

# A Bayesian Deep Learning Technique with an Advanced Divergence Scheme for Multi-Step Ahead Solar Generation Forecasting

Devinder Kaur, Shama Naz Islam, and Md. Apel Mahmud

**Abstract** - In this paper, we propose an improved Bayesian bidirectional long-short term memory (BiLSTM)-based deep learning technique for multi-step ahead (MSA) solar generation forecasting. The proposed probabilistic technique applies alpha-beta divergence for a more appropriate consideration of outliers in solar generation data and resulting variability of the weight parameter distribution in Bayesian neural networks. The proposed method is examined on highly granular solar generation data from Ausgrid using probabilistic evaluation metrics such as Pinball loss and Winkler score. Moreover, a comparative analysis is provided to test the effectiveness of the improved MSA forecasting method for variable forecasting horizons. The numerical results clearly demonstrate that the improved Bayesian BiLSTM with alpha-beta divergence outperforms standard Bayesian BiLSTM and other benchmark methods for MSA forecasting in terms of error performance.

**Keywords:** Multi-step ahead forecasting, Bayesian deep learning, power systems, renewable generation forecasting.

## I. INTRODUCTION

With the rapid increase in carbon emissions, there is a dire need of clean and sustainable energy systems in the world today. In this regard, solar power is one of the fastest growing renewable energy sources (RES) penetrating into the power grid around the globe [1]. With the substantial deployment of smart meters in the modern power grid, a huge amount of data related to energy generation and electricity demand can be acquired [2]. This spatio-temporal energy data can be utilized by forecasting methods to predict future energy demand and generation to support grid operations [3].

Over the past decades, statistical and machine learning (ML) methods such as autoregressive [4] and support vector regression (SVR) [5] respectively, have been widely utilized to forecast future energy trends related to customer electricity demand [6] and renewable power generation involving rooftop solar panels [7]. While majority of existing literature in energy forecasting involves single-step ahead (SSA) predictions over short time horizons, it is not sufficient for the long-term planning applications such as generation scheduling where bids need to be submitted in a day-ahead manner [8]. Moreover, having multiple values predicted in long- and short-term horizons, the variability and projected abnormality can be analyzed more efficiently [9]. Thus, it is crucial to forecast multi-step ahead (MSA) energy generation with longer time horizons [10].

Main challenges of MSA forecasting include accumulated errors and increased uncertainty in the predictions, contrary to SSA point forecasting [11]. To achieve the desired accuracy, different MSA approaches need to be contemplated in addition to a robust forecasting framework. In this regard, Taieb *et al.* presented a comparative study of various MSA approaches namely direct, recursive, multi-input multi-output (MIMO), and direct MIMO (DIRMO) [12]. Among these approaches, DIRMO has a unique advantage in retaining model flexibility and avoiding accumulated errors, which will be an important factor for deep learning (DL)-based MSA forecasting techniques. Similarly, Wang *et al.* [13] investigated aforementioned MSA approaches to make wind speed predictions over longer time horizons and observed better accuracy with the DIRMO approach.

Furthermore, direct, MIMO, and recursive approaches are inspected in [14] using SVR, where MIMO-SVR is found to have the least mean absolute percentage error (MAPE) despite a computational load trade-off. In a similar manner, [15] inspected a delay-embedding-based ML scheme for solar irradiance and wind speed forecasting for 100 and 150 time-steps ahead, respectively. However, the presented scheme relies on the availability of multi-dimensional spatial-temporal data which may not always be accessible. In this context, DL methods can manage the applications with data insufficiency more effectively [16].

Notably, artificial neural networks (ANN)-based DL techniques such as recurrent neural networks (RNN) have set a performance benchmark for energy forecasting applications, as they can deal with temporal and nonlinear features of energy datasets very efficiently over different time horizons [17]. In this context, Ahmed *et al.* presented a six-steps (half hourly with an interval of 5 min.) ahead wind forecasting scheme using autoregressive neural networks using direct and recursive methodologies [18], however the scheme could not be extended to forecasting horizons greater than six-time steps. Recently, authors in [19] have compared various DL methods such as RNN, long-short term memory (LSTM), gated recurrent unit (GRU), and convolutional neural networks (CNN) for solar photovoltaic (PV) power forecasting in both SSA and MSA manner and claimed LSTM to be best performing DL model amongst all. Furthermore, BiLSTM as a bidirectional component of LSTM serves as a more effective model due to its ability of back-and-forth learning [20]. In this regard, Gairaa *et al.* also utilized multiple linear regression (MLR) and DL-based MSA approach for six-hours ahead solar irradiance prediction but the accuracy has not been investigated for sites with high variability [21].

Although LSTM and BiLSTM provide superior perfor-

D. Kaur, S. N. Islam and M. A. Mahmud are with School of Engineering, Deakin University, Australia. emails: devinderkaur@deakin.edu.au, shama.i@deakin.edu.au, apel.mahmud@deakin.edu.au.

A shorter version of this paper has been presented in STPEC 2021.

mance for both SSA and MSA predictions [22], [23] as compared to other state-of-the-art ML methods, the accuracy for MSA-DL frameworks can be further improved and the consideration of longer time horizons also remains an open challenge [11]–[13]. In addition, standard DL-based forecasting techniques are deterministic in nature and suffer from uncertainty in model parameters (weights) and prediction outputs [24], which become significant when multiple time-steps are considered [25]. In this context, probabilistic approaches have been widely used as effective methods to manage uncertainties in the renewable generation data [26]. Thus, there is a scope for an improved DL algorithm incorporated with probabilistic methods for effective MSA generation forecasting, which can tackle the uncertainty challenges of RESs and weight parameters efficiently.

