*Environmental Data Science* (2022), 4: 1–9
doi:10.1017/eds.2022.xx

RESEARCH ARTICLE

# Week-ahead Solar Irradiance Forecasting with Deep Sequence Learning

Saumya Sinha[1*], Bri-Mathias Hodge[1,2] and Claire Monteleoni[1]

[1]University of Colorado, Boulder, CO, USA
[2]National Renewable Energy Laboratory (NREL), Boulder, CO, USA
*Corresponding author. Email: saumya.sinha@colorado.edu

**Received:** 28 January 2022; **Accepted:** 08 May 2022

**Keywords:** uncertainty estimation; deep sequence models; week-ahead forecasting; renewable energy

## Abstract

In order to enable widespread integration of solar energy into the power system, there is an increasing need to reduce the uncertainty associated with solar power output which requires major improvements in solar irradiance forecasting. While most recent works have addressed short-term (minutes or hours ahead) forecasting, through this work, we propose using deep sequence learning models for forecasting at longer lead-times such as a week in advance, as this can play a significant role in future power system storage applications. Along with point forecasts, we also produce uncertainty estimates through probabilistic prediction and showcase the potential of our machine learning frameworks for a new and important application of longer lead-time forecasting in this domain. Our study on the SURFRAD data over seven US cities compares various deep sequence models and the results are encouraging, demonstrating their superior performance against most benchmarks from the literature and a current machine learning based probabilistic prediction baseline (previously applied to short-term solar forecasting).

### Impact Statement

We show the promise of machine learning for longer-term solar forecasting with probabilistic predictions, an area that has not been sufficiently explored in the literature. Our encouraging results suggest such methods could play a larger role in future power system operations, when greater shares of renewable energy resources will require operational planning at these timescales. For example, these methods could inform the operation of hybrid power plants with storage capabilities, where information about expected future renewable power generation would weigh into decisions on storage charging and discharging.

## 1. Introduction

Renewable energy sources, like solar, wind, tidal, or geothermal energy, have the potential of reducing the world's dependency on fossil fuel. These resources are not only abundantly available in nature but they are also clean energy sources, reducing greenhouse gas emissions that lead to global warming. However, many of these resources are variable and uncertain, posing challenges for integration into a power system which is predicated upon dispatchable supply. There is therefore a growing need for accurate renewable energy forecasting to ease integration into electric grids. Solar photovoltaics (PV) systems are experiencing exponential growth in deployment and the output of PV systems is highly dependent on solar irradiance [1]. A number of physical and statistical models have been used for making solar forecasts at different timescales from intra-hour to a few days-ahead [19, 21]. Statistical methods have been shown to perform well at forecasting at very short time horizons, with numerical weather prediction models (NWP) outperforming them in the hours to days-ahead timeframe [19].

Most physical models in this domain are based on NWP simulations that traditionally provide more accurate forecasts at hours to days-ahead lead times [19]. However, due to their computational expense, NWP model outputs are updated less frequently and with coarser resolution at longer prediction lead times, such as week(s) ahead. This motivates the need for data-driven machine learning models that can provide forecasts at longer periods in advance at a finer (1 hour) resolution (as opposed to *e.g.*, the 12 hour resolution in the case of the European Centre for Medium-Range Weather Forecasts (ECMWF) model predictions). As a part of our study, we not only perform a direct comparison with the NWP baseline for our one-week ahead forecasts, but we also evaluate our models' performance when they incorporate NWP outputs as input features to see if it improves their forecasting ability.

Probabilistic forecasting provides a distribution over the prediction, this additional knowledge of uncertainty estimates can provide advantages over point forecasting. For example, knowing about future time periods of low and high uncertainty in advance can be very useful in planning plant maintenance [23]. Until recently, probabilistic forecasting for solar energy had not received as much attention as for wind energy, as observed in [7]. In their work, [7] introduces probabilistic benchmarks to evaluate probabilistic methods, which we will utilize in this work.

