Subseasonal Solar Power Forecasting via Deep Sequence Learning

Saumya Sinha[*]
University of Colorado
Boulder, CO, USA

Bri-Mathias Hodge
University of Colorado
National Renewable Energy Laboratory (NREL)
Boulder, CO, USA

Claire Monteleoni
University of Colorado
Boulder, CO, USA

Abstract

To help mitigate climate change, power systems need to integrate renewable energy sources, such as solar, at a rapid pace. Widespread integration of solar energy into the power system requires major improvements in solar irradiance forecasting, in order to reduce the uncertainty associated with solar power output. While recent works have addressed short lead-time forecasting (minutes to hours ahead), week(s)-ahead and longer forecasts, coupled with uncertainty estimates, will be extremely important for storage applications in future power systems. In this work, we propose machine learning approaches for these longer lead-times as an important new application area in the energy domain. We demonstrate the potential of several deep sequence learning techniques for both point predictions and probabilistic predictions at these longer lead-times. We compare their performance for subseasonal forecasting (forecast lead-times of roughly two weeks) using the SURFRAD data set for 7 stations across the U.S. in 2018. The results are encouraging; the deep sequence learning methods outperform the current benchmark for machine learning-based probabilistic predictions (previously applied at short lead-times in this domain), along with relevant baselines.

1 Introduction

Renewable energy resources such as wind and solar are abundantly available in nature and have the potential to reduce society's dependence on fossil fuels. However, 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 enable reliable 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 [18, 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 [18].

The works in [1, 13, 9, 6] show the potential of Long Short-Term Memory (LSTMs) for solar energy forecasting. Convolution neural network (CNN)-based models that use dilated and causal convolutions along with residual connections (also referred as Temporal CNNs) were designed specifically for

[*]Corresponding author: Saumya Sinha, saumya.sinha@colorado.edu

Tackling Climate Change with Machine Learning workshop at NeurIPS 2021

sequential modeling [4, 14]. Temporal CNNs have recently been applied to forecasting day-ahead PV power output, outperforming both LSTMs and multi-layer feed forward networks [12] . In this work, we study a significantly longer forecast horizon, and compare Temporal CNNs, Temporal CNNs with an added attention layer [16], and the Transformer model [19], against LSTMs.

Probabilistic forecasting provides a distribution over the prediction, this additional knowledge of uncertainty estimates is an advantage over point forecasting. Knowing about future time periods of low and high uncertainty in advance can be very useful in planning [23]. Until recently, probabilistic forecasting for solar energy had not received as much attention as for other renewable energy sources, as observed in [7], which introduces probabilistic benchmarks to evaluate probabilistic methods, which we will utilize in this work. Recently, [23] show 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. We now consider NGBoost, *i.e.*, Natural Gradient Boosting algorithm [8] 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, 16].

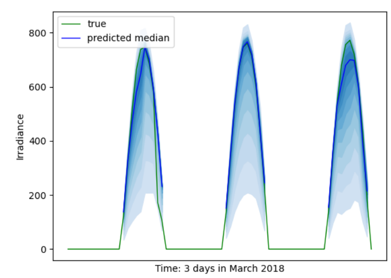

Figure 1: Fan plot showing the best performing (Transformer) model's prediction intervals from 5% to 95% percentile on three March days at the Boulder station.

Most physical models in this domain are based on NWP simulations that traditionally provide more accurate forecasts at hours to days-ahead lead times [18]. 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.*, 12 hour resolution in the case of European Centre for Medium-Range Weather Forecasts (ECMWF) model predictions). We train and test only on daylight hours (for relevance to the domain, as in [7]), and our forecasts are made at 168 daytime hours ahead. This is an approximately two week lead time, (depending on the number of daylight hours per day), a prediction lead-time known in the climate science literature as **subseasonal forecasting** [20].

Contributions We propose deep sequence learning for this subseasonal solar forecasting task, and demonstrate methods that outperform previous machine learning methods applied to solar forecasting. The results are comparable to the complete history persistence ensemble (Ch-PeEN) benchmark [7] in terms of CRPS scores, but better in terms of forecast sharpness.