In this regard, Bayesian inference integrated with neural layers [27], specifically LSTM [28] and BiLSTM [20] has emerged as a potential solution to deal with the problem of weight and prediction uncertainty in the energy forecasting domain. Contrary to the standard DL models, these methods can generate prediction intervals (PI), which can provide lower and upper bounds for the future prediction values with a given probability. Our previous work [29] has considered Bayesian DL (BDL) for MSA predictions which can be highly useful for renewable energy forecasting applications such as solar generation forecasting. However, existing research on BDL forecasting relies on standard variational inference (VI) techniques based on Kullback-Leibler (KL) divergence. Standard BDL methods underestimate the variance of weight parameter distribution and exhibit poorer performance for datasets having outliers [30], which can be a critical factor when there is high variability in the renewable energy data and in addition, MSA forecasting is performed. As a result, the standard VI techniques used in [29] are not sufficient to ensure the required accuracy for MSA solar generation forecasting and an improved approach needs to be developed which can quantify uncertainty with more sharpness and calibration.

As a result, we make following contributions in this paper:

- An improved Bayesian probabilistic method based on alpha-beta divergence integrated with Bidirectional LSTM (BiLSTM) is developed to approximate posterior distribution of weight parameters in the presence of uncertainties. The *improved*-Bayesian BiLSTM can quantify uncertainties and provide probabilistic forecasts using PIs more accurately.
- The proposed Bayesian method is implemented to generate MSA forecasts using DIRMO approach for one day ahead solar generation, i.e., 48-time steps ahead (half hourly). Furthermore, the developed method has also been tested for different forecasting horizons and it is evident from the results that it can manage longer forecasting horizons more effectively.
- A comparative case study is presented using Ausgrid solar generation dataset to demonstrate the superior performance of the proposed MSA forecasting framework against the benchmark SSA and MSA forecasting methods.

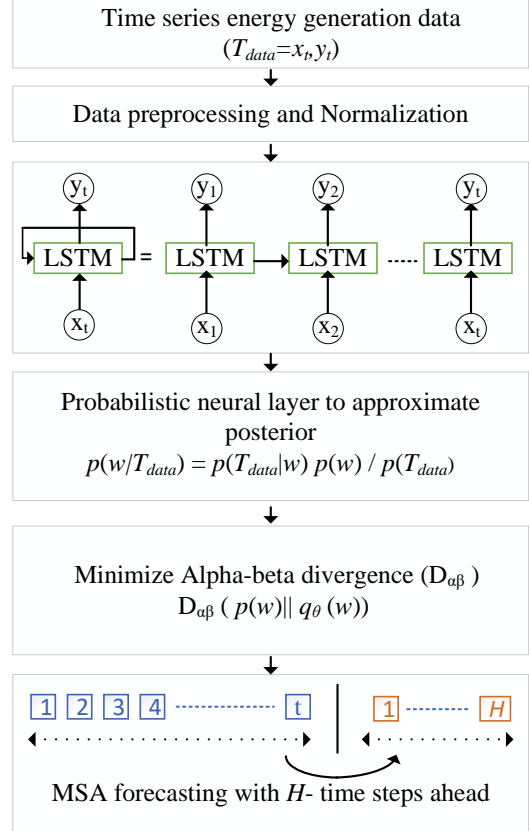


Fig. 1. Flowchart of the proposed *Improved*-Bayesian BiLSTM

The rest of the paper is organized in the following manner. Section II describes the problem formulation for optimizing the BDL model using BiLSTM in the presence of parameter uncertainties. Section III outlines the improved Bayesian BiLSTM approach integrated with direct MIMO (DIRMO) for MSA solar generation forecasting. Section IV highlights the implementation results. Finally, Section V concludes the paper.

## II. PRELIMINARIES AND PROBLEM FORMULATION

In this section, we discuss preliminaries on BiLSTM architecture and formulate the parameter optimization problem for BDL model considering uncertainties in model parameters and solar generation.

### A. Preliminaries on Bidirectional LSTM

This subsection briefly outlines the architecture of one LSTM cell employed in a bidirectional setting (BiLSTM) to process time series solar generation data with an intensive learning approach for improved MSA forecasting. Each LSTM cell involves four layers (cell state and three gates) in addition to RNN's recurrent loop structure. The three types of gates namely input ( $i_t$ ), output ( $o_t$ ), and forget ( $f_t$ ) gates enable the network to remove or add the information at  $t$  time stamp via cell state ( $c_t$ ) with the help of *sigmoid* function [31]. The equations for gating mechanism [32] are given as:

$$i_t = \sigma(w_i[y_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(w_f[y_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(w_o[y_{t-1}, x_t] + b_o) \quad (3)$$

where  $\sigma$  denotes the *sigmoid* function. Furthermore,  $w_i$ ,  $w_f$ , and  $w_o$  are the weight matrices and  $b_i$ ,  $b_o$ ,  $b_f$  denote the bias vectors corresponding to each gate. For a current input  $x_t$  the information at previous state ( $y_{t-1}$ ) is assessed by forget gate and is either discarded (using  $\sigma$ ) or updated (using  $\sigma$  and  $\tanh$ ) to the input gate using a candidate cell state ( $\tilde{c}_t$ ) as:

$$\tilde{c}_t = \tanh(w_c[y_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = (c_{t-1} * f_t) + (i_t * \tilde{c}_t) \quad (5)$$

where  $\tanh$  represents the hyperbolic tangent activation function. Note that,  $w_c$  and  $b_c$  refer to weight and bias corresponding to cell state. The filtered output ( $y_t$ ) for  $x_t$  is obtained using:

$$y_t = o_t * \tanh(c_t) \quad (6)$$

To train LSTM network, back-propagation through time (BPTT) is used as a training algorithm. The error derivatives are computed over weight parameters as a difference between predicted and estimated values. The total error ( $\epsilon$ ) for total number of samples ( $tot$ ) is calculated as:

$$\frac{d\epsilon}{dw} = \sum_{t=1}^{tot} \frac{d\epsilon_t}{dw} \quad (7)$$

The  $\epsilon$  needs to be minimized as:

$$\arg \min_{\phi} (\epsilon) \quad (8)$$

where  $\phi$  represents the hyperparameters such as learning rate, dropout rate, train-test and validation split ratio, number of epochs, etc. to tune model parameters in a neural network.