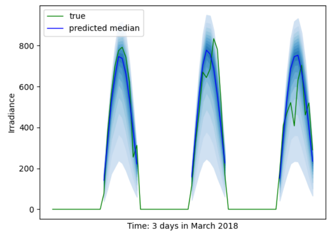



***Figure 1.*** *Fan plot showing the Temporal CNN (TCN) model's prediction intervals from 5% to 95% percentile on three March days at the Boulder station.*

**Contributions** We propose deep sequence learning for this longer lead-time (one week-ahead) solar irradiance forecasting task, that provides point as well as probabilistic predictions. Overall, these deep learning pipelines outperform several benchmarks from the literature including numerical weather prediction (NWP) models and a machine learning-based probabilistic prediction method. The results fall slightly behind the complete history persistence ensemble (Ch-PeEN) benchmark [7] in terms of continuous ranked probability score (CRPS), but are better in terms of forecast sharpness.

## 2. Related Work

A variety of deep learning approaches have been proposed for learning from sequence data, some of which have been applied in the solar energy domain. Recurrent neural networks (RNNs), unlike fully connected neural networks, have the ability to capture temporal dependencies in sequences by incorporating feedback from previous time steps. Long Short-Term Memory (LSTMs) models are especially useful for a time series data when the inputs can have longer dependencies. The works in [1, 14, 9, 6] show the potential of LSTMs for solar energy forecasting, and they outperform fully connected networks and traditional machine learning models at short forecasting lead times. Convolution neural network (CNN)-based models that use dilated and causal convolutions along with residual connections (also referred to as Temporal CNNs) were designed specifically for sequential modeling [4, 15]. They are autoregressive prediction models based on the recent WaveNet architecture [15]. Temporal CNNs have recently been applied to forecasting day-ahead PV power output, outperforming both LSTMs and multi-layer feed forward networks [13]. They are able to exploit a longer history in the time series, enabling more accurate forecasts. In this work, we study a significantly longer forecast horizon that challenges the limits of NWP forecasting and is expected to have emerging applications as power systems evolve. We compare LSTMs, Temporal CNNs, Temporal CNNs with an added attention layer [17], and the Transformer model [20].

Recently, [23] showed how probabilistic models such as Gaussian processes, neural networks with dropout for uncertainty estimation, and NGBoost [8] compare when making short-term solar forecasts. They explored *post hoc* calibration techniques for improving the forecasts produced by these models. NGBoost or Natural Gradient Boosting algorithm [8] is a gradient boosting pipeline that is extended to give a probabilistic distribution as an output (the parameters of the distribution are the regressed outputs). We now consider NGBoost with a Gaussian output distribution, to be a machine learning benchmark in this domain, since it showed superior performance for intra-hour and hourly resolution forecasting [23]. Deep learning-based probabilistic prediction models are, however, yet to be fully explored [21]. In this paper, we extend the deep learning point prediction models mentioned above to yield predictions at multiple quantiles (see Figure 1), as quantile regression is a non-parametric approach to obtain probabilistic forecasts [21, 17].

## 3. Data

We use open-source NOAA's SURFRAD network (Surface radiation budget network for atmospheric research) [2] that provides the ground-truth solar irradiance and meteorological measurements from seven sites across the US in different climatic zones (https://gml.noaa.gov/grad/surfrad/). Models are trained on measurements from the years 2016-2017, and then evaluated on the year 2018. The test data (year 2018) is kept hidden and the rest of the data is split into training and validation sets (70/30 split). Data is converted to an hourly resolution and only the day time values are considered for training and testing of all models including benchmarks (for relevance to the domain, as in [7]). Days with less than 24 hours of data points due to missing data were dropped. Following standard practice, we take a ratio of the ground-truth Global Horizontal Irradiance (GHI) (Watts/m$^2$) with respect to the "clear sky" GHI value (these are irradiance estimates under cloud-free conditions, obtained from CAMS McClear Service [11]), to produce a clearness index, such as in [23, 14, 7] that is used as the prediction label for training. While trained on the clearness index, the models are evaluated on the GHI.

