



## Article

# Probabilistic Intraday PV Power Forecast Using Ensembles of Deep Gaussian Mixture Density Networks

Oliver Doelle <sup>1,2,\*</sup> , Nico Klinkenberg <sup>3</sup>, Arvid Amthor <sup>4</sup> and Christoph Ament <sup>2</sup> <sup>1</sup> Siemens AG, Digital Industries—Data Visions, Siemenspromenade 1, 91058 Erlangen, Germany<sup>2</sup> Faculty of Applied Computer Science, University of Augsburg, 86159 Augsburg, Germany<sup>3</sup> Faculty of Business Studies and Information Technology, Westphalian University of Applied Sciences, 46397 Bocholt, Germany<sup>4</sup> Siemens AG, Technology, Günther-Scharowsky-Str. 1, 91058 Erlangen, Germany

\* Correspondence: oliver.doelle@siemens.com

**Abstract:** There is a growing interest of estimating the inherent uncertainty of photovoltaic (PV) power forecasts with probability forecasting methods to mitigate accompanying risks for system operators. This study aims to advance the field of probabilistic PV power forecast by introducing and extending deep Gaussian mixture density networks (MDNs). Using the sum of the weighted negative log likelihood of multiple Gaussian distributions as a minimizing objective, MDNs can estimate flexible uncertainty distributions with nearly all neural network structures. Thus, the advantages of advances in machine learning, in this case deep neural networks, can be exploited. To account for the epistemic (e.g., model) uncertainty as well, this study applies two ensemble approaches to MDNs. This is particularly relevant for industrial applications, as there is often no extensive (manual) adjustment of the forecast model structure for each site, and only a limited amount of training data are available during commissioning. The results of this study suggest that already seven days of training data are sufficient to generate significant improvements of 23.9% in forecasting quality measured by normalized continuous ranked probability score (NCRPS) compared to the reference case. Furthermore, the use of multiple Gaussian distributions and ensembles increases the forecast quality relatively by up to 20.5% and 19.5%, respectively.

**Keywords:** PV power; probabilistic forecast; MDN; Monte Carlo dropout; deep ensemble



**Citation:** 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>

Academic Editor: Hee-Je Kim

Received: 30 November 2022

Revised: 28 December 2022

Accepted: 1 January 2023

Published: 5 January 2023



**Copyright:** © 2023 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/>).

## 1. Introduction

### 1.1. Motivation and Background

Forecasting is becoming paramount for incorporating the increasing number of variable renewable energy (VRE) technologies into power systems. In particular, forecasts for photovoltaics (PV) systems are attracting increasing attention, after being declared the most immature area of energy forecasting by world-renowned energy forecasters as recently as 2016 [1]. This is primarily driven by the projected strong increase in grid penetration of PV systems. According to the current grid expansion plan of the German Federal Network Agency, for instance, the PV capacity in Germany is expected to increase 5.8 times (+268 GW) over the next 15 years [2]. Moreover, under the EU-wide European Solar Roof Initiative, all new residential, public and commercial buildings will be required to install PV roof systems by 2029 [3].

However, the inherent uncertainty of forecasting results in prediction errors, which in turn may lead to suboptimal operational schedules and unnecessarily high opportunity costs. A feasible solution is to estimate and quantify the prevailing uncertainty through probabilistic forecasts, allowing for improved market bidding strategies [4] and better operational planning [5]. Consequently, probabilistic forecasts can be beneficial for both grid operators and market participants. For instance, stochastic, scenario-based or robust model predictive controls [6,7] can be used to improve the robustness of operational

schedules by, e.g., forming adaptive reserve algorithms [8] or better economic dispatch models [9].

Forecasts for PV power are usually generated either directly or in a two-step approach by initially creating a forecast of the global horizontal irradiance (GHI) based on numerical weather prediction (NWP), followed by a conversion to PV power using static models [10]. This conversion is mostly achieved by first transforming the GHI with separation and transposition models into the solar irradiance on the plane of array and subsequently determining the PV power using a physical motivated model of the PV panel [10]. However, especially for the increasing number of local systems (e.g., rooftop or building-integrated systems with mixed orientation), the commissioning engineer often may not know all technical PV parameters (e.g., orientation, angle) required for the indirect approach. Moreover, measuring devices for local irradiance (e.g., pyranometers) need to be installed to calibrate the models, and shading issues have to be modelled separately. Given the mentioned drawbacks and the fact that potential users of the forecasts are grid operators and plant owners, this paper focuses on the direct prediction of PV power and its probability distribution.

### *1.2. Aleatoric and Epistemic Uncertainty*

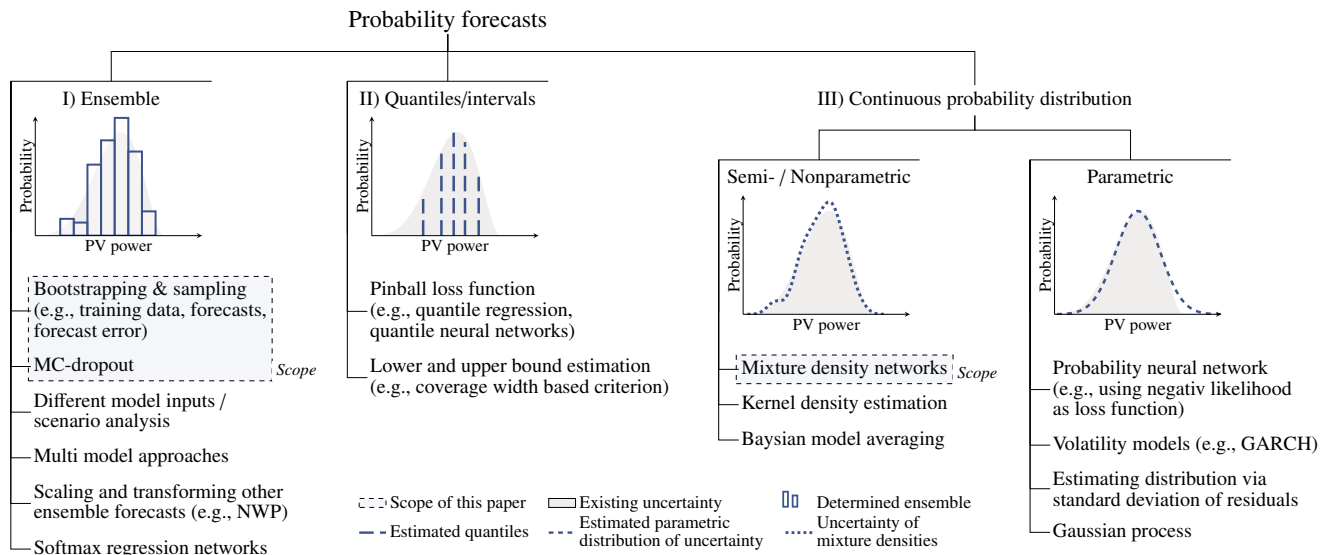
The inherent uncertainty of forecasts can be divided into epistemic and aleatoric uncertainty [11]. The latter denotes the stochastic component of the modelled process, such as noise in the data, which prevents a complete deterministic relation between the chosen model in- and outputs [12]. Epistemic uncertainty, in turn, is caused by the lack of knowledge about the perfect predictor and therefore encompasses e.g., the uncertainty caused by suboptimal model structures and suboptimal estimated parameters [12]. In terms of PV power forecasting, sources of aleatoric uncertainty may be inaccuracies in the input signals, such as NWP, and the omission of possible relevant information, such as wind speed and wind direction, in the model inputs. Epistemic uncertainty, meanwhile, can be caused by an insufficient amount of available training data, a (too) high model complexity, or the inability to find the global minimum during training, due to nonlinear models with respect to their parameters (e.g., neural networks). Correspondingly, epistemic uncertainty declines with more training data, whereas this has no effect on the aleatoric uncertainty [12]. Consequently, the former is modeled by placing a probability distribution over the model parameters, while the aleatoric uncertainty is modeled by placing a distribution over the model output [11]. The distinction between these two uncertainties is important as neglecting one of them will likely lead to an underestimation or misrepresentation of the overall uncertainty. For instance, in practice, a detailed manual adjustment of the forecast model is not always feasible and therefore a standard approach (e.g., a model with default structure and hyperparameters based on previous locations) is used instead. This, in turn, may considerably increase epistemic uncertainty and, therefore, if not taken into account, results in prediction intervals that are too narrow. However, several current studies do not consider a combined assessment of both uncertainties [13].