Pathways to Climate Impact 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 . Efficient energy planning will become increasingly important as transportation switches to electric vehicles.

2 Methods

Models We focus on showcasing the potential of three deep multi-variate sequence models: Temporal CNNs, Temporal CNNs 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] and LSTM, along with benchmarks from the literature (as described in the Results section).

	Ngboost		**LSTM**		**TCN**		**TCN+Attention**		**Transformer**	
	SP	HC	SP	HC	SP	HC	SP	HC	SP	HC
Sioux Falls, SD	28.5	17.53	19.08	6.51	**29.59**	**18.79**	28.51	17.54	28.09	16.91
Fort Peck, MT	28.02	28.8	23.21	23.83	30.02	30.85	**30.66**	**31.48**	29.99	30.56
Bondville, IL	27.66	12.1	17.59	-0.06	29.23	14.33	**29.95**	**15.2**	26.97	11.26
Penn State, PA	26.88	14.48	22.42	9.26	**26.91**	**14.51**	26.19	13.67	25.27	12.6
Boulder, CO	30.69	15.72	26.42	10.65	28.01	12.45	29.93	14.8	**30.79**	**15.95**
Desert Rock, NV	28.1	40.46	22.45	35.56	25.07	37.95	29.25	41.41	**32.23**	**43.68**
Goodwin Creek, MS	31.8	18.26	24.09	8.75	31.54	17.94	30.82	17.09	**32.95**	**19.4**

Table 1: Results of the point forecasting pipeline. Results are in terms of skill scores (%) based on RMSE, using Smart Persistence (SP) and Hourly Climatology (HC) as the baselines. The higher the skill score, the better the model.

Temporal CNN (TCN): Temporal CNN is an autoregressive prediction framework (based on the WaveNet architecture [14]). It consists of multiple layers of dilated causal convolution filters (as explained in Figure 2) that are responsible for learning long term dependencies efficiently [4, 12].
Temporal CNN with Attention: Attention [3] has been used for sequential modeling and time series prediction problems [15]. 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: Transformer models have been adapted for the task of time series forecasting as they work very well with longer sequences [17, 22]. For this work, we use the encoder structure of Transformer and we work with a single stack of two-headed self attention modules (as it gives the best results) and other layers based on [19].
Probabilistic prediction: For probabilistic forecasts, the above models are modified to output predictions at multiple quantiles (from 5% to 95%, as in Figure 1)). While the point models are trained with mean-squared-error losses, their probabilistic counterparts are trained using quantile loss.

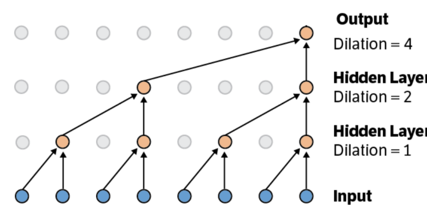

Figure 2: Dilated kernels as shown in [14, 5]. Figure shows how dilations help to increase (exponentially) the receptive field of a kernel. This makes the model capable of learning correlations between data points far apart in the past.

Hyperparameters are tuned on a validation dataset. 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]. Along with architectural parameters, the number of previous timesteps (sequence length) to look at when making the prediction was also tuned for each of the models. Adding past values of the input variables enable making better forecasts at longer lead times.

Data We use NOAA's SURFRAD network [2] that provides the ground-truth solar irradiance and meteorological measurements from seven sites across the US. Models are trained on measurements from the years 2016-2017, and then evaluated on the year 2018. Inputs are converted to an hourly resolution and only the day time values are considered for training and testing of all models (including benchmarks). We take a ratio of the ground-truth Global Horizontal Irradiance (GHI) (Watts/m2) with respect to the "clear sky" GHI value (these are irradiance estimates under cloudfree conditions, obtained from CAMS McClear Service [11]), to produce a clearness index like in [23, 13, 7] that is used as the prediction label for training. Important predictor variables available with 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 15 input variables overall). While trained on the clearness index, the models are evaluated on the GHI.