Important predictor variables available in the data, such as solar zenith angle, hour of the day, month of the year, wind, pressure, temperature, and relative humidity, are included, along with the clearness index at the hour (a total of 16 input variables overall). These inputs are scaled (standardized) before the modeling procedure. All the sequence models take in a 3D input, where every row is a sequence of input feature vectors corresponding to previous timesteps: we use a history of $12 \times 7$ past daylight hours for all our models. Each row in the time series at hour $h$ is assigned a label which is the clearness index value at the hour $h$+*one-week*.

## 4. Methods

We focus on showing the potential of the following deep multi-variate sequence models: LSTM, Temporal CNN, Temporal CNN with an attention layer, and Transformer, for point and probabilistic solar irradiance forecasting. We compare them to the NgBoost method [8] that has been shown to outperform various probabilistic models for short-term solar forecasting [23], along with benchmarks from the literature (as described in the next section). Hyperparameters were tuned on the validation dataset.

**LSTM**: We use a simple LSTM pipeline; a single hidden layer with a dimension of 25.
**Temporal CNN (TCN)**: Temporal CNN consists of 1D dilated convolution filters and residual layers that are responsible for learning long term dependencies efficiently [4, 13]. Figure 2 shows how dilations help to increase (exponentially) the receptive field of a kernel. This makes the model capable of

learning correlations between data points far apart in the past. The convolutions are also causal, meaning that while convolving, outputs at time t only convolve with time t and earlier from the previous layer. Our TCN architecture is comprised of 3 levels, size of the hidden layer is 25 and kernel size is 3 with dilation factors d = 1,2,4.

**Temporal CNN with Attention**: The Attention mechanism [3] has been used for sequential modeling and time series prediction problems [16]. It has the ability to model dependencies in long sequences without regard to their distance [20]. We add a self-attention layer (adapted from [24]) on the convolution maps generated from the Temporal CNN network and observe the prediction outcomes. This enables the model to "pay attention" to various important parts of the feature maps that can help in making more accurate predictions.

**Transformer**: Transformers are architectures that are comprised of only the attention layers, leaving out any recurrence or convolutions entirely [20]. They have been adapted for the task of time series forecasting as they work very well with longer sequences [18, 22]. For this work, we use the encoder structure of Transformers and work with a single stack of two-headed self attention modules and other standard layers based on [20].

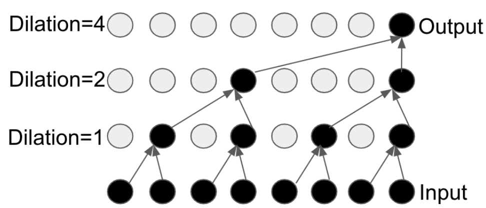



***Figure 2.*** *Dilation in kernels [15, 5] .*

**Probabilistic prediction**: For probabilistic forecasts, the above models are modified to output predictions at multiple quantiles (from 5% to 95%). While the point models are trained with mean-squared-error losses, their probabilistic counterparts are trained using quantile loss.

A fully connected layer at the end of each model is modified to produce either a single output (for point) or multiple outputs (for probabilistic). Ngboost is trained with default parameters and 2000 estimators as in [23].

## 5. Evaluation

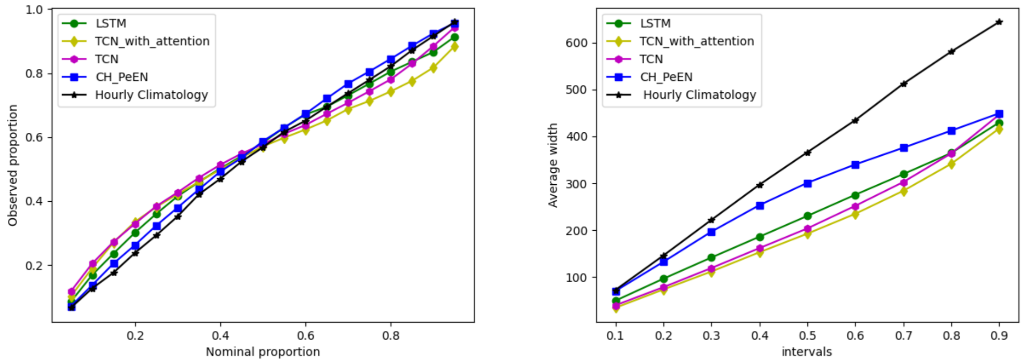
We provide the results of our experiments over all 7 SURFRAD stations for the test year (2018) in Table 1 and Table 2. The benchmarks from the solar energy literature (derived from [7]) are:

**Hourly Climatology (HC)** is a model that assigns the irradiance at a certain hour in 2018, to be the average of all irradiance values at the same hour of every day in the training data. For the probabilistic forecast evaluation, we do not use the average but the cumulative distribution function (CDF) over these values.

**Complete History Persistence ensemble (CH-PeEN):** CH-PeEN is used as a probabilistic prediction benchmark, where for a certain forecast hour, we take a CDF over the clearness indices at the same hour of every day from the training data and these are further converted to irradiance measures.

**Numerical weather prediction (NWP) ensembles:** We are using the ECMWF 51-member ensemble as our NWP outputs. These members are updated only twice a day for one-week ahead forecasts, and hence had to be repeated for the rest of the hours of the day (to be consistent with other forecasts). For point forecasts, we take the ensemble mean, while for probabilistic prediction we take a empirical CDF over them.

**Smart Persistence (SP)** is a model that assumes the clearness index (ratio of GHI/clear-sky GHI) at time $t$+lead-time to be the same as at time $t$, and uses that to obtain the irradiance at $t$+lead-time. This is a common benchmark from the short-term point forecasting literature, which we would not expect to perform well at longer forecast lead times, but include for the sake of completeness.




(a) (b)

***Figure 3.*** *Reliability and Sharpness plots at Penn State station.*

**Evaluation metrics**: The evaluation metrics for point forecasting are the RMSE (root mean squared error) scores of each model. For probabilistic forecasting, we use CRPS or Continuous Ranked Probability Score. CRPS is a widely used metric for evaluating probabilistic forecasts as it balances reliability, resolution and sharpness which are other criteria to measure the quality of probabilistic outputs [10]. Intuitively, CRPS measures the area between the predicted and the observed CDF, the observed (true) CDF being a step function at the observation [23]. The lower the CRPS, the better the model. To evaluate our probabilistic forecasts, i.e., when our model outputs predictions at different quantiles ($\xi \in (0,1)$), the CRPS score can be expressed as an integral over quantile scores($QS$) at all quantiles (from [7]):

$$\texttt{CRPS} = \int_0^1 \frac{1}{T}\sum_{t=1}^{T} QS_\xi(P^{-1}(\xi,t), y(t))\, d\xi \tag{1}$$

where $y$ is the observation, 1 is an indicator function, $P$ the predicted CDF distribution and $T$ the number of data points. $QS$ at a particular $\xi$ is defined as:

$$QS_\xi = 2(1\left\{y(t) \leq P^{-1}(\xi,t)\right\} - \xi)(P^{-1}(\xi,t) - y(t)) \tag{2}$$

Reliability looks at the statistical consistency between the forecast distribution and observed distribution, while sharpness looks at the concentration (narrowness) of the forecast [7, 10, 12]. These characteristics are best observed with a visual analysis and we follow the work in [7] to visualize both reliability and sharpness. The sharpness plot in Figure 3b is where we plot the average forecast width at 10%,20%,30%,... central intervals. As we can see, sharpness doesn't look at the observation, it just considers the narrowness of the prediction interval. We also provide a reliability diagram (Figure 3a) where we compare the proportion of the observations that lie within a given quantile output, vs the quantile or nominal proportion itself (in an ideal scenario both are expected to be equal).

