INL/CON-24-77082-Revision-0

# Entropy-Infused Deep Learning Loss Function for Capturing Extreme Values in Wind Power Forecasting

May 2024

Changing the World's Energy Future

Mucun Sun, Sergio Valdez, Juan M. Perez, Kevin.Garcia , Gael Galvan, Cesar Cruz, Yifeng Gao, Li Zhang

INL/CON-24-77082-Revision-0

# Entropy-Infused Deep Learning Loss Function for Capturing Extreme Values in Wind Power Forecasting

**Mucun Sun, Sergio Valdez, Juan M. Perez, Kevin.Garcia , Gael Galvan, Cesar Cruz, Yifeng Gao, Li Zhang**

**May 2024**

**Idaho National Laboratory**
**Idaho Falls, Idaho 83415**

**http://www.inl.gov**

**Prepared for the**
**U.S. Department of Energy**
**Under DOE Idaho Operations Office**
**Contract N/A**

# Entropy-Infused Deep Learning Loss Function for Capturing Extreme Values in Wind Power Forecasting

Mucun Sun
*Idaho National Laboratory*
Idaho Falls, ID, USA
Mucun.Sun@inl.gov

Sergio Valdez, Juan M. Perez, Kevin.Garcia,
Gael Galvan, Cesar Cruz, Yifeng Gao, Li Zhang
*The University of Texas Rio Grande Valley*
Edinburg, TX, USA

***Abstract*—Extreme scenarios in wind power generation occur with higher frequency and larger magnitude in the recent years due to the ever-increasing extreme meteorological factors. Accurate forecasting of the occurrence of extreme values in wind power generation is of great concern to ensure reliable power system operation. Recently, deep learning models have surged in popularity for wind power forecasting, with the mean squared error (MSE) loss function being commonly used. However, the MSE loss function, being sensitive to extreme values, disproportionately penalizes larger errors, cannot adequately capture the extreme values present in wind energy data, and novel loss functions have seldom been tailored for wind power forecasting. To this end, in this paper, we introduce a novel loss function specifically crafted to capture extreme values in wind power forecasting. The experimental results with four fundamental deep learning methods on open source wind power dataset validate that the new loss function is efficient and superior in all cases compared to MSE in capturing extreme values while maintaining forecasting performance.**

***Index Terms*—Loss function, wind power forecasting, extreme values, mean squared error**

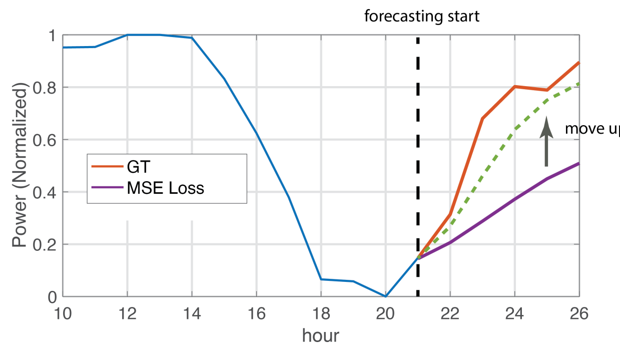


Fig. 1: Motivation of Proposed Work. Deep learning forecasting obtained by MSE loss (purple line) often has a large error in extreme value region (ground truth in red line). This paper proposes a novel loss function named Extropy Infused Forecasting loss (EIF), encouraging the model to fit extreme values while maintaining the overall performance.

## I. Introduction

With 50 GW annual installations in 2018 and 591 GW total installations worldwide, wind power contributes nearly a half of renewable power capacity (not including hydropower) by 2018. Accurate wind power forecasts benefit power system stakeholders from different perspectives. For example, wind power plants are subjected to a discounted price or even penalty for the underestimated or overestimated wind power generation during the market biding process. From a system operator's viewpoint, accurate wind power forecasting helps reduce the amount of operating reserves that are needed to balance generation and load [1]. For policymakers and investors of energy sector, accuracy wind power forecasting is important for wind energy resource assessment and wind farm design [2]. However, the inherent variability and uncertainty in wind power are the major concerns for reliable and economic power system operations with a high wind energy penetration.

Deep learning model, as a promising type of machine learning capable for discovering the inherent nonlinear features and high-level invariant structures in data, has been frequently adopted in the literature to improve wind power forecasting accuracy [3]. A crucial aspect of deep learning models is the tuning of parameters. Typically, this is achieved by optimizing a loss function during training. The primary purpose of the loss function is to measure the discrepancy between a model's predictions and the actual target values. The learning algorithm of the model strives to minimize this loss function, utilizing the gradients derived from it to refine the model's parameters.