### B. Problem formulation for Bayesian deep learning models

The proposed probabilistic scheme utilizes the concept of Bayesian neural networks (BNN). The idea is to learn the distributions of weight parameters instead of deterministically learning specific weights (and bias) in a deep neural network [33]. The weight distributions are further used to sample the targeted output from the input data [34]. For a time series power generation dataset  $T_{data} = (x_t, y_t)$ , the posterior distribution of weight parameters in a BNN [35] is defined as:

$$p(w|T_{data}) = \frac{p(T_{data}|w)p(w)}{p(T_{data})} \quad (9)$$

where  $y_t$  represents the actual solar generation data for the  $t^{th}$  time interval. Further,  $p(w|T_{data})$  is known as the posterior distribution function of weight parameters ( $w$ ) and  $p(w)$  is the prior distribution symbolizing the assumed probability of  $w$  prior to any evidence is obtained. Using evidence  $T_{data}$ , prior is updated by multiplying likelihood  $p(T_{data}|w)$  with  $p(w)$  as given in the numerator of (9).

Note that computing posterior is the key element to obtain a full distribution for  $w$  and it often represents the epistemic or

weight uncertainty in Bayesian neural layers [36]. To compute full posterior, the product of likelihood and priors is evaluated over all parameters and further normalized using the technique of marginalization. The probability of evidence ( $p(T_{data})$ ) acts as a normalization constant used for marginalization and is computed using integration on each parameter setting as:

$$p(T_{data}) = \int p(T_{data}|w)p(w)dw \quad (10)$$

Considering the large sample space generated using (10), computing the exact posterior analytically becomes impossible. To solve this, variational inference (VI) is utilized to obtain an approximated posterior distribution in the modern Bayesian deep learning domain [37]. Thus, the network is trained using VI to learn the parameters of posterior distribution and optimized using negative log likelihood (NLL) cost function. Generally, the deviation between actual and approximated posterior is determined using divergence techniques [38] such as KL divergence as:

$$D(p||q) = \log p(T_{data}) - \mathbb{E}_{q_{\theta}(w)} \left[ \log \frac{p(T_{data}, w)}{q_{\theta}(w)} \right] \quad (11)$$

where  $q_{\theta}(w)$  denotes the approximated posterior over  $\theta$  parameters using VI, signifying the mean and variance of the approximated posterior distribution. Therefore, the objective function for the Bayesian deep learning model aims to minimize  $D_{KL}$  between approximated and actual posterior distribution as shown below:

$$\arg \min_{q_{\theta}(w)} D(p||q) \quad (12)$$

## III. PROPOSED MSA SOLAR GENERATION FORECASTING APPROACH

In this section, we propose an improved Bayesian deep learning approach for MSA solar generation forecasting. Fig. 1 represents a conceptual flow chart of the proposed *Improved-Bayesian BiLSTM* method for MSA forecasting. The proposed scheme utilizes Bayesian BiLSTM model integrated with an improved divergence technique and applies DIRMOMethod for MSA forecasting. MSA forecasting involves predicting a sequence of future half-hourly interval solar generation values in a horizon  $H$  for a given time series with historical solar generation data. A time series  $X_{data}$  for historical solar generation values at half-hourly intervals can be presented as:

$$X_{data} = [x_1, x_2, x_3, \dots, x_t] \quad (13)$$

where  $x_t$  is the solar generation at the  $t^{th}$  interval. The SSA predictions (next half-hour solar generation) at the  $(t+1)^{th}$  time interval is defined as:

$$\hat{y}_{t+1} = f(x_1, x_2, x_3, \dots, x_t) \quad (14)$$

Similarly,

$$\hat{y}_{t+2} = f(\hat{y}_{t+1}) \quad (15)$$

and so on. A segment  $X_{data(t-k)}^t$  from the training dataset  $X_{data}$ , representing solar generation data from  $(t-k)^{th}$  interval to  $t^{th}$  interval for  $k < t$ , can be defined as:

$$X_{data(t-k)}^t = [x_{t-k}, x_{t-k+1}, x_{t-k+2}, \dots, x_t] \quad (16)$$

Here  $k$  represents the number of past observations of solar generation which are used to generate the MSA forecasts. The function to obtain MSA forecasts for  $h$ -time steps ahead is formulated as:

$$[\hat{y}_{t+1}^{t+h}] = f(\hat{y}_t^{t+h-1}, X_{data(t-k)}^t) \forall h \in [1, \dots, H] \quad (17)$$

where  $\hat{y}_{t+1}^{t+h}$  ( $\hat{y}_t^{t+h-1}$ ) are the solar generation output forecasts for the time intervals  $(t+1)$  to  $(t+h)$  ( $t$  to  $t+h-1$ ). The proposed MSA probabilistic deep learning forecasting provides the distribution of multiple forecasts over a prediction horizon. So, the  $T_{data}$  in (9) becomes:

$$p(T_{data}) = p(y_{t+1}, y_{t+2}, \dots, y_{t+h} | x_1, x_2, \dots, x_t) \quad (18)$$

Since the proposed methodology incorporates BiLSTM for  $h$ -time steps ahead forecasting, the objective function in (8) becomes:

$$\arg \min_{\phi} (\epsilon_{t+1}^{t+h}) \quad (19)$$

where

$$\epsilon_{t+1}^{t+h} = \sum_{t+1}^{t+h} \epsilon \quad (20)$$

To manage the uncertainties associated with optimizing the model parameters of the aforementioned MSA forecasting problem, a Bayesian deep learning approach is integrated with the BiLSTM model. In this approach, the prior probability distribution of the weight parameters is modeled as a standard normal distribution  $\mathcal{N}(0, \sigma_i^2)$  with zero mean and standard deviation as  $\sigma_i$ . On the other hand, the posterior distribution is modeled by drawing random samples from  $\mathcal{N}(0, \mathbf{I})$  with zero mean and unity variance, shifting the samples by  $\mu$  and then scaling by  $\sigma_p$ . Here  $\mu$  and  $\sigma_p$  represent the mean and standard deviation of the approximated posterior distribution of the weight parameters.  $\sigma_p$  is modeled as a softplus function. Then the difference between true and approximated posterior is minimized using divergence techniques.