### *1.3. Probabilistic Forecasts*

Three different basic concepts (see Figure 1) are widely used for the representation and generation of probabilistic forecasts: (1) creation of ensemble forecasts, (2) identifying a discrete cumulative distribution function by e.g., quantiles, or (3) determining a continuous probability function via a parametric distribution or non-parametric depiction (e.g., kernels). In the following, these different approaches are discussed in detail using several examples.

Ensemble forecasts consist of different ensemble members—typically point forecasts—which are generated by e.g., bootstrapping [14], multi-model approaches [15,16], determining possible input deviations [17] or by selecting outputs from comparable situations in the past [18]. They can therefore assume any occurring distribution function. However, ensembles may require a post-processing to calibrate the prediction interval, and are often in comparison computationally more demanding [17]. Furthermore, they commonly only model one type of uncertainty. Training data bootstrapping, for instance, only depicts

epistemic uncertainty [19]. In addition to their standalone use, ensemble concepts can also be combined with other probabilistic approaches (e.g., an ensemble of quantile forecasts) to enable the modelling of all types of uncertainty.



**Figure 1.** Overview of different representation types for probabilistic forecasts and exemplary methods to generate them. While continuous probability distributions can be both parametric and non-parametric, ensembles and quantile representations never assume a parametric distribution. A combination of several approaches and the conversion of the different representation forms into each other is possible. The approaches explored in this paper are highlighted.

Estimating the cumulative distribution function (CDF) of probabilistic prediction discretely using, e.g., quantiles, is the most common approach for probabilistic forecasts [20]. In doing so, for each quantile  $\tau \in [0, 1]$ , a forecast  $\hat{y}_\tau(t)$  is estimated for the signal  $y(t)$ , where the probability of  $y(t)$  to be lower than  $\hat{y}_\tau(t)$  is exactly  $\tau$ :

$$P(y(t) < \hat{y}_\tau(t)) = \tau. \quad (1)$$

While in classical deterministic forecasts, the parameters are estimated by minimizing the mean squared error (MSE) of the residuals; in quantile regression, an asymmetrically weighted error is used for the loss function, e.g., the so-called pinball loss. Hence, this approach can be applied to several algorithms and is consequently often used to easily transform an existing deterministic forecast model into a probabilistic one [21]. For example, Refs. [19,20] provide a comparison and an overview of different linear prediction models based on quantile regression. The authors in [22], in turn, applied quantile regression to an encoder–decoder architecture that uses long short-term memory (LSTM) neural networks in combination with a multilayer perceptron. However, quantile regression only depicts the aleatoric uncertainty and therefore the uncertainty in the data, as the adjusted loss function only characterizes a distribution over the model output. In addition, for each to be a determined quantile, a separate model training is usually performed and some subsequent application methods (e.g., stochastic optimization) require a continuous representation of the CDF [7].

Alternatively, the complete probability distribution can also be estimated directly. In parametric approaches, a set distribution (e.g., a Gaussian distribution) is assumed a priori for the uncertainty and its parameters are subsequently estimated. This can be accomplished with additive volatility models such as generalized autoregressive conditional heteroscedasticity (GARCH) models, which provide a probabilistic extension for any deterministic prediction but require minimal computational effort [23]. Another approach is to estimate a normal distribution using the standard deviation of the residuals,

which has been implemented in many publicly available R (e.g., *tsibble*) and Python (e.g., *pmdarima*) forecasting packages thanks to its simplicity. However, this approach only considers the deviation and therefore uncertainty of the random error term and assumes homoscedasticity, which can lead to an underestimation of prediction intervals of up to 25% [24]. Analogous to quantile regression, the cost function of the prediction model can also be adapted in order to apply a parametric approach to different model structures (e.g., distributional neural networks). By using the negative log likelihood of a Gaussian distribution as the cost function, for example, the mean and the standard deviation can be estimated directly. Another prominent method is the Gaussian process, which has been used in [25] for the probabilistic prediction of PV power. Although this is a non-parametric forecasting method and does not assume a fixed number of model parameters, the resulting distribution is Gaussian and therefore parametric. However, parametric approaches have a disadvantage. Previous studies on irradiation forecasts have shown that the assumption of a fixed Gaussian distribution of the error terms is not always supported by the data due to, e.g., a lack of symmetry, resulting in lower forecast quality [26].

Flexible density estimation, e.g., via kernel functions, can solve this challenge. The authors of [27] used Bayesian Model Averaging as a post-processing method to generate a probabilistic mixture model out of NWP ensembles, combining a discrete component for power clipped at the inverter rating and a continuous portion for the lower output. The authors of [28], in turn, used a coupled input and forget gate network in combination with quantile regression to initially generate individual quantiles and afterwards converted them into a continuous probability distribution using kernel density estimation.

Another promising approach, which has been successfully applied to other forecasting domains [29], are mixture density networks (MDNs). They can be seen as extensions of distributional neural networks, as they combine different distributions using the sum of weighted negative log likelihoods of kernel functions as minimizing objectives. Consequently, they can estimate flexible uncertainty distributions with almost all neural network structures and thereby benefit from the advances in machine learning, e.g., deep neural networks. Moreover, with a sufficiently high number of Gaussian distributions, it is theoretically possible to represent any other distribution form [30]. Ref. [29] implemented it, for instance, for the probabilistic forecast for regional wind power. Furthermore, in [31], an MDN with four distributions was able to achieve a significantly better deterministic forecast than a linear transformation model for solar irradiation. Analogous to quantile regression, MDNs only depict the aleatoric uncertainty by changing the minimizing objective.

There are various possibilities to consider both aleatoric and epistemic uncertainty in neural networks. Nevertheless, the inclusion of the epistemic counterpart in MDNs is overall still largely unexplored [32] and, to the authors' knowledge, has not yet been applied to PV power forecasts. For instance, Ref. [32] addressed the epistemic uncertainty of MDNs only for short-term load forecasting through a Bayesian approach of estimating weight distributions over the model parameters instead of specific values. A disadvantage of this approach, however, is the relatively high computational effort.

A computationally less demanding alternative is Monte Carlo (MC) Dropout [33]. Here, dropout is not only activated during training, but also during testing of the network. By randomly dropping various units during forecasting, different results are generated. These can even be interpreted overall as a deep Gaussian process approximation when dropout is applied to each hidden layer [33]. Another epistemic extension is the repeated model training with different parameter initializations. In combination with the limitation to only one single Gaussian distribution, this approach is often termed a deep ensemble. Using a deep ensemble, Ref. [13] has achieved better results in probabilistic load forecasting for one and 24 h ahead than with a linear quantile regression model or even a deep Gaussian process. The effectiveness of the ensemble methods can be illustrated by the popular bagging [34] approach. It reduces the epistemic uncertainty by aggregating the ensemble members, which, in turn, improves the deterministic accuracy.

#### 1.4. Contributions and Scope

To the best of the authors' knowledge, a comprehensive study of probabilistic PV power, forecasts with MDN has not been conducted before. This study aims to close this gap and advance the field of probabilistic PV power forecasting, in particular by:

- Analyzing the probabilistic forecast quality of MDN PV power forecasts and extending them with different approaches to encompass the epistemic uncertainty;
- Examining the influence of different architectural configurations (e.g., number of mixture components/distributions and ensemble members);
- Assessing the significance of the different uncertainty types for industrial applications by varying the number of available past data for training and validation (7 and 182 days/half a year);

In the analysis, emphasis is placed on ensuring that practical conditions are met. Frequently, only a simple division between training and test set or a low (e.g., 4-fold) cross-validation is carried out. A variation of the number of training data, for instance to be able to estimate the performance during commissioning, does not take place. In industrial applications, however, the model structure and hyperparameters are rarely adapted over time or for different locations due to capacity reasons. Therefore, a solution that also takes into account the additional uncertainties involved should be preferred. For this purpose, 24 different commissionings of the forecast per location are simulated for 7 and 182 training days.