The most common loss function in deep learning is the Mean Square Error (MSE). However, MSE might not be the best fit for wind power forecasting due to their inherent properties. MSE-based prediction tends to heavily penalize large errors by squaring them, making MSE Loss less robust to extreme values. This characteristic can be problematic in wind power forecasting, where the variability of wind energy sources often results in extreme values. Such extreme values might emerge from scenarios like high seasonal variance or sensor malfunctions induced by severe weather conditions. In summary, as illustrated in Figure 1, we aim to propose a novel loss that achieves better performance in the extreme value region while achieving similar overall performance compared with popular MSE loss.

To address the aforementioned cold-start problem, several studies aim to develop new metrics or altering on existing metrics to assist wind power or wind speed forecasting. For example, Loutfi *et al.* [4] propose a novel loss function for electricity market price forecasting. To overcome the shortcomings of MSE while training neural networks, several alternative loss functions has been proposed. Huber loss was proposed as a piecewise function of both MSE and MAE, in which a boundary hyper-parameter $\delta$ determines which one of the two should be used. Unfortunately, finding the right value for $\delta$ increases the complexity of training neural networks, which already have a sufficiently large number of other hyperparameters that must be finetuned. [5]. Chen *et al.* proposed a kernel MSE loss function designed for wind speed forecasting. The evaluation of the loss function is shifted from the original feature space to a higher-dimensional Hilbert space to capture the ubiquitous nonlinearity of deep learning errors. Despite these aforementioned methods, current loss functions remain inadequately tailored for addressing extreme values in energy forecasting. Consider, for example, the Texas electric power crisis in February 2021, triggered by unexpected and severe cold weather [6], or the obscured Global Horizontal Irradiance (GHI) forecasting in the New York Independent System Operator (NYISO) area due to the Canadian wildfires in 2021 [7], or the annular solar eclipse in 2023 which affect solar forecasting [8]. Underestimation of wind power generation can lead to poor utility service, including potential blackouts, and may force utility companies to purchase energy from external sources at a higher cost. Therefore, it becomes compelling to explore the development of a loss function that is specifically sensitive to and can effectively account for these extreme values in the wind power forecasting.

In this study, our approach to the problem is systematic. Initially, we evaluate four fundamental deep learning methods on open source wind power dataset, using the traditional MSE as the loss function. This serves to highlight MSE's limitations, specifically its inability to capture extreme values in wind power forecasting. Subsequently, we delineate the criteria our ideal loss function should meet. The key research questions of this paper are: Does minimizing MSE truly capture extreme values in wind power forecasting? And which promising metrics can serve as the foundation for a candidate loss function to address these shortcomings? Which properties should the proposed loss function have in pursuit of extreme value capturing?

To answer these research questions, we performed a conceptual and an empirical analysis. The paper is organized as follows. In section II, we formulate the problem and underscore the shortcomings of MSE. We validate these limitations by employing the MSE loss function across various deep neural network models on an open-source wind energy datasets, demonstrating that this issue is existing in wind power time series forecasting. In Section III, we outline the properties that our ideal loss function should meet. Additionally, we review the current literature to explore potential solutions for refining loss functions. In this paper, our objective is to propose a loss function that mitigates the limitations of MSE in accurately capturing extreme values in wind power forecasting, while maintaining the overall forecasting performance. Section IV applies the developed loss function to the same wind energy dataset and compares the model performance across several benchmark deep neural network forecasting models. Concluding remarks and future work are discussed in Section V.

## II. Problem Formulation

### A. Basic Deep Learning Models

To highlight MSE's limitations, specifically its inability to capture extreme values in wind power forecasting, we evaluate four different popular used deep neural network on wind energy time series dataset at Dallas, using the traditional MSE as the loss function. Selected deep neural network models include: multilayer perceptron (MLP), Convolutional neural network (CNN), Gated recurrent units (GRU), and Long short-term memory (LSTM). These deep learning algorithms have been shown to perform well at different forecasting horizons in the literature for wind power forecasting [3]. The selected models have also shown a similar level of performance and acceptable computational cost in the literature [3]. Since this paper concentrates on discussing loss functions in deep learning models, other algorithms such as linear regression and statistical methods are not included. The four selected deep learning algorithms are briefly introduced in this section.