Standard BDL methods use KL divergence to optimize the posterior distribution of the weight parameters. However, KL divergence faces the challenge of underestimating the posterior variance [39], especially when the solar generation data has high variability. As a result, a more robust divergence technique which can manage the variability appropriately is required for optimizing the posterior distribution effectively. To address this issue, the proposed method integrates an improved divergence technique, namely, alpha-beta divergence ( $D_{\alpha\beta}$ ) as a more scalable alternative of KL divergence to approximate posterior using parameters  $\alpha$  and  $\beta$ . It gives more freedom for parameter selection according to the probability density and can be derived as:

$$D_{\alpha\beta}(p||q) = \left( \frac{\alpha + \beta - 1}{\beta(\alpha + \beta)} - \frac{\alpha + \beta - 1}{\alpha\beta} - \frac{1}{\alpha(\alpha + \beta)} + \frac{1}{\alpha} \right) \times \mathbb{E}[\log(\theta)] \quad (21)$$

The detailed derivation for  $D_{\alpha\beta}$  is illustrated in the Appendix through (28)-(31). Thus, the weight parameter optimization problem transforms to:

$$\arg \min_{q_{\theta}(w)} D_{\alpha\beta}(p||q) \quad (22)$$

---

**Algorithm 1** An *Improved-Bayesian BiLSTM* technique for MSA solar generation forecasting

---

**Input:**  $T_{data} = (x_t, y_t)$

**Output:**  $\hat{y}_t, \dots, \hat{y}_{t+h}, \hat{y}_{t,\tau}, \hat{y}_{t+h,\tau}$  for the  $\tau^{th}$  percentile

**Hyperparameters:**  $\phi$

**Parameters:**  $w$

**Performance metrics:**  $Q_l, W_l, RMSE, MAE$

---

- 1: Obtain  $T_{data}$  from half-hourly solar generation dataset;
  - 2: Scale and normalize  $T_{data}$ ;
  - 3:  $x_t \implies$  explanatory variable (2D);
  - 4:  $y_t \implies$  observed variable (2D);
  - 5: Convert  $x_t \rightarrow$  3D;
  - 6: Split  $T_{data}$  into training and testing subsets;
  - 7: Follow steps 9-18 to train Bayesian inference;
  - 8: **for** ( $ep=1; ep \leq N; ep++$ ) **do**
  - 9:  $p(w) \leftarrow \mathcal{N}(0, \sigma_i^2)$ ;
  - 10:  $q_{\theta}(w) \leftarrow \sigma_p \odot \mathcal{N}(0, \mathbf{I}) + \mu$ ;
  - 11: Compute  $D_{\alpha\beta}(p||q)$  using (21);
  - 12: Minimize  $D_{\alpha\beta}$  using (22);
  - 13: Input  $x_t$  from training subset to BiLSTM layer;
  - 14: Calculate  $i_t, f_t$ , and  $\tilde{c}_t$  using (1), (2) and (4), respectively;
  - 15: Update cell state using (5);
  - 16: Update  $o_t$  using (3);
  - 17: Update output ( $y_t$ ) of LSTM block using (6);
  - 18: Pass the inputs in forward and backward directions and update  $w$ ;
  - 19: **end for**
  - 20: Obtain the distribution of  $t^{th}$  to  $(t+h)^{th}$  predictions as output using (17);
  - 21: Compute the prediction intervals based on  $\hat{y}_{t,\tau}$ ;
  - 22: Draw samples from the distributions of the forecasting model outputs;
  - 23: Compute the predictive mean of the generation forecast outputs;
  - 24: To evaluate probabilistic forecasting performance, follow steps 25 and 26 on test subset;
  - 25: Compute  $Q_l$  using (23);
  - 26: Compute  $W_l$  using (24);
  - 27: To evaluate deterministic forecasting performance, follow steps 28 and 29 on test subset;
  - 28: Compute  $RMSE$  using (25);
  - 29: Compute  $MAE$  using (26);
- 

Furthermore, Algorithm 1 presents the pseudo-code for the proposed MSA Bayesian BiLSTM forecasting approach. As illustrated, the obtained time series data ( $T_{data}$ ) is preprocessed and normalized before dividing it into the training and testing subsets (lines 1-6). In addition, the two-dimensional (2D) explanatory variable ( $x_t$ ) is converted into three-dimensional (3D) array representing historical generation data as time lags (samples, time-lags, features) (line 5). This 3D data from training set is then fed to the *improved-Bayesian BiLSTM* deep neural network to generate multi-time steps ahead predictions via output layer. The proposed model is trained for

TABLE I  
COMPARATIVE ANALYSIS OF PROPOSED METHOD FOR MSA SOLAR GENERATION FORECASTING FOR  $(t + 48)$

Sr. no.	Method	RMSE (kWh)	MAE (kWh)	Pinball (avg) (kWh)	Winkler (kWh)	Time (s)
1	Standard BiLSTM (point forecasting)	0.2634	0.1485	-	-	435.83
2	Standard LSTM (point forecasting)	0.4041	0.2109	-	-	222.04
3	Standard RNN (point forecasting)	0.5621	0.4129	-	-	139.67
4	Bayesian RNN [27]	0.1422	0.1113	0.3431	3.1330	323.39
5	Bayesian LSTM [28]	0.1011	0.0535	0.2693	2.9820	204.56
6	Bayesian BiLSTM [20]	0.0975	0.0486	0.2806	1.7400	412.33
7	<i>Improved</i> -Bayesian BiLSTM	0.0860	0.0387	0.0161	0.0143	174.37

$N$  iterations called epochs (lines 9-18). From probabilistic perspective, for each iteration, the posterior distribution is approximated for  $w$  using variational inference by optimizing  $D_{\alpha\beta}$  between prior and posterior distribution (lines 11-12).

