE*, Skopje, Macedonia, 2017. 623-625
26. N. Wilson, "Algorithms for dempster-shafer theory," in *Handbook of Defeasible Reasoning and Uncertainty Management Systems: Algorithms for Uncertainty and Defeasible Reasoning*, 2000, pp. 421-475. 626-627
27. H. Wang, Y.-M. Zhang, J.-X. Mao and H.-P. Wan, "A probabilistic approach for short-term prediction of wind gust speed using ensemble learning," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 202, 2020. 628-629
28. M. Yang, Y. Lin and X. Han, "Probabilistic wind generation forecast based on sparse Bayesian classification and Dempster-Shafer theory," *IEEE Transactions on Industry Applications*, vol. 52, no. 3, pp. 1998-2005, 2016. 630-631
29. A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes," *Advances in Neural Information Processing Systems*, vol. 14, 2001. 632-633
30. A. Blázquez-García, A. Conde, U. Mori and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1-33, 2021. 634-635
31. T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing*, vol. 21, pp. 137-146, 2011. 636-637
32. T. Hastie, R. Tibshirani, J. H. Friedman and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, New York: Springer, 2009. 638-639
33. A. Katsounidis and I. Karali, "Dempster-Shafer Theory: How Constraint Programming Can Help," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems: 18th International Conference, IPMU 2020*, Lisbon, 2020. 640-641
34. D. W. Van der Meer, J. Munkhammar and J. Widén, "Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals," *Solar Energy*, vol. 171, pp. 397-413, 2018. 642-644
35. J. Juban, N. Siebert and G. N. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation," in *2007 IEEE Lausanne Power Tech*, 2007. 645-646
36. H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather and Forecasting*, vol. 15, no. 5, pp. 559-570, 2000. 647-648
37. C. A. Ferro, "Comparing probabilistic forecasting systems with the Brier score," *Weather and Forecasting*, vol. 22, no. 5, pp. 1076-1088, 2007. 649-650
38. P. Lauret, M. David and P. Pinson, "Verification of solar irradiance probabilistic forecasts," *Solar Energy*, vol. 194, pp. 254-271, 2019. 651-652
39. C. Li, G. Tang, X. Xue, X. Chen, R. Wang and C. Zhang, "The short-term interval prediction of wind power using the deep learning model with gradient descend optimization," *Renewable Energy*, vol. 155, pp. 197-211, 2020. 653-654
40. S. Alessandrini, L. D. Monache, S. Sperati and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy*, vol. 157, pp. 95-110, 2015. 655-656
41. Solar Energy Technologies Office (SETO), "American-Made Solar Forecasting Prize," 2021. [Online]. Available: <https://www.energy.gov/eere/solar/american-made-solar-forecasting-prize>. [Accessed 2023]. 657-658
42. "NYSERDA DER Integrated Data System," [Online]. Available: <https://der.nyserda.ny.gov/reports/view/performance/?project=318>. [Accessed 26 08 2023]. 659-660
43. R. H. Inman, H. T. Pedro and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535-576, 2013. 661-662
44. "Weather Data Services | Visual Crossing," [Online]. Available: <https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition>. [Accessed 26 08 2023]. 663-664
45. T. Hsing, L.-Y. Liu, M. Brun and E. R. Dougherty, "The coefficient of intrinsic dependence (feature selection using el CID)," *Pattern Recognition*, vol. 38, no. 5, pp. 623-636, 2005. 665-666
46. Z. Liu and Z. Zhang, "Solar forecasting by K-Nearest Neighbors method with weather classification and physical model," in *2016 North American Power Symposium (NAPS)*, Denver, CO, USA, 2016. 667-669