MLP is a fundamental type of neural network architecture used extensively in the field of machine learning. Characterized by its layered structure, an MLP consists of an input layer, one or more hidden layers, and an output layer. Each layer is made up of nodes, or neurons, which are interconnected and transmit information using weighted connections. These networks are adept at learning complex patterns and relationships in data through a process known as backpropagation, making them particularly useful for timeseries forecasting.

CNN is one of the most widely used deep learning models for extracting hierarchical spatially invariant features [**?**]. CNN usually consists of convolutional layers, pooling layers, and fully-connected layers. At the convolutional layers, a convolution operation is applied to the input feature maps and then these feature maps are transformed through a non-linear activation function. Through the convolution operation, most revealing features of the input can be extracted. The convolution operation process can be expressed as:

$$x_m^o = \Phi\Big(\sum_{n \in D_m} x_n^{o-1} * k_{mn}^o + b_m^o\Big) \tag{1}$$

where $x_m^o$ denotes the $m$th feature map of the $o$th layer, $*$ denotes the convolution operation, $D_m$ is the inputs pool, $b_m^o$ denotes the bias, $k_{mn}$ is the kernel of the $o$th layer, and $\Phi(\cdot)$ is the activation function.

The goal of pooling layer is to progressively reduce the spatial size of the output feature maps from the convolutional layer through a down sampling function, which is expressed as:

$$x_m^o = d(\cdot)(x_m^{o-1}) \tag{2}$$

where $d(\cdot)$ denotes the down sampling function. The feature map is downsampled in such a way that the maximum feature response within a given sample size is retained.

The fully-connected layer connects every neuron in one layer to every neuron in another layer, thus combining all local features into global features to form the output:

$$x^o = \Phi(\omega^o x^{o-1} + b^o) \tag{3}$$

where $\omega^o$ and $b^o$ are the weight matrix and bias vector of the $o$th layer, respectively.

GRU is a type of recurrent neural network (RNN) architecture renowned for its efficiency in processing sequential data. Developed as a more streamlined alternative to the traditional LSTM networks, GRUs simplify the structure while retaining the capability to handle long-term dependencies in data. This is achieved through the use of specialized gating units, which control the flow of information and allow the network to retain or discard data based on its relevance. This mechanism helps in addressing the common problem of vanishing gradients in standard RNNs.

LSTM is also a special RNN architecture for time series modeling and forecasting, which has the capability of learning and memorizing long-term dependencies within the time-series data. The standard RNN have one hidden layer, which could only trace back to few time steps due to the vanishing gradient effect [9]. To better capture the long-term dependencies, LSTM introduces different gates which could regulate the gradient flow of the network. A standard LSTM unit has input gate, forget gate, and output gate. LSTM updates its hidden state $c_t$ by using the current input $x_t$ and the previous state $c_{t-1}$. The final state $h_t$ is determined by $c_t$ and $o_t$. The weights are optimized by minimizing the difference between the LSTM outputs and training samples.

## III. METHODOLOGY

### A. Identification of Extreme Wind Power Values

The definition of extreme wind power values is a fundamental issue in our work. Different definitions may lead to somewhat different extreme wind power value identification, which may affect the design of loss function. Generally, extreme wind power generation refers to such as low (turbines produce little or no net power) or high (turbines produce their rated maximum power or closer) generation and ramps in generation [10]. In this paper, we only focus on low or high generation. Specifically, a couple of definition methods were used in the literature to define low and high wind power. For example, in [11], a $3\sigma$ criterion is adopted to identify extreme value. However, the shape of wind power is generally skewed toward lower power, featuring a long tail that extends toward the maximum power output, which the $3\sigma$ criterion doesn't fit. In this paper, for simplicity, we approximate the extreme low and high wind power values to be 20% and 80% of the rated wind power, respectively.

### B. Properties of Loss Function

To address these shortcomings illustrated in II, we outline a set of desired properties that our potential alternative loss function should meet as follows.

1) **Property 1**: Less sensitive to extreme low and high wind power values.
2) **Property 2**: Maintain forecasting performance using traditional MSE as loss function.

*1) Traditional MSE loss function:* In deep neural network-based models, the loss function is used to evaluate the difference between the ground truth and model prediction. The unknown hyperparameters in the model are estimated by minimizing the loss function. Traditional loss function such as MSE is expressed as follow:

$$L_{MSE} = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2 \tag{4}$$

where, $N$ denotes the number of samples, while, $\hat{y}_n$ and $y_n$ denote the prediction and observations, respectively. A higher MSE error value corresponds to poorer prediction performance. While MSE address errors in a linear space, they fall short in efficiently capturing the nonlinearity of errors [12]. In addition, the MSE loss function squares errors, therefore, large errors from extreme values can disproportionately dominate the optimization, misleading the optimization direction.