As we observe from Table 1, for the majority of the stations, all of our proposed deep learning models including LSTM, TCN, TCN+Attention and Transformers outperform Smart Persistence, Hourly Climatology and NWP for point forecasts. LSTM, TCN and TCN+Attention perform very well for point prediction. Ngboost performs comparably, or better, but falls behind in probabilistic evaluation. The probabilistic prediction results in Table 2 shows that TCN (and TCN+Attention) obtain superior results against all benchmarks except CH-PeEN in terms of the CRPS scores. Overall, LSTMs perform equally well. Transformers however do not come close to the other proposed models for both point and probabilistic evaluation (except for Desert Rock station). The CH-PeEN benchmark consistently performs slightly better than the best performing probabilistic models. To investigate this, we refer to the

| | SP | HC | NWP | Ngboost | LSTM | TCN | TCN+Attention | Transformer |
|---|---|---|---|---|---|---|---|---|
| Sioux Falls, SD | 222.34 | 220.14 | 288.84 | 177.01 | 205.22 | 190.9 | 199.63 | 229.67 |
| Fort Peck, MT | 203.11 | 225.12 | 285.92 | 161.11 | 149.14 | 150.64 | 158.73 | 172.94 |
| Bondville, IL | 269.6 | 230.99 | 298.91 | 196.78 | 210.83 | 205.32 | 211.21 | 242.24 |
| Penn State, PA | 235.57 | 225.67 | 280.43 | 187.04 | 192.71 | 199.02 | 193.69 | 227.08 |
| Boulder, CO | 242.01 | 220.52 | 329.78 | 177.21 | 184.68 | 182.56 | 187.59 | 213.74 |
| Desert Rock, NV | 146.83 | 196.15 | 379.63 | 110.76 | 104.26 | 110.75 | 137.52 | 117.52 |
| Goodwin Creek, MS | 252.01 | 230.06 | 327.24 | 184.37 | 184.49 | 186.61 | 187.75 | 210.28 |

***Table 1.*** *Results of the point forecasting pipeline. Results are in terms of RMSE scores (lower the better). Comparisons are made with the Smart Persistence (SP), Hourly Climatology (HC) and Numerical weather prediction (NWP) benchmarks.*

| | HC | CH-PeEN | NWP | Ngboost | LSTM | TCN | TCN+Attention | Transformer |
|---|---|---|---|---|---|---|---|---|
| Sioux Falls, SD | 126.04 | 87.75 | 218.42 | 106.41 | 93.96 | 94.03 | 98.14 | 153.29 |
| Fort Peck, MT | 129.37 | 77.37 | 217.76 | 97.83 | 80.05 | 76.65 | 77.83 | 95.59 |
| Bondville, IL | 131.79 | 100.84 | 225.15 | 119.07 | 117.79 | 107.87 | 110.68 | 134.31 |
| Penn State, PA | 126.9 | 97.74 | 199.22 | 114.38 | 101.37 | 107.43 | 103.3 | 131.04 |
| Boulder, CO | 126.46 | 88.65 | 256.79 | 102.91 | 92.42 | 91.47 | 91.25 | 114.77 |
| Desert Rock, NV | 112.46 | 44.78 | 311.33 | 55.13 | 45.21 | 46.43 | 56.43 | 51.11 |
| Goodwin Creek, MS | 130.4 | 95.91 | 249.16 | 111.05 | 95.26 | 98.6 | 100.63 | 120.91 |

***Table 2.*** *Results of the probabilistic forecasting pipeline. Results are in terms of CRPS scores. Comparisons are made with the probabilistic Hourly Climatology (HC), Complete history persistence ensemble (CH-PeEN) and Numerical weather prediction (NWP) benchmarks. The lower the CRPS, the better the model.*

reliability and sharpness diagrams for the station Penn State in Figure 3. We clearly note all our proposed models have better sharpness (as their curves are lower) in their forecasts than CH-PeEN, even though it very reliable.