The investigations are performed with data from three PV systems from different locations in Germany and Austria.

The forecast horizon is six hours, as this time span is needed to also include the planning of the thermal side (e.g., storages, heat pumps) in the optimal dispatch calculation of on-site energy systems [5]. A time resolution of 15 min is chosen, which corresponds to the imbalance settlement period and therefore predominantly billing resolution in the European Union [35].

#### 1.5. Organization

The remainder of this paper is organized as follows: Section 2 briefly presents an overview of the used data. Subsequently, Section 3 introduces the generation of ensemble MDNs in detail. In Section 4, the metrics and benchmark used to evaluate the probabilistic forecasts are discussed. Afterwards, a detailed analysis of the achieved probabilistic forecasts with a critical discussion of their behavior is carried out in Section 5. Finally, Section 6 summarizes the findings of this study and presents possible next steps.

## 2. Data

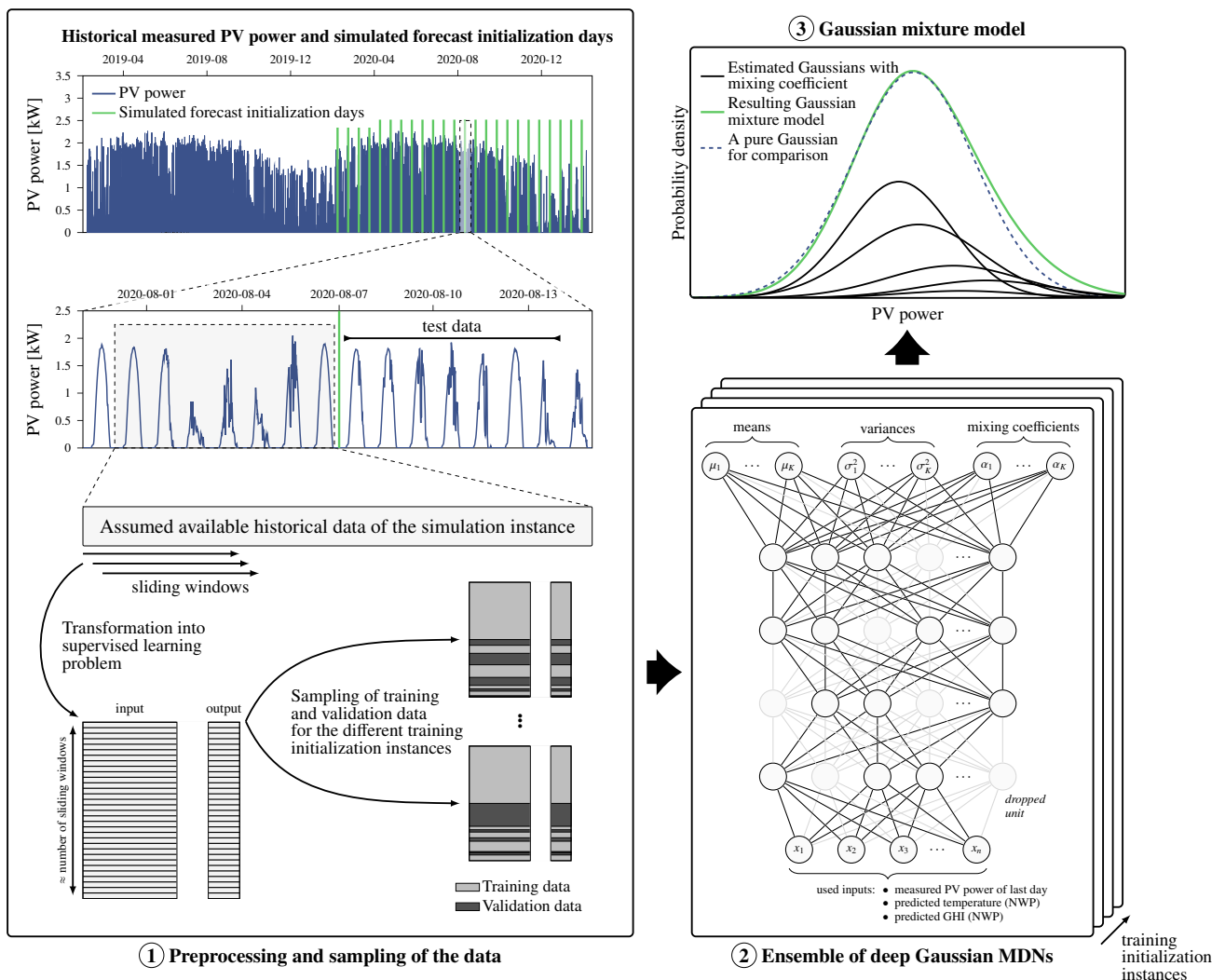
An overview of the studied PV systems is shown in Table 1. As all three of them are rooftop systems, their generated PV power is more volatile compared with ground-mounted systems due to the relatively small local spread (missing low pass effect). Alongside the measured PV power, GHI and outside temperature predicted by an external provider on the previous day serve as input signals for the forecasts. To achieve realistic practical conditions, the forecasts of the provider Meteonorm were continuously recorded at midnight of the respective previous day over the entire study period.

During preprocessing, gaps of less than 30 min were linearly interpolated and days with gaps larger than 30 min or days with snow on the panel were excluded from the analysis. Subsequently, 24 evenly distributed instances were selected and, for each of them, a forecast initialization was simulated for the next seven days using both 7 and 182 available days of training data, respectively (see also Figure 2).



**Table 1.** Main information concerning the data used in this work.

	North Bavaria (Germany)	South Bavaria (Germany)	Vienna (Austria)
Elevation [m]	280	725	150
Annual GHI [MWh/m <sup>2</sup> ]	1.77	1.88	1.79
Study period	08/19–02/21	01/19–09/21	05/17–04/19
Sample rate [min]	15	15	15
Ratio of missing days [%]	1.3	4.9	12.2
Number of samples	50,112	89,280	55,559
Solar variability ( $\sigma\Delta kt$ )	0.188	0.186	0.195
$P_{\text{peak}}$ , daily of PV panel [kW]	1.29	14.584	14.95
Installed capacity [kW]	2.2	24	27



**Figure 2.** Schematic illustration of the performed forecast approach. For each of the 24 different simulated forecast initializations, the past data (7 or 182 days) are used to train the model ①. To estimate the epistemic uncertainty, MC dropout and multiple training initializations with different samples of training and validation data are used ②. Subsequently, the ensembles of Gaussian distributions estimated by the Deep Mixture Density Networks are combined according to their mixture coefficients to estimate the overall forecast uncertainty ③. With a sufficient number of hidden units and a sufficient number of mixture distributions, this approach can theoretically approximate any probability distribution [30].

### 3. Methodology

First, the basic principle of MDN is presented in this chapter, followed by a detailed description of the extensions applied to also encompass the epistemic uncertainty. Figure 2 summarizes the entire concept of this study graphically for a clearer overall understanding.