	HC	CH-PeEN	Ngboost	LSTM	TCN	TCN+Attention	Transformer
Sioux Falls, SD	123.08	91.49	98.58	94.51	91.68	93.03	**87.42**
Fort Peck, MT	126.78	80.88	84.6	82.75	78.58	78.28	**77.69**
Bondville, IL	129.5	103.11	108.63	120.97	101.23	**100.52**	104.71
Penn State, PA	123.93	100.9	106.32	111.19	103.13	102.7	**100.52**
Boulder, CO	122.26	**90.11**	95.76	98.96	94.29	94.58	91.5
Desert Rock, NV	104.35	44.33	49.63	**43.56**	46.3	45.37	44.99
Goodwin Creek, MS	124.45	95.66	99.0	105.26	97.24	97.97	**95.48**

Table 2: Results of the probabilistic forecasting pipeline. Results are in terms of CRPS scores. Comparisons are made with the probabilistic Hourly Climatology (HC) and CH-PeEN benchmarks. The lower the CRPS, the better the model.

3 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 three benchmarks from the solar energy literature (derived from [7]) are:
Smart Persistence (SP): 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.
Hourly Climatology (HC): 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 very similar to the probabilistic version of Hourly Climatology, except that a CDF is taken over the clearness indices (and not the irradiance itself) from the training data at the same hour.

The evaluation metrics for point forecasting are *skill scores*: the percentage improvement of the RMSE (root mean squared error) of each learned model (TCN, TCN+attention, Transformer, Ngboost and LSTM) over that of the benchmark (Smart Persistence and Hourly Climatology, respectively). For probabilistic forecasting, we use CRPS (Continuous Ranked Probability Score) scores to compare the models. Intuitively, CRPS measures the area between the predicted and the observed CDF, the observed (true) CDF being a step function at the observation [23]. CRPS is a widely used metric for evaluating probabilistic forecasts as it gives a single score that takes both reliability and sharpness into account [10]. 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].

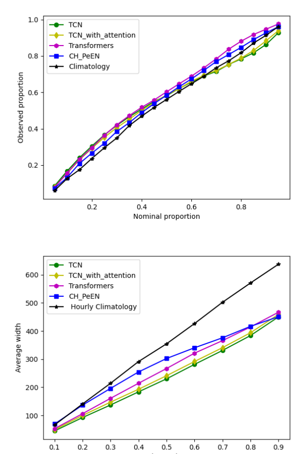

Figure 3: Reliability and Sharpness plots.

Overall, the three models (TCN, TCN+Attention, Transformer) outperform Ngboost, LSTM, and the hourly climatology for both point and probabilistic forecasts. Transformer outperforms TCN and TCN+Attention for probabilistic forecasts for most stations, except for Desert Rock, where the abundance of clear sky days in the desert make improvements difficult. Ngboost is close to the proposed deep learning methods in point prediction but falls behind in probabilistic evaluation. The CH-PeEN benchmark consistently performs very close to the best performing probabilistic models. To investigative this, we plot the reliability and sharpness diagrams for the station Penn State in Figure 3. We see that the TCN methods and Transformer have more sharpness (as their curves are lower) in their forecasts than CH-PeEN, even though it has comparable CRPS and reliability. The strong performance of CH-PeEN suggests that incorporating features based on the climatology into the machine learning models could further boost performance.

4 Discussion

We show the valuable potential of deep learning methods for subseasonal solar irradiance forecasting. Temporal CNNs are a faster alternative to training LSTMs, and in this application they show superior performance over LSTMs and Ngboost, especially for probabilistic forecasts. Attention mechanisms proved useful when used in conjunction with TCNs, and even more so with the Transformer. Notably, the observed performance did not require any NWP inputs. Future steps would be to include the NWP model ensemble outputs (available for standard forecast periods of $\sim$7-10 days lead-time) as input features to our existing deep models to potentially enhance performance. 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.

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] 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.

[13] 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.

[14] 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.

[15] 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`.

[16] 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.

[17] 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.

[18] 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.

[19] 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`.

[20] Frédéric Vitart and Andrew W Robertson. The sub-seasonal to seasonal prediction project (s2s) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1(1):1–7, 2018.

[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.