*2) Proposed loss function:* In this subsection, we will introduce our proposed Entropy Infused Forecasting Loss (EIF Loss). Intuitively, in addition to basic MSE loss, we will add an 'external' penalty term, which is modeled by cross entropy between low and high wind power stages, to guide the model to additional focus on extreme cases — extreme low and high wind power capacity.

First, we will define our target extreme set:

$$\mathcal{E} = \{y_n| \quad y_n \leq \delta_1 \quad or \quad y_n \geq \delta_2\} \tag{5}$$

where $\delta_1$ is the threshold of the wind power capacity that we consider extremely high and $\delta_2$ is the threshold of the wind power capacity that we consider extremely low. Similarly, we could define:

$$\mathcal{E}_l = \{y_n| \quad y_n \leq \delta_1\} \tag{6}$$

$$\mathcal{E}_u = \{y_n| \quad y_n \geq \delta_2\} \tag{7}$$

which denotes low and high extreme wind power cases.

Then, the proposed loss function is defined as:

$$L(\theta, \eta) = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2 + \lambda\mathbb{1}\{y_n \in \mathcal{E}\}CE(\hat{y}_n, c_n) \tag{8}$$

where $\lambda$ is the hyper-parameter which controls the penalty strength and $\mathbb{1}$ is the indicator function defined as:

$$\mathbb{1}\{x = a\} = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases} \tag{9}$$

TABLE I: 6-hour ahead wind power forecasts

| | | MLP | | CNN | | GRU | | LSTM | |
|---|---|---|---|---|---|---|---|---|---|
| $i^{th}$ hour | | Vanilla | Ours | Vanilla | Ours | Vanilla | Ours | Vanilla | Ours |
| 1 | RMSE | 0.1306 | 0.1319 | 0.1419 | 0.1481 | 0.1296 | 0.1347 | 0.1329 | 0.1372 |
| | MAE | 0.0876 | 0.0814 | 0.0978 | 0.1043 | 0.0863 | 0.0907 | 0.0885 | 0.089 |
| 2 | RMSE | 0.1702 | 0.1715 | 0.1774 | 0.1798 | 0.1719 | 0.1758 | 0.1768 | 0.1813 |
| | MAE | 0.1191 | 0.1121 | 0.1256 | 0.1267 | 0.1206 | 0.1212 | 0.1238 | 0.1223 |
| 3 | RMSE | 0.1964 | 0.1978 | 0.2031 | 0.2031 | 0.1993 | 0.2020 | 0.2076 | 0.2105 |
| | MAE | 0.1415 | 0.1347 | 0.1486 | 0.1449 | 0.1455 | 0.1435 | 0.1526 | 0.1493 |
| 4 | RMSE | 0.2165 | 0.2186 | 0.2238 | 0.2220 | 0.2202 | 0.2226 | 0.2317 | 0.2339 |
| | MAE | 0.1607 | 0.1550 | 0.1685 | 0.1623 | 0.1662 | 0.163 | 0.1769 | 0.1732 |
| 5 | RMSE | 0.2320 | 0.2343 | 0.2399 | 0.2374 | 0.2356 | 0.238 | 0.2501 | 0.2523 |
| | MAE | 0.1771 | 0.1721 | 0.1848 | 0.1773 | 0.1813 | 0.1787 | 0.1967 | 0.1926 |
| 6 | RMSE | 0.2453 | 0.2483 | 0.2538 | 0.2504 | 0.2488 | 0.2519 | 0.2660 | 0.2685 |
| | MAE | 0.1906 | 0.1873 | 0.1986 | 0.1900 | 0.1931 | 0.192 | 0.2139 | 0.2088 |
| MAE Improvement% | | - | **4.31%** | **-** | **1.17%** | **-** | **0.04%** | **-** | **1.56%** |
| RMSE Improvement% | | - | - 0.94% | - | -0.42% | **-** | -1.81% | **-** | **1.65%** |

.

$\mathbb{1}\{y_n \in \mathcal{E}\}$ indicate that only the samples $y_i$ that belong to $\mathcal{E}$ set will add the right side penalty term.