#### 3.1. Mixture Density Networks

MDNs were initially introduced by Bishop in [30] to estimate general distribution functions. For this purpose, the conditional probability distribution  $p(\mathbf{y}|\mathbf{x})$  of the target variable  $\mathbf{y}$  given the input features  $\mathbf{x}$  is represented as a linear combination of kernel functions  $\phi_k(\mathbf{y}|\mathbf{x})$ :

$$p(\mathbf{y}|\mathbf{x}) = \sum_{k=1}^K \alpha_k(\mathbf{x}) \phi_k(\mathbf{y}|\mathbf{x}), \quad \text{with: } \sum_{k=1}^K \alpha_k(\mathbf{x}) = 1, \quad (2)$$

where  $K \in \mathbb{Z}^+$  is the amount, and  $k \in \mathbb{Z}^+$  is the respective number of the considered components in the mixture model. Furthermore,  $\alpha_k(\mathbf{x}) \in \mathbb{R}^+$  constitutes the weighting of the respective mixture component also called a mixing coefficient. In the present study, Gaussian kernels are used, since a neural network with a sufficient number of hidden units and a mixture model with a sufficient number of kernel functions can theoretically approximate any conditional density function [30]. Consequently,  $\phi_k(\mathbf{y}|\mathbf{x})$  is formulated as follows:

$$\phi_k(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_k^2(\mathbf{x})}} \exp\left(-\frac{\|\mathbf{y} - \mu_k(\mathbf{x})\|^2}{2\sigma_k^2(\mathbf{x})}\right), \quad (3)$$

with  $\mu_k(\mathbf{x}) \in \mathbb{R}$  as the mean and  $\sigma_k^2(\mathbf{x}) \in \mathbb{R}^+$  as the variance of the  $k$ th mixture component. To estimate the neural network parameter, the negative log likelihood is used as a minimization objective  $\mathcal{L}$ . Given (2) and (3), this results in:

$$\begin{aligned} \mathcal{L} &= -\log(p(\mathbf{y}|\mathbf{x})) \\ &= -\log\left(\sum_{k=1}^K \frac{\alpha_k(\mathbf{x})}{\sqrt{2\pi\sigma_k^2(\mathbf{x})}} \exp\left(-\frac{\|\mathbf{y} - \mu_k(\mathbf{x})\|^2}{2\sigma_k^2(\mathbf{x})}\right)\right). \end{aligned} \quad (4)$$

However, in practice, several measures must be implemented to guarantee that the Gaussian parameters comply with their mathematical constraints and that no numerical underflow occurs. As the logarithm of the product of exponential functions leads to an underflow issue for low values, the so-called log-sum-exp trick

$$\log \sum_{i=1}^w e^{\zeta_i} = \max_j \zeta_j + \log \sum_{i=1}^w e^{\zeta_i - \max_j \zeta_j}, \quad \forall w, i, j \in \mathbb{Z}^+ \quad (5)$$

is adopted, where  $\zeta \in \mathbb{R}^n$  denotes arbitrary values. For this purpose, the exponential function in (4) is reformulated:

$$\mathcal{L} = -\log\left(\sum_{k=1}^K \exp\left(\log(\alpha_k(\mathbf{x})) - \overbrace{\frac{1}{2} \log(2\pi)}^{\text{constant}} - \frac{1}{2} \log(\sigma_k^2(\mathbf{x})) - \frac{\|\mathbf{y} - \mu_k(\mathbf{x})\|^2}{2\sigma_k^2(\mathbf{x})}\right)\right). \quad (6)$$

For the mixing coefficients  $\alpha_k(\mathbf{x})$ , a softmax activation function is used, since they must be positive and sum to unity (see (2)). In addition, clipping ( $\alpha_k \in [1 \times 10^{-12}, 1], \forall k \in K$ ) is performed beforehand to guarantee that the mixing coefficients are for numerical stability purposes not too low. To ensure a positive variance, Bishop [30] initially suggested an exponential activation function. However, as such a function becomes unstable for large values, a softplus function with an additional constant minimum variance term was added

instead. This approach was also used in [36] for multiple regression tasks with a deep ensemble and only a single Gaussian distribution.

Consequently, the output layer of the neural network corresponds to a parameter vector  $[h_{[\mu_k]}, h_{[\sigma_k^2]}, h_{[\alpha_k]}]^T, \forall k \in [1, K]$ , which must be post-processed to get the parameters of the Gaussian mixture model (GMM) for both the loss function and the forecast, as follows:

$$\mu_k = h_{[\mu_k]} \quad (7a)$$

$$\sigma_k^2 = \log\left(1 + \exp\left(h_{[\sigma_k^2]}\right)\right) + 1 \times 10^{-6} \quad (7b)$$

$$\alpha_k = \frac{\exp\left(h_{[\alpha_k]}^*\right)}{\sum_{j=1}^K \exp\left(h_{[\alpha_j]}^*\right)}, \forall j \in \mathbb{Z}^+ \text{ with:} \quad (7c)$$

$$h_{[\alpha_k]}^* = \begin{cases} 1 \times 10^{-12}, & \text{if } h_{[\alpha_k]} \leq 1 \times 10^{-12} \\ 1, & \text{if } h_{[\alpha_k]} \geq 1 \\ h_{[\alpha_k]}, & \text{otherwise} \end{cases}. \quad (7d)$$

### 3.2. Epistemic Extension of Mixture Density Networks

As already outlined, MDNs only depict aleatoric uncertainty, as the uncertainty of the estimated model parameters is not considered. These uncertainties can be taken into account by determining multiple conditional distributions  $p_{\theta_m}(y|\mathbf{x}, \theta_m)$ , also called ensemble members, based on different estimated model parameters  $\theta_m \in \mathbb{R}$ . Analogous to bagging, these ensemble members can subsequently be averaged. This means for the present approach that the conditional distributions are combined in a higher-level mixture model. Considering (2) and (3), the overall distribution for the approach can thus be defined as:

$$\begin{aligned} p(y|\mathbf{x}) &= \frac{1}{M} \sum_{m=1}^M p_{\theta_m}(y|\mathbf{x}, \theta_m) \\ &= \sum_{m=1}^M \sum_{k=1}^K \frac{\alpha_{k,m}(\mathbf{x})}{M \sqrt{2\pi\sigma_{k,m}^2(\mathbf{x})}} \exp\left(-\frac{\|\mathbf{y} - \mu_{k,m}(\mathbf{x})\|^2}{2\sigma_{k,m}^2(\mathbf{x})}\right), \end{aligned} \quad (8)$$

where  $M \in \mathbb{Z}^+$  is the overall quantity of ensemble members, and  $m \in \mathbb{Z}^+$  indicates the respective ensemble member number. According to (8), the ensembles also influence the shape of the distribution function. Since the overall number of distribution is  $K$  times  $M$ , the density flexibility increases. For instance, even with only one distribution used at the output ( $K = 1$ ) and multiple ensembles ( $M > 1$ ), the result is a GMM. However, with well-chosen hyperparameters, the distributions between the generated ensemble members are expected to differ less than the generated output distributions within a single forecast.

For the generation of ensemble members, two different approaches are compared in this study. First, MC dropout is applied. In principle, dropout is used in neural networks only during training as a regularization method to avoid overfitting [37]. For this purpose, each unit or neuron  $i$  in a hidden layer is randomly switched off with the dropout probability  $p_i \in [0, 1]$  during each training sample. Consequently, it can be interpreted as if numerous thinned networks with shared weights are implicitly trained [37]. In the case of MC dropout, dropout is also applied during testing. Consequently, it can be expected that the network structure is different for each prediction leading to varying forecasts. Applied to each layer, this approximates a deep Gaussian process [33]. The primary advantage is that multiple ensemble members can be created despite a single model training.

In the second approach, multiple ensemble members are generated by repeated model training with different initial model parameters. This captures the deviation from the different local minima of the objective function (6) reached during model training. When compared to MC dropout, better results were achieved with this approach in [36] for



different regression tasks. Moreover, this approach often performs better in practice than Bayesian neural networks [38]. To account for the uncertainty caused by limited training data, the allocation to the training and evaluation dataset is carried out randomly before each training.