## 6. Discussion

Our encouraging results demonstrate that deep sequence learning algorithms hold promise for producing improved week-ahead forecasts as they outperform most of the literature benchmarks. Our methods also provide a distribution over the prediction, and this additional knowledge of uncertainty can be extremely important in efficient power system and generator planning. Our proposed models outperform a machine learning-based approach [8] in probabilistic forecasting.

| | SP | HC | Ngboost | LSTM | TCN | TCN+Attention | Transformer |
|---|---|---|---|---|---|---|---|
| Sioux Falls, SD | 222.34 | 220.14 | 171.53 | 187.37 | 183.71 | 189.97 | 215.57 |
| Fort Peck, MT | 203.11 | 225.12 | 156.82 | 147.31 | 164.02 | 153.48 | 176.53 |
| Bondville, IL | 269.6 | 230.99 | 196.36 | 193.85 | 212.27 | 198.57 | 229.75 |
| Penn State, PA | 235.57 | 225.67 | 186.48 | 196.78 | 206.61 | 213.77 | 218.41 |
| Boulder, CO | 242.01 | 220.52 | 177.53 | 183.75 | 196.69 | 190.98 | 224.79 |
| Desert Rock, NV | 146.83 | 196.15 | 111.83 | 104.68 | 120.23 | 130.51 | 125.34 |
| Goodwin Creek, MS | 252.01 | 230.06 | 180.41 | 190.9 | 209.99 | 206.94 | 218.45 |

***Table 3.*** *Results of the point forecasting pipeline with NWP ensemble included as features in our models. Results are in terms of RMSE scores.*

While Temporal CNNs are a faster alternative to training LSTMs, they show an almost equal performance in this application, especially for probabilistic forecasts. The Attention mechanism proved useful when used in conjunction with TCNs but notably, not as much when we dispensed the convolution layers and used a Transformer which is entirely based on attention.

Furthermore, as our part of our study, we wanted to look into the potential performance of our existing deep learning models if they are provided with an additional input feature of the NWP model

| | HC | CH-PeEN | Ngboost | LSTM | TCN | TCN+Attention | Transformer |
|---|---|---|---|---|---|---|---|
| Sioux Falls, SD | 126.04 | 87.75 | 102.85 | 88.26 | 106.89 | 121.44 | 131.73 |
| Fort Peck, MT | 129.37 | 77.37 | 97.2 | 75.03 | 83.51 | 81.76 | 95.77 |
| Bondville, IL | 131.79 | 100.84 | 119.29 | 109.3 | 113.11 | 113.61 | 131.93 |
| Penn State, PA | 126.9 | 97.74 | 114.35 | 99.61 | 118.22 | 117.47 | 135.36 |
| Boulder, CO | 126.46 | 88.65 | 103.16 | 92.2 | 103.03 | 100.83 | 120.93 |
| Desert Rock, NV | 112.46 | 44.78 | 55.15 | 44.02 | 47.01 | 60.04 | 48.59 |
| Goodwin Creek, MS | 130.4 | 95.91 | 112.11 | 96.13 | 121.14 | 120.4 | 128.56 |

***Table 4.*** *Results of the probabilistic forecasting pipeline with NWP ensemble included as features in our models. Results are in terms of CRPS scores.*

ensemble. Table 3 and 4 provide the results obtained when the 51 member ensemble is incorporated into our models. With the poor temporal resolution of these NWP predictors, we did not expect to see a huge performance improvement in the forecasts. We do observe an overall slight enhancement in performance with LSTM but not a clear trend with the TCN, TCN+Attention and Transformer models. Further investigation is left to future work.

## 7. Conclusion

We provide a quantitative study and demonstrate the valuable potential of deep learning methods for one week-ahead solar irradiance forecasting, especially when such longer-term predictions are ill-served by existing NWP models. Week-ahead and longer forecasts, coupled with uncertainty estimates, can be very significant for future power systems operations when efficient energy planning will become increasingly important with greater shares of renewable energy penetrating into power systems. We hope this paper will encourage future work leveraging machine learning for long-term point and probabilistic forecasting, not only for solar power, but also for other renewables and applications mitigating climate change.

**Funding Statement.** No funding declared.

**Competing Interests.** Claire Monteleoni is the Editor-in-Chief of Environmental Data Science: This paper was independently reviewed and accepted for publication.

**Data Availability Statement.** We use open-source NOAA's SURFRAD network (Surface radiation budget network for atmospheric research) [2] that provides the ground-truth solar irradiance and meteorological measurements from seven sites across the US in different climatic zones (https://gml.noaa.gov/grad/surfrad/).