And $c_n$ is defined as:

$$c_n = \begin{cases} 1 & y_n \in \mathcal{E}_u \\ 0 & y_n \in \mathcal{E}_l \end{cases} \tag{10}$$

which each $c_n$ can be considered as the 'label' of high power (when equal to 1) and low power (when equal to 0). And $CE(.,.)$ is the cross entropy loss:

$$CE(\hat{y}_i, c_i) = \mathbb{1}\{c_i = 1\}\log(\hat{y}_i) + \mathbb{1}\{c_i = 0\}\log(1-\hat{y}_i) \tag{11}$$

Intuitively, $CE(y_n, c_n)$ reaches minimal value when $y_n = c_n$. The loss can achieve two advantages. First, the loss will discourage output stay close to 0.5, which guides the model to move quickly to fit high and low power when there is a sudden change existed. Second, the loss has a more steep gradient than MSE Loss when it is close to the actual $c_n$, which encourages the model to forecast $y_n \in \mathcal{E}$ more accurately than the regular region.

Note that in the experiment, we will set $\delta_1 = 0.2$, $\delta_2 = 0.8$ according to our discussion in Section III-A.

## IV. Experiments and Case Studies

### A. Dataset Summary

In this study, we use wind data from Wind Integration National Dataset (WIND) Toolkit [13]. Particularly, we use the data from Dallas, Texas. The wind dataset includes meteorological information (e.g., wind direction, wind speed, air temperature, surface air pressure, density at hub height), and synthetic actual wind power. The dataset comprises six years of hourly time series data. We normalized the wind power data against the wind turbine capacity.

### B. Experiment Setting

The effectiveness of proposed EIF loss was tested on four baseline deep learning models: MLP, CNN, GRU, and LSTM, which have shown good performance on wind power forecasting in the literature [3]. We conducted experiments using our proposed EIF loss in comparison to the standard MSE loss function. Forecasting performance was evaluated by RMSE and MAE metrics. Moreover, the performance on extreme set region $\epsilon$ was explored. The goal is to show the effectiveness of EIF loss function capturing extreme values while maintaining overall forecasting performance across different deep learning models.

In our experiments, we utilize 48-hour of time series data to predict the subsequent 6-hour period wind power generation. The train/validation/test ratio is set to 50%, 20%, 30%. The hyperparameter $\lambda$ is set to be $0.1$.

### C. Forecasting Performance Evaluation

Two evaluation metrics are adopted to evaluate our proposed EIF loss function on wind power forecasting, which are the mean absolute error (MAE) and the root mean square error (RMSE). The mathematical expressions of the two metrics are expressed as:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|\hat{y}_n - y_n| \tag{12}$$

$$RMSE = \frac{1}{N}\sqrt{\sum_{n=1}^{N}(\hat{y}_n - y_n)^2} \tag{13}$$

where $\hat{y}_n$, and $y_n$, are the forecast wind power and wind power observation, respectively. For these metrics, a smaller value indicates better forecasting performance.

Table I shows the results of comparing our proposed EIF loss with the vanilla MSE loss across four baseline models. Results indicate that our proposed EIF loss function enhances

TABLE II: 6-hour ahead wind power forecasts with distinguished extreme value regions

| Regions | | MLP | | CNN | | GRU | | LSTM | |
|---|---|---|---|---|---|---|---|---|---|
| | | Vanilla | Ours | Vanilla | Ours | Vanilla | Ours | Vanilla | Ours |
| $y_n \in \mathcal{E}_u$ | RMSE | 0.2389 | **0.2302** | 0.2439 | **0.2234** | 0.3413 | **0.3326** | **0.3276** | 0.3332 |
| | MAE | 0.1758 | **0.1551** | 0.1787 | **0.1404** | 0.2523 | **0.2317** | 0.2337 | **0.2326** |
| $y_n \in \mathcal{E}_l$ | RMSE | 0.1906 | **0.1889** | 0.2103 | **0.1982** | 0.1019 | **0.1039** | 0.1017 | **0.0994** |
| | MAE | 0.1317 | **0.121** | 0.1549 | **0.1238** | **0.0562** | 0.0564 | 0.0557 | **0.0553** |
| $y_n \in \mathcal{E}$ | RMSE | 0.2125 | **0.2074** | 0.2252 | **0.2093** | 0.1442 | **0.1435** | 0.1409 | **0.1407** |
| | MAE | 0.1504 | **0.1335** | 0.165 | **0.1309** | 0.0754 | **0.0736** | 0.0732 | **0.0709** |

the overall MAE performance across all four baseline models, with improvement up to $4.3\%$. On the other hand, the RMSE values remain similar. In two out of the four baseline models, the change in RMSE value is less than $1\%$. One model (GRU) decreases $1.81\%$ and the other model (LSTM) increases $1.65\%$. The result demonstrates that the proposed EIF loss function has not only slightly improved overall MAE performance but also maintains similar RMSE performance compared with MSE Loss.