### 3.3. Training and Network Parameter

The implemented Gaussian MDN is a multilayer perceptron consisting of four hidden layers with 75 neurons in each layer. As recommended in [37], dropout is combined with a max-norm regularization of two at each layer to prevent the network weights from growing too large. Furthermore, different neural networks were trained for each sample of the forecast horizon as the number of network outputs already increases to  $k(c + 2)$  for each forecast horizon sample  $c \in \mathbb{Z}^+$  in comparison to the deterministic forecast. An overview of the used model and training hyperparameters is shown in Table 2. Due to the relatively long computing times, they were based on sample investigations using grid search of one location and, as is common in practice, adopted for all sites. Accordingly, marginally better results could be expected, if the model structure was individually tailored for each location.

**Table 2.** Summarized information of training- and MDN hyperparameters.

Architecture hyperparameters	
Number of hidden layers	4
Number of units per hidden layer	75
Activation function	ReLU
Input feature information [number of samples]	
PV power	~last day [97]
Predicted temperature (NWP)	forecast horizon [24] last 3 h [12]
Predicted GHI (NWP)	forecast horizon [24] ~last day [97]
Regularization techniques	
Dropout rate (in each hidden layer)	0.35
Weigh constraint of max norm	2
Early stopping with patience level	150
Training hyperparameters	
Minibatch size	32
Number of epochs	500
Optimizer	Adam
Validation split	0.3

## 4. Evaluation Framework

The quality of a probabilistic forecast should be assessed both on reliability and on sharpness [39]. The latter refers to the width of the probability distribution or the distance between ensemble members and is a measure of information density. Reliability, in turn, characterizes whether the predicted distribution corresponds to the observed distribution. This study follows the common approach from [26,39], which both recommend a quantitative evaluation based on an error metric and a qualitative graphical evaluation based on, e.g., a rank histogram. For this purpose, a normalization of the continuous ranked probability score (CRPS) is used, which is proper and can be denoted as:

$$\text{NCRPS} = \frac{1}{\bar{P}_{\text{peak, daily}}} \frac{1}{N} \sum_{l=1}^N \int_{-\infty}^{\infty} [\hat{F}_l(\tilde{y}) - \mathbb{1}(y_l - \tilde{y})]^2 d\tilde{y}, \quad (9)$$

where  $\hat{F}_l$  represents the predicted CDF of the variable of interest  $\tilde{y}$  (PV power) for the  $l$ th forecast and observation pair.  $\mathbb{1}$  is the Heaviside step function, which changes from 0 to 1, when the observation  $y_i > \tilde{y}$  and  $N \in \mathbb{Z}^+$  is the number of respective forecast and observation pairs of the analysis. The normalization is performed by dividing the CRPS

with the mean maximum daily produced power  $\overline{P_{\text{peak, daily}}}$  (see Table 1) of the PV panels. The CRPS has the same dimension as the predicted variable, and a lower value indicates a higher forecasting quality.

To assess the performance of the respective probabilistic methods against a reference, this study uses, inter alia, the complete-history persistence ensemble (CH-PeEn) (see Algorithm 1). The authors in [40] recommend it based on a comparative study against other benchmark methods for solar irradiation.

---

**Algorithm 1:** Complete-history persistence ensemble

---

- 1 Calculate the clear sky index  $\kappa$  for the PV power,  $P_{\text{PV}}$  at time  $t$ :  $\kappa(t) = P_{\text{PV}}(t)/P_{\text{PV, csp}}(t)$ , where the clear sky profile  $P_{\text{PV, csp}}$  is determined by a moving horizon of the maximum values in the past seven days of the same time of day
  - 2 Generate a forecast ensemble by using all past values of  $\kappa$  in the same hour
  - 3 Multiply the ensembles of clear sky indices with  $P_{\text{PV, csp}}$  to obtain the complete-history persistence ensemble
- 

In addition, the influence of the number of mixture distributions is also evaluated. Consequently, the solution with only one Gaussian output distribution, often referred to as deep ensemble, automatically serves as an additional, more advanced, state-of-the-art benchmark. For instance, in [13], this approach had achieved better results than, e.g., a deep Gaussian process for load prediction.

For a clear evaluation of the improvement related to different hyperparameters (e.g., number of distributions, ensemble generation method) and a better comparison to the benchmark methods, this study also uses the skill score, which is calculated as follows:

$$SS = \left(1 - \frac{\text{NCRPS}_{\text{forecast}}}{\text{NCRPS}_{\text{ref}}}\right) \cdot 100\%. \quad (10)$$

A higher value of the SS signifies a relatively greater improvement of the prediction accuracy with regard to the respective used reference case  $\text{NCRPS}_{\text{forecast}}$ , with the threshold of one indicating a perfect forecast. Conversely, negative values indicate a worse forecast performance than the benchmark.

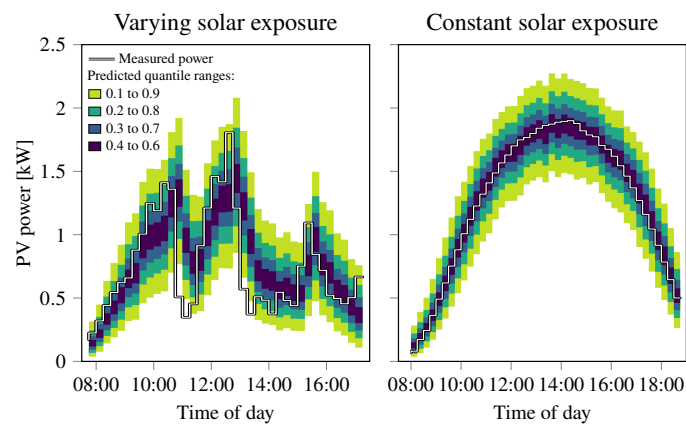
Time instances with marginal PV power generation (PV power < 3% of the respective  $P_{\text{peak, daily}}$ ) are neglected for the evaluation, since they would disproportionately impact the relative error metrics [10]. Nevertheless, the remaining considered time points possess more than 97% of the generated electric energy.

## 5. Results

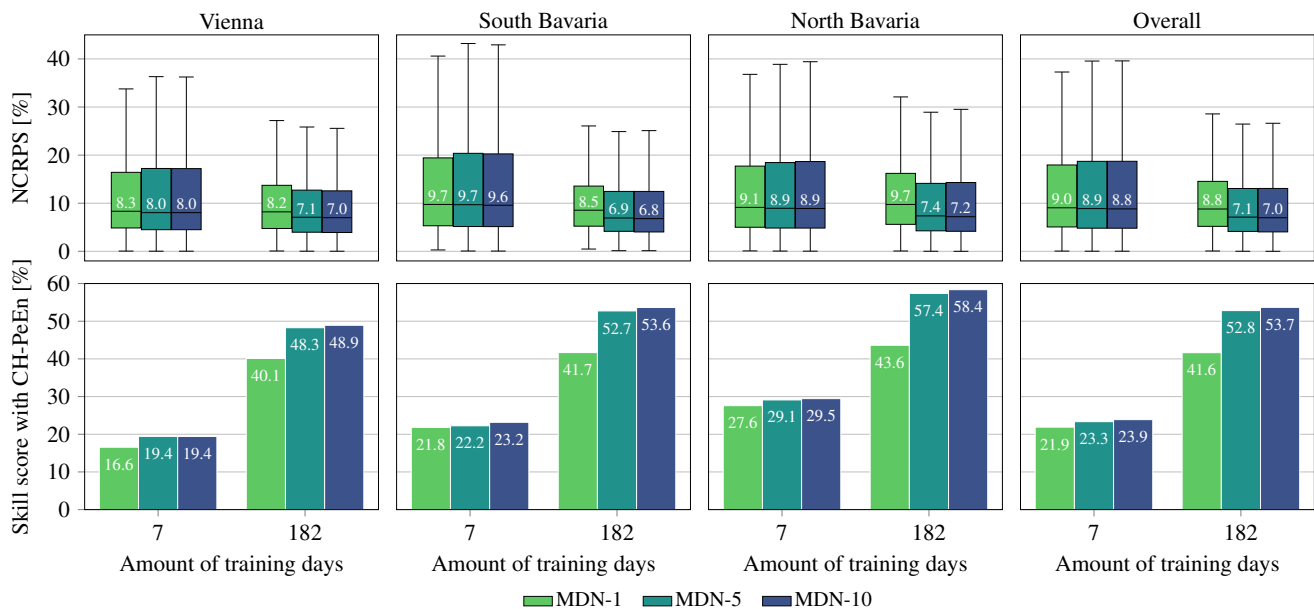
First, the influence of the number of mixture distributions is analyzed in detail, followed by the comparison of the different approaches to capture the epistemic uncertainty in MDN models. The temporal behavior of the forecasts are shown in Figure 3 for a better comprehension of the probabilistic results.