The training data is passed forward and backward in multiple slices through the network. The network adjusts weights based on the Bayesian inference. The model is trained until the value of the cost function cannot be further reduced. After achieving a global minima, the predictions for  $t^{th}$  to  $(t + h)^{th}$  intervals are computed (line 20). Note that  $t + h$  number of neurons are passed in the output layer to generate multiple outputs at each step contributing towards DIRM approach. Then, predictive mean of the forecast outputs for all time intervals within horizon  $H$  are computed (line 22). Moreover, prediction intervals (PIs) are generated using  $\hat{y}_{t,\tau}$  for desired percentiles ( $\tau$ ) values. Furthermore, to evaluate the forecasting performance in deterministic and probabilistic format, different error values are computed (lines 23-28).

#### IV. RESULTS AND DISCUSSIONS

This section includes the implementation results for the proposed technique carried out on solar generation data (in kWh) from 300 individual rooftop PV panels in Australia with 30 minutes interval available from the Ausgrid dataset [40]. For training purposes, first 9 months of generation data from house number 2076 for the year 2011 – 2012 is utilized with a validation split of 0.2 and optimizer *Adam*. The trained model is then tested on remaining 3 months of energy generation data by making 48-time steps ahead predictions. The past one day (48 intervals) of generation data is passed to the network as historical input to make the algorithm learn. The results are obtained using python's open source library Tensorflow, 16 GB RAM, graphics processing unit (GPU), and an *i7* processor.

##### A. Evaluation metrics

The proposed scheme is evaluated on both deterministic and probabilistic error metrics. For Bayesian probabilistic performance, Pinball loss ( $Q_i$ ) and Winkler score ( $w_i$ ) are computed to evaluate the accuracy and sharpness of forecasted quantiles [22] as:

$$Q_i(y_t, \hat{y}_{t,\tau}, \tau) = \begin{cases} \tau(y_t - \hat{y}_{t,\tau}) & y_t \geq \hat{y}_{t,\tau} \\ (1 - \tau)(\hat{y}_{t,\tau} - y_t) & y_t \leq \hat{y}_{t,\tau} \end{cases} \quad (23)$$

where  $\hat{y}_{t,\tau}$  symbolizes  $t^{th}$  forecast for percentile value  $\tau$  against actual solar generation value  $y_t$ . Furthermore, for confidence level  $(1 - \gamma) \times 100$ , Winkler score is computed as:

$$W_i = \begin{cases} \delta & lb_t \leq y_t \leq ub_t \\ \delta + 2(lb_t - y_t)/\gamma & lb_t \geq y_t \\ \delta + 2(y_t - ub_t)/\gamma & ub_t \leq y_t \end{cases} \quad (24)$$

where  $lb_t$  and  $ub_t$  represent the lower and upper bounds of probabilistic forecasts at  $t$ , respectively. And,  $\delta = ub_t - lb_t$  is defined as the PI width. A lower Pinball loss and Winkler score imply better probabilistic estimation.

Furthermore, forecasting accuracy at deterministic level is evaluated by computing root-mean square error (RMSE) and mean absolute error (MAE) of the differences between actual generation values and predictive mean as given below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (25)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (26)$$

where  $n$  denotes the total number of samples in the test data.

##### B. Comparative analysis

To evaluate the forecasting performance, the proposed Bayesian scheme is compared with other variants of standard BDL and benchmark point forecasting methods including RNN, unidirectional LSTM, and BiLSTM architectures. The numerical results for the 48-times steps ahead solar generation forecasting from the proposed and comparative methods are demonstrated in Table I using deterministic and probabilistic evaluation metrics.

While point forecasting methods exhibit poor performance for MSA forecasts by demonstrating higher RMSE and MAE values, Bayesian methods with recurrent neural layers, namely Bayesian-RNN, Bayesian-LSTM, and Bayesian-BiLSTM manage to achieve significant decrease in the afore-said deterministic errors with a competent time complexity. The proposed *Improved*-Bayesian BiLSTM method demonstrates the lowest RMSE and MAE values compared to standard point forecasting and BDL methods. In addition, standard point forecasting methods are not capable to quantify uncertainty in deep learning models and cannot provide probabilistic

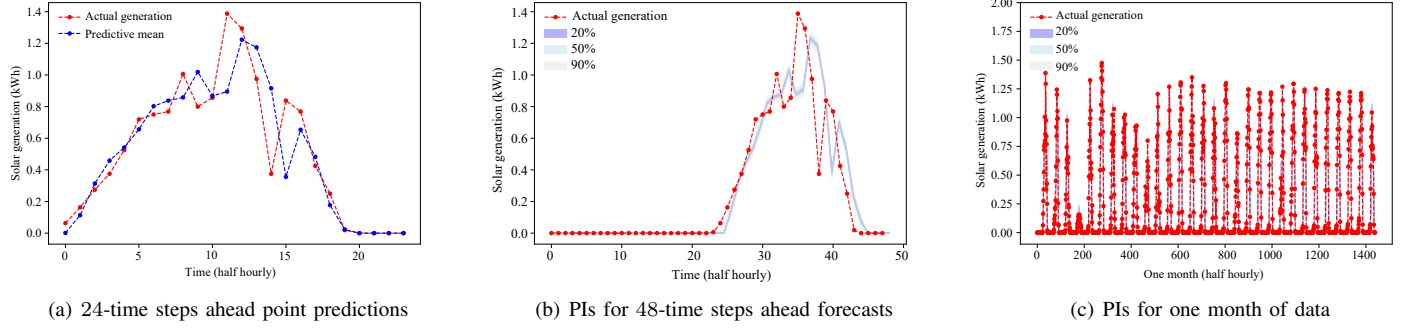


Fig. 2. Performance analysis of the proposed *Improved*-Bayesian BiLSTM for solar generation forecasting