**Author Contributions.** Conceptualization: S.S; B.H; C.M. Methodology: S.S; C.M. Data Curation: B.H; S.S. Supervision: C.M; B.H. Writing original draft: S.S; C.M; B.H. Writing – review and editing: S.S; C.M; B.H. All authors approved the final submitted draft.

## References

1. Ahmad Alzahrani, Pourya Shamsi, Cihan Dagli, and Mehdi Ferdowsi. Solar irradiance forecasting using deep neural networks. *Procedia Computer Science*, 114:304–313, 2017.
2. John A Augustine, John J DeLuisi, and Charles N Long. Surfrad–a national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society*, 81 (10):2341–2358, 2000.
3. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1409.0473.
4. Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
5. Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*, 2017.

6. Banalaxmi Brahma and Rajesh Wadhvani. Solar irradiance forecasting based on deep learning methodologies and multi-site data. *Symmetry*, 12(11):1830, 2020.
7. Kate Doubleday, Vanessa Van Scyoc Hernandez, and Bri-Mathias Hodge. Benchmark probabilistic solar forecasts: Characteristics and recommendations. *Solar Energy*, 206:52–67, 2020.
8. Tony Duan, Avati Anand, Daisy Yi Ding, Khanh K Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. Ngboost: Natural gradient boosting for probabilistic prediction. In *International Conference on Machine Learning*, pages 2690–2700. PMLR, 2020.
9. André Gensler, Janosch Henze, Bernhard Sick, and Nils Raabe. Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks. In *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 002858–002865. IEEE, 2016.
10. Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2): 243–268, 2007.
11. Claire Granier, Sabine Darras, Hugo Denier van der Gon, Doubalova Jana, Nellie Elguindi, Galle Bo, Gauss Michael, Guevara Marc, Jukka-Pekka Jalkanen, Jeroen Kuenen, et al. *The Copernicus atmosphere monitoring service global and regional emissions (April 2019 version)*. PhD thesis, Copernicus Atmosphere Monitoring Service, 2019.
12. Philippe Lauret, Mathieu David, and Pierre Pinson. Verification of solar irradiance probabilistic forecasts. *Solar Energy*, 194:254–271, 2019.
13. Yang Lin, Irena Koprinska, and Mashud Rana. Temporal convolutional neural networks for solar power forecasting. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.
14. Sakshi Mishra and Praveen Palanisamy. Multi-time-horizon solar forecasting using recurrent neural network. In *2018 IEEE Energy Conversion Congress and Exposition (ECCE)*, pages 18–24. IEEE, 2018.
15. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
16. Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison W. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 2627–2633. ijcai.org, 2017. doi: 10.24963/ijcai.2017/366. URL https://doi.org/10.24963/ijcai.2017/366.
17. Moumita Saha, Bhalchandra Naik, and Claire Monteleoni. Probabilistic and Point Solar Forecasting Using Attention Based Dilated Convolutional Neural Network. In *EGU General Assembly Conference Abstracts*, EGU General Assembly Conference Abstracts, page 12818, May 2020.
18. Huan Song, Deepta Rajan, Jayaraman J Thiagarajan, and Andreas Spanias. Attend and diagnose: Clinical time series analysis using attention models. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
19. Aidan Tuohy, John Zack, Sue Ellen Haupt, Justin Sharp, Mark Ahlstrom, Skip Dise, Eric Grimit, Corinna Mohrlen, Matthias Lange, Mayte Garcia Casado, et al. Solar forecasting: Methods, challenges, and performance. *IEEE Power and Energy Magazine*, 13(6):50–59, 2015.
20. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
21. Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, 198:111799, 2019.

22. Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*, 2020.
23. Eric Zelikman, Sharon Zhou, Jeremy Irvin, Cooper Raterink, Hao Sheng, Jack Kelly, Ram Rajagopal, Andrew Y Ng, and David Gagne. Short-term solar irradiance forecasting using calibrated probabilistic models. *arXiv preprint arXiv:2010.04715*, 2020.
24. Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International conference on machine learning*, pages 7354–7363. PMLR, 2019.