### 5.1. Influence of Training Data and Mixture Distributions

Figure 4 illustrates both the probabilistic accuracy and the relative improvement of the MDN compared to the reference forecast, each depending on the amount of training data and the number of mixture distributions. A significant improvement of 16.6% to 27.6% compared to the benchmark can already be observed with seven days of training data and a single Gaussian output distribution. MDNs can therefore be used with relatively limited available training data, e.g., during commissioning. Prediction accuracy increases considerably with half a year of training data, as can be seen by the reduced interquartile range and median in the box plots. The median NCRPS for ten mixture distributions, for instance, decreases from 8.8% to 7.0%, corresponding to a relative improvement of 20.5% and an improvement over the CH-PeEn benchmark of 53.7%.



**Figure 3.** Probabilistic forecast of the MDN for a day with varying and constant solar exposure. For better illustration, the estimated probability distributions of the mixture distributions are converted into intervals, and only the next step ahead forecast (forecast horizon corresponds to 15 min) is shown. The width of the intervals depends on both the level and the volatility of the PV power.

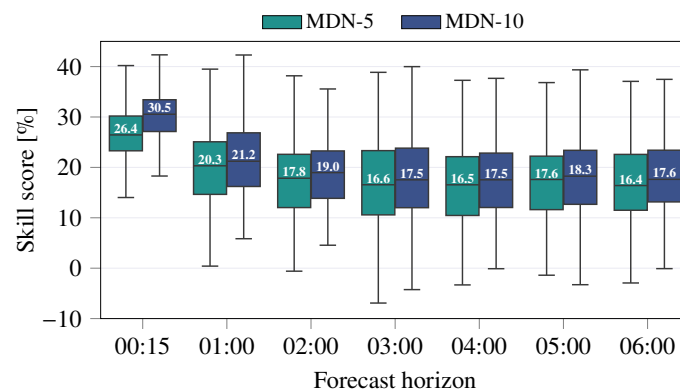


**Figure 4.** Accuracy of the probabilistic forecast for the different location and overall—for different amounts of available training data and different numbers of mixture distributions (top row). The comparison to the benchmark forecast CH-PeEn is also illustrated by the skill scores (bottom row). The results are based on a forecast horizon of 6 h and MDN ensembles with 15 training initializations, which in turn have 15 dropout ensembles each. The number after the abbreviation MDN in the legend indicates the respective distribution quantity, e.g., MDN-1 means one output distribution. The whiskers in the box plot span 1.5 times the interquartile range, which extends from the 25th to the 75th percentile.

The use of additional output distributions in the mixture model also enhances the prediction quality. Accordingly, an improvement of the MDN to the deep ensemble reference forecast with only one Gaussian output distribution can also be observed. This effect is more distinct with half a year of training data, resulting in a relative improvement of e.g., 20.4% from one to five mixture distributions. Hence, firstly, the more complex model structure is better at utilizing the additional information provided by the extra data and, secondly, a certain amount of training data are required to exploit the potential of the more flexible distribution mixture. From five to ten distributions, no significant additional improvement occurred, indicating that the underlying uncertainty distribution of the forecast

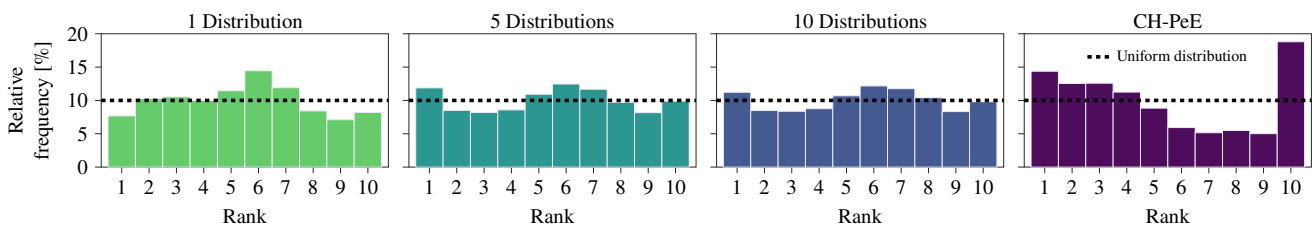
can already be estimated relatively accurately with five output distributions. Moreover, it should not be forgotten that slightly different Gaussian distributions are already included in the mixture model due to the ensemble members.

The increase in forecast quality due to additional distributions is further illustrated by the more detailed analysis regarding the forecast horizon in Figure 5. For shorter forecast horizons, both five and ten output mixture distributions provide even better improvements of up to 30.5% compared with a single Gaussian distribution, which suggests that the added value of MDNs is even greater at short forecast horizons. However, with increasing forecast horizon, the improvement by the additional mixture distributions and also the difference from five to ten distributions decreases. This is probably caused by the fact that the determination of uncertainty becomes more difficult at distant time horizons. The accompanying lower forecast quality makes a significant distinction harder.

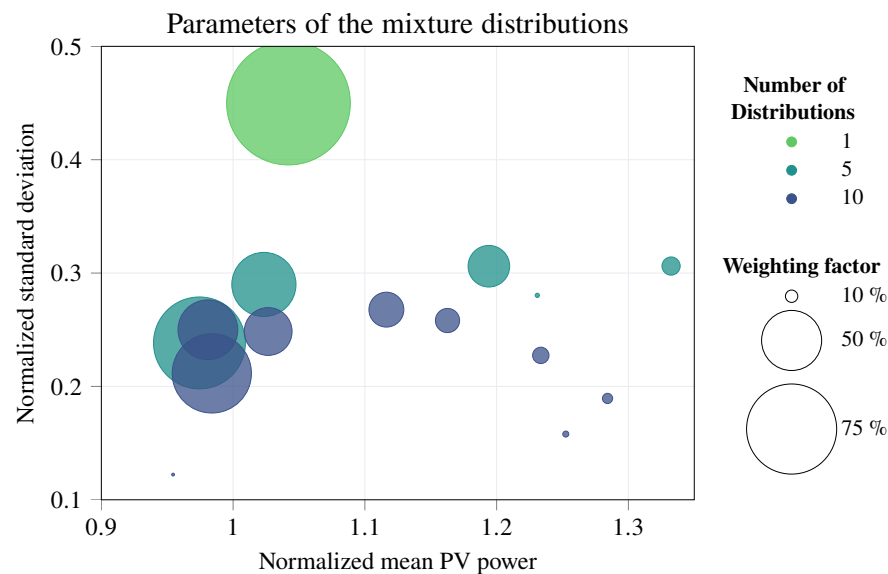


**Figure 5.** Forecast quality of the MDN depending on the forecast horizon for different numbers of mixture distributions of 182 days of training data. The forecast quality of the MDN with a single output distribution serves in each case as a reference value for the skill score. Consequently, the values represent the percentage improvement of the forecast quality caused by the additional used mixture distributions.