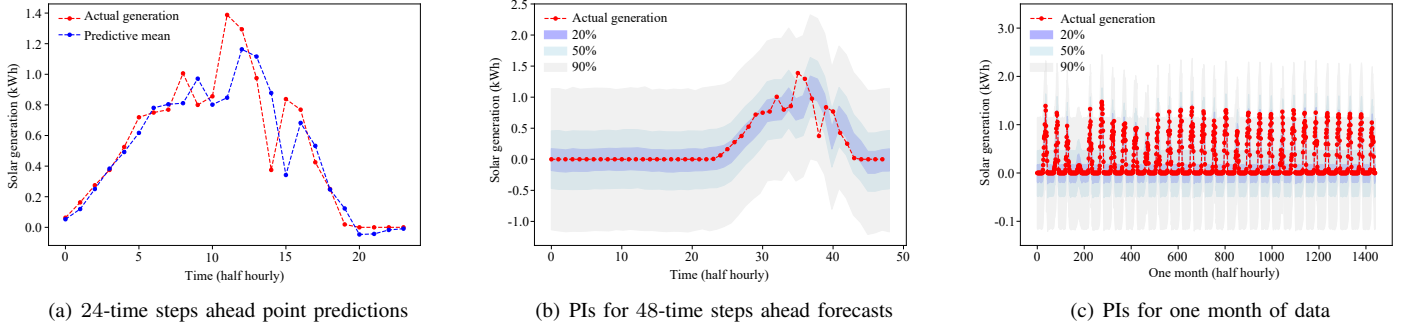


Fig. 3. Performance analysis of the standard Bayesian BiLSTM for solar generation forecasting

forecasts, thus Pinball and Winkler scores are not reported for the first three comparative methods in Table I.

From the probabilistic metrics, it is evident that the proposed *Improved*-Bayesian BiLSTM achieves least Pinball and Winkler scores (in kWh), i.e., 0.0161 and 0.0143, respectively which is more than 10 fold improvement compared to Bayesian-RNN, Bayesian-LSTM, and Bayesian-BiLSTM. This improvement is entailed by alpha-beta divergence employed in the proposed framework, which can manage uncertainties in the solar generation data and optimize deep learning model parameters more effectively. On the other hand, standard Bayesian BiLSTM uses KL divergence which underestimates the variance of the approximated posterior. Also, it is important to note that BiLSTM with point and probabilistic approach takes additional execution time due to bidirectional processing involved in the learning process. However, the *Improved*-Bayesian BiLSTM requires less computation time compared to standard Bayesian BiLSTM as the solutions can converge effectively when the weight parameters are optimized using a more accurate divergence technique.

Furthermore, Fig. 2 and Fig. 3 represent graphical illustrations for solar generation forecasts using *Improved* and standard Bayesian BiLSTM, respectively. Figs. 2(a) and 3(a) reflect similar point forecasting patterns between actual and predictive mean with a time lag during the peak irradiance hours. On the other hand, Fig. 2(b) and Fig. 3(b) highlight the probabilistic forecasts for next day ahead (48-time steps ahead) solar generation with 20%, 50%, and 90% PIs using *Improved*-Bayesian and standard Bayesian BiLSTM method, respectively. Given that the *Improved*-Bayesian BiLSTM

estimates a more accurate approximation of the posterior distribution of weight parameters, Fig. 2(b) demonstrates that the proposed method improves the future generation probabilities for different percentiles, by providing tighter PI bounds. Furthermore, Fig. 2(c) and Fig. 3(c) provide MSA probabilistic forecasts on one month of data using both methods ensuring the capability of proposed method to quantify uncertainty more precisely for longer horizons.

TABLE II  
MSA PROBABILISTIC FORECASTING RESULTS FOR  $(t + 24)$

Method	RMSE	MAE	Pinball	Winkler
Bayesian BiLSTM	0.0963	0.0449	0.2770	1.7050
<i>Improved</i> -Bayesian BiLSTM	0.0854	0.0370	0.0148	0.0123

TABLE III  
MSA PROBABILISTIC FORECASTING RESULTS FOR  $(t + 12)$

Method	RMSE	MAE	Pinball	Winkler
Bayesian BiLSTM	0.0954	0.0475	0.2653	1.6170
<i>Improved</i> -Bayesian BiLSTM	0.0851	0.0357	0.0138	0.0121

We further extend the experiments to three more variations involving 24-, 12-, and 1-time steps ahead generation forecasts conducted on standard Bayesian BiLSTM and the proposed *Improved*-Bayesian BiLSTM. Tables II, III, and IV illustrate numerical results for respective variations using deterministic and probabilistic forecasting evaluation metrics. It can be observed that RMSE and MAE values for the proposed method

TABLE IV  
SSA PROBABILISTIC FORECASTING RESULTS FOR  $(t + 1)$

Method	RMSE	MAE	Pinball	Winkler
Bayesian BiLSTM	0.0928	0.0469	0.1446	1.4170
<i>Improved</i> -Bayesian BiLSTM	0.0850	0.0350	0.0127	0.0115

are lower than those of the Bayesian BiLSTM method from all three cases. In addition, *improved* technique achieves a more than 10-fold improvement in the Pinball and Winkler scores for all three variations of MSA solar generation forecasting. Also, it is worthwhile to observe that with the decrease in forecasting time horizons, i.e., from 48 to SSA (30 minutes ahead), the forecasting accuracy increases in terms of lower error values. To be specific, there is a 17% (50%) degradation in the Winkler score (Pinball loss) for the standard Bayesian BiLSTM method when forecasts are generated 48-time steps ahead rather than 1-time step ahead. On the other hand, the *improved* method suffers a degradation of only 6.5% (14%) in the Winkler score (Pinball loss). Thus, the proposed method is more suitable for generating MSA forecasts as compared to standard Bayesian BiLSTM.