Table II presents the performance comparison of our proposed EIF loss against the vanilla MSE Loss in the extreme value regions across four benchmark deep learning models. We report the result in three types of extreme value regions: $\mathcal{E}_u$, $\mathcal{E}_l$, and overall $\mathcal{E}$. From the result, the proposed loss enhances the performance in 2 out of 24 comparison cases. The result indicates that our proposed model can enhance the performance in the extreme region.

It is important to note that in this paper, the thresholds defining extremely low wind power and extremely high wind power values in the loss function are set at 20% and 80% of the wind turbine's rated power, respectively. Ideally, wind turbines would convert 100% of the wind they encounter into energy. However, due to factors like friction, achieving this efficiency is impractical in real-world scenarios. As a consequence, wind turbines typically operate at a much lower of their capacity. This operational range varies depending on the turbine type and its power factor, which allows for the adjustment of these thresholds as needed. Consequently, this loss function is adaptable and can be tailored to suit various wind power forecasting scenarios (e.g., different turbine type, wind farm location).

## V. Conclusion

This paper developed a novel entropy infused loss function for wind power forecasting to better capture extreme value while maintain or enhance forecasting performance. Results of the case study showed that: (i) the proposed loss function improves MAE and generates comparable RMSE relative to the use of MSE as loss function; (ii) the proposed EIF loss excels in accurately capturing extreme values. Future work will explore: (i) adaptive integrate extreme value threshold as a parameter in the loss function, and (ii) generalize the loss function to extreme events (e.g., wind power ramp) instead of extreme value.

## References


[1] A. Botterud, Z. Zhi, J. Wang, R. J. Bessa, H. Keko, J. Mendes, J. Sumaili, and V. Miranda, "Use of wind power forecasting in operational decisions." Argonne National Lab.(ANL), Argonne, IL (United States), Tech. Rep., 2011.

[2] N. Izadyar, H. C. Ong, W. Chong, and K. Leong, "Resource assessment of the renewable energy potential for a remote area: a review," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 908–923, 2016.

[3] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, 2019.

[4] A. A. Loutfi, M. Sun, I. Loutfi, and P. B. Solibakke, "Empirical study of day-ahead electricity spot-price forecasting: Insights into a novel loss function for training neural networks," *Applied Energy*, vol. 319, p. 119182, 2022.

[5] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, pp. 1–26, 2020.

[6] G. Zhang, H. Zhong, Z. Tan, T. Cheng, Q. Xia, and C. Kang, "Texas electric power crisis of 2021 warns of a new blackout mechanism," *CSEE journal of Power and Energy Systems*, vol. 8, no. 1, pp. 1–9, 2022.

[7] T. W. Juliano, P. A. Jiménez, B. Kosović, T. Eidhammer, G. Thompson, L. K. Berg, J. Fast, A. Motley, and A. Polidori, "Smoke from 2020 united states wildfires responsible for substantial solar energy forecast errors," *Environmental Research Letters*, vol. 17, no. 3, p. 034010, 2022.

[8] S. Wilcox, B. Roberts, M. Sengupta, A. Habte, J. Yang, and T. Stoffel, "Clear-sky probability for the 2023 annular solar eclipse and the 2024 total solar eclipse using the nrel national solar radiation database," National Renewable Energy Laboratory (NREL), Golden, CO (United States), Tech. Rep., 2023.

[9] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[10] D. J. Cannon, D. J. Brayshaw, J. Methven, P. J. Coker, and D. Lenaghan, "Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in great britain," *Renewable Energy*, vol. 75, pp. 767–778, 2015.

[11] Y. Shi, Y. Liu, and X. Gao, "Study of wind turbine fault diagnosis and early warning based on scada data," *IEEE Access*, vol. 9, pp. 124 600–124 615, 2021.

[12] X. Chen, R. Yu, S. Ullah, D. Wu, Z. Li, Q. Li, H. Qi, J. Liu, M. Liu, and Y. Zhang, "A novel loss function of deep learning in wind speed forecasting," *Energy*, vol. 238, p. 121808, 2022.

[13] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015.