The rank diagrams in Figure 6 show a good reliability in comparison to the benchmark method and a slight improvement with the added mixture distributions. Since the underlying uncertainty at the output does not exactly resemble a normal distribution, a slight bias occurs at the output with only one distribution. Consequently, the sixth percentile is slightly overestimated. This context is demonstrated further by Figure 7, which shows the parameters (standard deviation, mean value, weighting factor) of the individual output distributions normalized by the measured PV power for different numbers of mixture distributions. In the case of a single distribution, the mean is slightly overestimated, and the standard deviation is comparatively higher, in order to include and represent also extreme values in the uncertainty estimation. With multiple mixture models, these extreme values can, in turn, be modeled by the additional distributions with higher normalized mean values and lower standard deviations and weighting factors. Thus, instead of one broad distribution, multiple narrower distributions are combined with each other. Thereby, the distributions with the smallest distance to the true value, which is in the normalized representation the value one, have the largest weighting factors. As the number of mixture distributions increases, their weighting factors decrease significantly. This also leads to the conclusion that additional distributions probably do not improve the forecast quality, and at the same time may cause numerical problems considering a possible underflow in (6).



**Figure 6.** The rank histogram for different numbers of distributions at the output in comparison to the benchmark method. The predictions are reliable if the observed quantiles occur with a frequency equal to their nominal probabilities, which is illustrated by the dashed line. Deviations from the nominal distribution entail an under- or overestimation of the specific quantile. Hence, a  $\cap$ -shape denotes overdispersed while a  $\cup$ -shape denotes underdispersed predictions.



**Figure 7.** Representation of the parameters of the different mixture distributions depending on their number in the model output. The size of the markers reflects their respective weighting factor in the mixture model. In order to enable comparability for varying power levels, both the standard deviation and the mean of the distributions were normalized by dividing the values with the respective measured PV power.

### 5.2. Benefit of Epistemic Estimation in MDNs

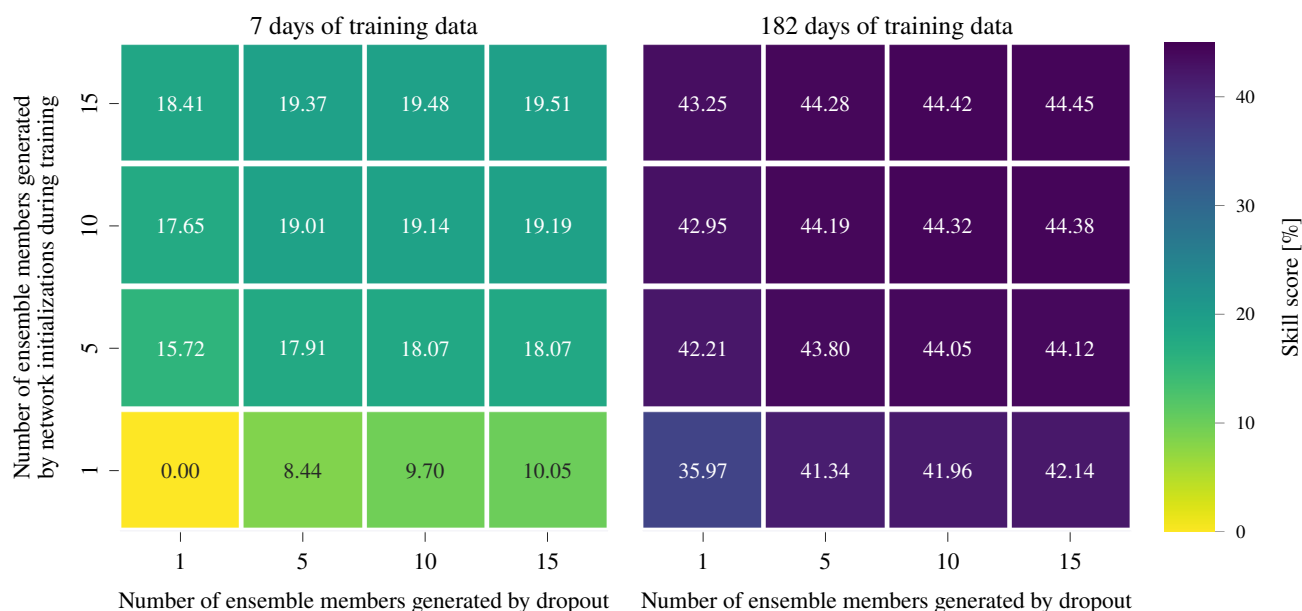
The influence of the extensions to estimate epistemic uncertainty are summarized in Figure 8. Both the network initializations and MC dropout have a significant positive effect on the forecast performance. For example, the use of dropout ensemble members alone improves the forecast quality by up to 10.05% and additional network initializations by up to 18.41% for seven days of training data. The impact of MC dropout is therefore slightly lower, probably because dropout ensemble members tend to focus on a single mode of the loss landscape, whereas different network initializations tend to explore diverse modes in the function space [38]. Moreover, since the training and validation data are also sampled, the variety and information in the training data are also higher for the multiple network initializations. Nevertheless, MC dropout needs less computational resources and is faster, as the model training does not have to be performed multiple times. For both methods, the added value decreases significantly as the number of ensemble members increases. At least a few ensemble members should therefore be considered in practice, since considerable great advances can already be achieved with relatively little effort.

The improvement of the forecast quality by the two approaches is larger for 7 days of training data than for 182 days, since the epistemic uncertainty decreases with increasing number of training data. Accordingly, the epistemic extensions to the MDN are particularly



recommended for applications in practice, when the number of training data may be limited during commissioning, and no corresponding individual adjustment of the network structure is made.

Prediction accuracy depending on the number and type of ensemble members



**Figure 8.** Influence of the number of MC dropout members and network initializations during training on the forecast quality depending on the amount of available training data. The data include all forecasts over the entire horizon (six hours) with ten mixed distributions. To enable a clearer analysis of the benefits of the epistemic extensions, the prediction with one dropout member and one initialization member was used as a reference value for the skill score.

## 6. Conclusions and Outlook

This paper broadens the field of probabilistic PV power forecasting by introducing MDNs and extending them with two epistemic estimation approaches. To emulate practical conditions, the investigations were carried out using 24 simulated forecast initializations at three different locations, each with a limited (7 days) and more extensive (182 days/half a year) amount of available training data. The analyzed forecast horizon was six hours with a sample rate of 15 min. For the estimation of the epistemic uncertainty, on the one hand, MC dropout and, on the other hand, multiple trainings initialization with a sampling of the training and validation data were used. Given those conditions, the major findings of this study are:

- By using multiple Gaussian distributions in a mixture model, the uncertainty distribution can be estimated more accurately. For example, when using ten Gaussian distributions instead of one, the NCRPS on average increases by 30.5% for a one step ahead prediction and by 20.5% averaged for the predictions over six hours. Consequently, the MDN can be seen as an improvement to a deep ensemble reference forecast with only one Gaussian output distribution;
- MDN forecasts achieve relatively good results even with a limited amount of training data. Already with 7 days, the MDN was 23.9% better than the CH-PeEn benchmark;
- Epistemic uncertainty has a major impact of up to 19.5% on the overall accuracy of MDN forecasts, especially when the amount of training data are limited. Therefore, compensation methods should always be considered in practice;
- Already a few ensemble members are capable of achieving significant forecast improvements. For example, five model initialization may result in a quality increase of up to 15.72%;

- Ensemble members generated by model initialization in combination with training data sampling improved the forecast quality relatively up to 83% more than ensemble members generated by MC dropout.

In the future, further investigations should be carried out by using more advanced model architectures. Particularly, temporal convolutional neural networks (TCNN) and models based on neural basis expansion analysis for time series (N-BEATS) could be combined with the MDN approach and different ensemble generating methods. These forecasts could then be compared to the feedforward model outlined in this paper.