TABLE V  
IMPACT OF ALPHA-BETA PARAMETERS ON *Improved*-BAYESIAN BiLSTM  $(t + 48)$

$\alpha$	$\beta$	Pinball	Winkler
1.0	2.0	0.0161	0.0143
2.0	3.0	0.0190	0.0210
3.0	4.0	0.0224	0.0514
2.0	6.0	0.0422	0.2085
3.0	8.0	0.0638	0.2663
5.0	15.0	0.1143	0.2805

Table V represents the impact of different  $\alpha$  and  $\beta$  parameters on Pinball and Winkler scores for 48-time steps ahead solar generation forecasting. It can be observed from the experiments that lower values of alpha and beta achieve lower errors, and thus provide better forecasting performance.

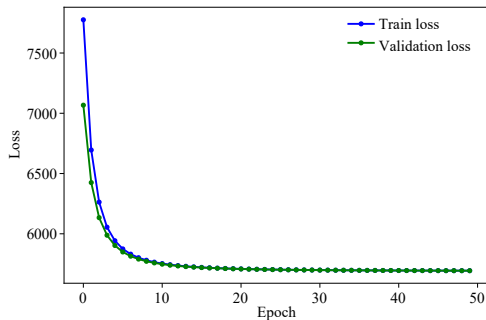


Fig. 4. Loss reduction using standard Bayesian BiLSTM

Furthermore, Fig. 4 and Fig. 5 illustrate the convergence of likelihood loss against number of epochs validated with 0.2 validation split during the network training phase. The aforementioned figures demonstrate effective learning process via Bayesian technique without overfitting or underfit-

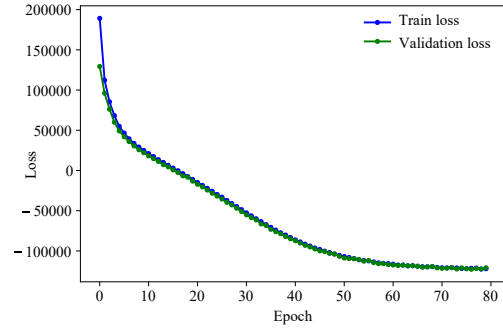


Fig. 5. Loss reduction using *Improved*-Bayesian BiLSTM

ting the training data. However, the proposed *Improved*-Bayesian BiLSTM achieves lower loss as compared to standard Bayesian method as number of epochs increase. This is because the loss function employed in the learning process is negative log likelihood which is more accurately treated by the alpha-beta divergence technique.

## V. CONCLUSION

In this paper, an improved MSA forecasting approach is presented using Bayesian probability integrated with alpha-beta divergence and BiLSTM deep learning method to predict future energy generation in smart grid systems. The proposed method is evaluated using solar generation data and forecasting results demonstrate the efficacy of point and probabilistic predictions in the terms of lower error values. Furthermore, impact of different time horizons on the forecasting performance of the proposed method has also been investigated. The proposed method outperforms standard Bayesian techniques for both SSA and MSA forecasting by demonstrating least prediction errors with less computational time. Future work will focus on exploring Bayesian BiLSTM for multidimensional energy forecasting problems.

## APPENDIX

This appendix presents the derivation for alpha-beta divergence ( $D_{\alpha\beta}$ ) given in the Eq. (21) as an extension to the following scale-invariant alpha-beta divergence from [30]:

$$\begin{aligned}
 D_{\alpha\beta}(q||p) &= \frac{1}{\alpha(\alpha + \beta)} \log \mathbb{E} \left[ \frac{p(\theta, w)^{\alpha+\beta}}{q(\theta)} \right] \\
 &+ \frac{1}{\beta(\alpha + \beta)} \log \mathbb{E} [q(\theta)^{\alpha+\beta-1}] \\
 &- \frac{1}{\alpha\beta} \log \mathbb{E} [q(\theta)^{\alpha+\beta-1} \left( \frac{p(\theta, w)}{q(\theta)} \right)^\beta] \quad (27)
 \end{aligned}$$

The aforementioned expression can be written as:

$$D_{\alpha\beta}(q||p) \geq \frac{1}{\alpha(\alpha+\beta)} \mathbb{E}[\log \frac{p(\theta, w)^{\alpha+\beta}}{q(\theta)}] + \frac{1}{\beta(\alpha+\beta)} \mathbb{E}[\log[q(\theta)^{\alpha+\beta-1}]] - \frac{1}{\alpha\beta} \mathbb{E}[\log q(\theta)^{\alpha+\beta-1} \left( \frac{p(\theta, w)}{q(\theta)} \right)^\beta] \quad (28)$$

$$= \frac{1}{\alpha(\alpha+\beta)} \mathbb{E}[\log p(\theta, w)^{\alpha+\beta} - \log q(\theta)] + \frac{1}{\beta(\alpha+\beta)} * (\alpha + \beta - 1) * \mathbb{E}[\log q(\theta)] - \frac{1}{\alpha\beta} \mathbb{E}[\log q(\theta)^{\alpha+\beta-1}] - \frac{1}{\alpha\beta} \log \left( \frac{p(\theta, w)}{q(\theta)} \right)^\beta \quad (29)$$

$$= \frac{1}{\alpha} \mathbb{E}[\log p(\theta, w)] - \frac{1}{\alpha(\alpha+\beta)} \log[q(\theta)] + \frac{\alpha+\beta-1}{\beta(\alpha+\beta)} \mathbb{E}[\log q(\theta)] - \frac{\alpha+\beta-1}{\alpha\beta} \mathbb{E}[\log q(\theta)] - \frac{1}{\alpha} \mathbb{E}[\log p(\theta, w)] + \frac{1}{\alpha} \log q(\theta) \quad (30)$$

$$= \left( \frac{\alpha+\beta-1}{\beta(\alpha+\beta)} - \frac{\alpha+\beta-1}{\alpha\beta} - \frac{1}{\alpha(\alpha+\beta)} + \frac{1}{\alpha} \right) \times \mathbb{E}[\log q(\theta)] \quad (31)$$

where (28) follows Jensen's inequality and (29)-(30) are derived using logarithmic operations. Finally, (31) is written after some mathematical manipulations.

## REFERENCES

- [1] [Online]. Available: <https://www.energy.gov.au/households/solar-pv-and-batteries>
- [2] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2018.
- [3] D. Kaur, S. N. Islam, M. Mahmud, Z. Dong *et al.*, "Energy forecasting in smart grid systems: A review of the state-of-the-art techniques," *arXiv preprint arXiv:2011.12598*, 2020.
- [4] J. W. Messner and P. Pinson, "Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1485–1498, 2019.
- [5] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 549–556, 2013.
- [6] R. Saeedi, S. K. Sadanandan, A. Srivastava, K. Davies, and A. Gebremedhin, "An adaptive machine learning framework for behind-the-meter load/pv disaggregation," *IEEE Transactions on Industrial Informatics*, 2021.
- [7] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.
- [8] A. G. Parlos, O. T. Rais, and A. F. Atiya, "Multi-step-ahead prediction using dynamic recurrent neural networks," *Neural networks*, vol. 13, no. 7, pp. 765–786, 2000.
- [9] H. Cheng, P.-N. Tan, J. Gao, and J. Scripps, "Multistep-ahead time series prediction," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2006, pp. 765–774.
- [10] G. Bontempi, "A monte carlo strategy for structured multiple-step-ahead time series prediction," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 853–858.
- [11] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, "A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition," *Expert systems with applications*, vol. 39, no. 8, pp. 7067–7083, 2012.
- [12] S. B. Taieb, G. Bontempi, A. Sorjamaa, and A. Lendasse, "Long-term prediction of time series by combining direct and mimo strategies," in *2009 International Joint Conference on Neural Networks*. IEEE, 2009, pp. 3054–3061.
- [13] J. Wang, Y. Song, F. Liu, and R. Hou, "Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 960–981, 2016.
- [14] Y. Bao, T. Xiong, and Z. Hu, "Multi-step-ahead time series prediction using multiple-output support vector regression," *Neurocomputing*, vol. 129, pp. 482–493, 2014.
- [15] H. Peng, P. Chen, and R. Liu, "Multi-step-ahead prediction from short-term data by delay-embedding-based forecast machine," *arXiv preprint arXiv:2005.07842*, 2020.
- [16] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting: novel pooling deep rnn," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2017.
- [17] D. Kaur, R. Kumar, N. Kumar, and M. Guizani, "Smart grid energy management using rnn-lstm: A deep learning-based approach," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.
- [18] A. Ahmed and M. Khalid, "Multi-step ahead wind forecasting using nonlinear autoregressive neural networks," *Energy Procedia*, vol. 134, pp. 192–204, 2017.
- [19] A. Mellit, A. M. Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renewable Energy*, vol. 172, pp. 276–288, 2021.
- [20] D. Kaur, S. N. Islam, M. Mahmud *et al.*, "A vae-based bayesian bidirectional lstm for renewable energy forecasting," *arXiv preprint arXiv:2103.12969*, 2021.
- [21] K. Gairaa, C. Voyant, G. Notton, S. Benkaciali, and M. Guermoui, "Contribution of ordinal variables to short-term global solar irradiation forecasting for sites with low variabilities," *Renewable Energy*, 2021.
- [22] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1203–1215, 2018.
- [23] H. Jahangir, H. Tayarani, S. S. Gougheri, M. A. Golkar, A. Ahmadian, and A. Elkamel, "Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional lstm network," *IEEE Transactions on Industrial Electronics*, 2020.
- [24] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural network," in *International Conference on Machine Learning*. PMLR, 2015, pp. 1613–1622.
- [25] Y. Gal, "Uncertainty in deep learning," *University of Cambridge*, vol. 1, no. 3, p. 4, 2016.
- [26] G. E. Constante-Flores and M. S. Illindala, "Data-driven probabilistic power flow analysis for a distribution system with renewable energy sources using monte carlo simulation," *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 174–181, 2019.
- [27] Y. Yang, W. Li, T. A. Gulliver, and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4703–4713, 2019.
- [28] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188–201, 2019.
- [29] D. Kaur, S. N. Islam, and M. A. Mahmud, "A bayesian probabilistic technique for multi-step ahead renewable generation forecasting," in *STPEC*. IEEE, 2021.
- [30] J.-B. Regli and R. Silva, "Alpha-beta divergence for variational inference," *arXiv preprint arXiv:1805.01045*, 2018.
- [31] [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [33] X. Zhang, F. Fang, and J. Wang, "Probabilistic solar irradiation forecasting based on variational bayesian inference with secure federated learning," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2020.
- [34] P. Izmailov, S. Vikram, M. D. Hoffman, and A. G. Wilson, "What are bayesian neural network posteriors really like?" *arXiv preprint arXiv:2104.14421*, 2021.
- [35] Y. Ho and R. Lee, "A bayesian approach to problems in stochastic estimation and control," *IEEE Transactions on Automatic Control*, vol. 9, no. 4, pp. 333–339, 1964.
- [36] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," *arXiv preprint arXiv:2002.08791*, 2020.
- [37] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.



- [38] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, “Advances in variational inference,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 2008–2026, 2018.
- [39] T. Minka *et al.*, “Divergence measures and message passing,” Technical report, Microsoft Research, Tech. Rep., 2005.
- [40] Ausgrid. [Online]. Available: <https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data>