**Author Contributions:** Conceptualization, O.D. and N.K.; methodology, O.D.; software, O.D. and N.K.; validation, O.D.; formal analysis, O.D.; investigation, O.D.; resources, O.D. and A.A.; data curation, O.D.; writing—original draft preparation, O.D.; writing—review and editing, O.D., N.K., A.A. and C.A.; visualization, O.D. and N.K.; supervision, A.A. and C.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The collection of the data used for this work was supported by the research team of Aspern Smart City Research (ASCR).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

CDF	Cumulative distribution function
CRPS	Continuous ranked probability score
CH-PeEn	Complete-history persistence ensemble
GHI	Global horizontal irradiance
GMM	Gaussian mixture model
MC	Monte Carlo
MDN	Mixture density networks
NCRPS	Normalized continuous ranked probability score
NWP	Numerical weather prediction
MSE	Mean squared error
PV	Photovoltaics
SS	Skill score
VRE	Variable renewable energy

## References

1. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [\[CrossRef\]](#)
2. Bundesnetzagentur. *Genehmigung des Szenariorahmens 2023–2037/2045*; Technical Report; Bundesnetzagentur: Bonn, Germany, 2022.
3. European Commission. *EU Solar Energy Strategy*; European Commission: Brussels, Belgium, 2022.
4. Dent, C.J.; Bialek, J.W.; Hobbs, B.F. Opportunity Cost Bidding by Wind Generators in Forward Markets: Analytical Results. *IEEE Trans. Power Syst.* **2011**, *26*, 1600–1608. [\[CrossRef\]](#)
5. El-Baz, W.; Tzscheutschler, P.; Wagner, U. Day-ahead probabilistic PV generation forecast for buildings energy management systems. *Sol. Energy* **2018**, *171*, 478–490. [\[CrossRef\]](#)
6. Gross, A.; Lenders, A.; Zech, T.; Wittwer, C.; Diehl, M. Using Probabilistic Forecasts in Stochastic Optimization. In Proceedings of the 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Liege, Belgium, 18–21 August 2020; IEEE: New York, NY, USA, 2020; pp. 1–6. [\[CrossRef\]](#)
7. Li, B.; Zhang, J. A review on the integration of probabilistic solar forecasting in power systems. *Sol. Energy* **2020**, *210*, 68–86. [\[CrossRef\]](#)
8. Bucksteeg, M.; Niesen, L.; Weber, C. Impacts of Dynamic Probabilistic Reserve Sizing Techniques on Reserve Requirements and System Costs. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1408–1420. [\[CrossRef\]](#)
9. Lorca, Á.; Sun, X.A. Adaptive Robust Optimization With Dynamic Uncertainty Sets for Multi-Period Economic Dispatch Under Significant Wind. *IEEE Trans. Power Syst.* **2015**, *30*, 1702–1713. [\[CrossRef\]](#)

10. Yang, D.; Alessandrini, S.; Antonanzas, J.; Antonanzas-Torres, F.; Badescu, V.; Beyer, H.G.; Blaga, R.; Boland, J.; Bright, J.M.; Coimbra, C.F.; et al. Verification of deterministic solar forecasts. *Sol. Energy* **2020**, *210*, 20–37. [[CrossRef](#)]
11. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2017; pp. 5575–5585.
12. Hüllermeier, E.; Waegeman, W. Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods. *arXiv* **2019**, arXiv:1910.09457.
13. Al-gabalawy, M.; Hosny, N.S.; Adly, A.R. Probabilistic forecasting for energy time series considering uncertainties based on deep learning algorithms. *Electr. Power Syst. Res.* **2021**, *196*, 107216. [[CrossRef](#)]
14. Grantham, A.; Gel, Y.R.; Boland, J. Nonparametric short-term probabilistic forecasting for solar radiation. *Sol. Energy* **2016**, *133*, 465–475. [[CrossRef](#)]
15. Zemouri, N.; Bouzgou, H.; Gueymard, C.A. Multimodel ensemble approach for hourly global solar irradiation forecasting. *Eur. Phys. J. Plus* **2019**, *134*, 594. [[CrossRef](#)]
16. Bracale, A.; Carpinelli, G.; De Falco, P. A Probabilistic Competitive Ensemble Method for Short-Term Photovoltaic Power Forecasting. *IEEE Trans. Sustain. Energy* **2017**, *8*, 551–560. [[CrossRef](#)]
17. Sperati, S.; Alessandrini, S.; Delle Monache, L. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. *Sol. Energy* **2016**, *133*, 437–450. [[CrossRef](#)]
18. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy* **2015**, *157*, 95–110. [[CrossRef](#)]
19. Doelle, O.; Kalysh, I.; Amthor, A.; Ament, C. Comparison of intraday probabilistic forecasting of solar power using time series models. In Proceedings of the 2021 International Conference on Smart Energy Systems and Technologies (SEST), Vaasa, Finland, 6–8 September 2021; IEEE: New York, NY, USA, 2021; pp. 1–6. [[CrossRef](#)]
20. Lauret, P.; David, M.; Pedro, H.T. Probabilistic solar forecasting using quantile regression models. *Energies* **2017**, *10*, 1591. [[CrossRef](#)]
21. van der Meer, D.; Widén, J.; Munkhammar, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1484–1512. [[CrossRef](#)]
22. Dumas, J.; Cointe, C.; Fettweis, X.; Cornelusse, B. Deep learning-based multi-output quantile forecasting of PV generation. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 27 June–2 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–6. [[CrossRef](#)]
23. David, M.; Ramahatana, F.; Trombe, P.J.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Sol. Energy* **2016**, *133*, 55–72. [[CrossRef](#)]
24. Hyndman, R.J.; Koehler, A.B.; Snyder, R.D.; Grose, S. A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecast.* **2002**, *18*, 439–454. [[CrossRef](#)]
25. Najibi, F.; Apostolopoulou, D.; Alonso, E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106916. [[CrossRef](#)]
26. David, M.; Luis, M.A.; Lauret, P. Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data. *Int. J. Forecast.* **2018**, *34*, 529–547. [[CrossRef](#)]
27. Doubleday, K.; Jascourt, S.; Kleiber, W.; Hodge, B.M. Probabilistic Solar Power Forecasting Using Bayesian Model Averaging. *IEEE Trans. Sustain. Energy* **2021**, *12*, 325–337. [[CrossRef](#)]
28. Liu, R.; Wei, J.; Sun, G.; Muyeen, S.M.; Lin, S.; Li, F. A short-term probabilistic photovoltaic power prediction method based on feature selection and improved LSTM neural network. *Electr. Power Syst. Res.* **2022**, *210*, 108069. [[CrossRef](#)]
29. Zhang, H.; Liu, Y.; Yan, J.; Han, S.; Li, L.; Long, Q. Improved Deep Mixture Density Network for Regional Wind Power Probabilistic Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 2549–2560. [[CrossRef](#)]
30. Bishop, C.M. *Mixture Density Networks*; Technical Report; Department of Computer Science and Applied Mathematics Aston University: Birmingham, UK, 1994.
31. Vallejo, D.; Chaer, R. Mixture Density Networks applied to wind and photovoltaic power generation forecast. In Proceedings of the 2020 IEEE PES Transmission & Distribution Conference and Exhibition—Latin America (T & D LA), Montevideo, Uruguay, 28 September–2 October 2020; IEEE: New York, NY, USA, 2020; pp. 1–5. [[CrossRef](#)]
32. Brusaferrri, A.; Matteucci, M.; Spinelli, S.; Vitali, A. Probabilistic electric load forecasting through Bayesian Mixture Density Networks. *Appl. Energy* **2022**, *309*, 118341. [[CrossRef](#)]
33. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; Volume 48, pp. 1651–1660.
34. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
35. European Commission. EB GL: Commission Regulation (EU) 2017/2195 establishing a guideline on electricity balancing. *Off. J. Eur. Union* **2017**, *2017*, 312/6–312/53.
36. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2016; pp. 6403–6414.
37. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

38. Fort, S.; Hu, C.H.; Lakshminarayanan, B. Deep Ensembles: A Loss Landscape Perspective. *arXiv* **2020**, arXiv:1912.02757.
39. van der Meer, D. Comment on “Verification of deterministic solar forecasts”: Verification of probabilistic solar forecasts. *Sol. Energy* **2020**, *210*, 41–43. [[CrossRef](#)]
40. Doubleday, K.; Van Scyoc Hernandez, V.; Hodge, B.M. Benchmark probabilistic solar forecasts: Characteristics and recommendations. *Sol. Energy* **2020**, *206*, 52–67. [[CrossRef](#)]

**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.