Ensuring Cyberattack-Resilient Load Forecasting with A Robust Statistical Method

Jieying Jiao, Zefan Tang, Peng Zhang, Meng Yue, Chen Chen, Jun Yan

***Abstract*—Cyberattacks in power systems can alter load forecasting models' input data. Although extreme outliers that fail to follow regular patterns can be easily identified, other more carefully-designed attacks can escape detection and seriously impact load forecasting. While existing work mainly focuses on enhancing attack detection, we propose a cyberattack-resilient load forecasting model that is based on an adaptation of classic Huber's robust statistical method. In a large-scale simulation study, the proposed method performed better than the classic method in various settings.**

***Index Terms*—Cyber security, power systems, load forecasting, Huber's robust method, regression model**

I. Introduction

As energy delivery systems evolve and become increasingly reliant on sophisticated forecasting data for efficient operations, they also become more vulnerable to cybersecurity issues. As one of the key elements for enabling utilities to make decisions or options such as purchasing and generating electric power and load switching, an accurate load forecasting will reduce the risk and unexpected cost, especially with the increasing uncertainties introduced by non-dispatchable distributed energy resources (DERs) such as solar and wind generation, demand-side management (DSM), and responsive load. Thus, it is crucial to make the load forecasting process resilient to possible cyberattacks.

Existing load forecasting models and techniques can be roughly divided between statistical approaches and artificial intelligence-based approaches [1]–[3]. We focus on statistical approaches to improving the robustness of load forecasting. Statistical load forecasting methods include point forecasting, probabilistic forecasting and ensemble forecasting. Regression models [4], exponential smoothing models [5], and various time series models such as autoregressive moving average (ARMA) [6] are most frequently used for point forecasting, which provide a single point forecast. Probabilistic load forecasting methods can provide the intervals, scenarios, density functions or probabilities about the desired future load [7], [8], allowing uncertainty assessment in decision-making. Ensemble methods combine several different forecasting methods together for better performance [9]–[11].

In spite of the voluminous literature on load forecasting, relatively little work has been done on cyberattack-resilient load forecasting. Some authors have proposed attack detection methods and different ways to treat the identified attacked data [12]–[14]. In general, the detection methods can be categorized as descriptive analytic methods and model based methods. Descriptive analytic methods do not rely on the load forecasting model. They start from the properties of the data to identify abnormal data. [15] introduced an outlier detection method that uses the Chebyshev inequality to produce upper and lower limits. [16] applied the property of second order difference (SOD) to this field. [17] combined the two methods mentioned above with symbolic aggregation approximation. Other detection methods are based on forecasting models where a model is fitted first and then the outliers are identified from investigating the residuals. [18] used a regression model and fixed threshold to detect residual outliers. [19] used a dynamic model to do short-term forecasting, which improved the detection method by updating the threshold for residuals as new data came in. Both works used the GEFCom2014 data [20] from the ISO New England.

Our contribution is a new robust load forecasting method in a multiple linear regression setting [21]. The proposed approach assigns weights to observations based on the extremeness of their residuals. Observations whose residuals are large in magnitude are downweighted. The tuning parameter of the method is the percentile of the absolute residuals at which the downweighting starts. With the GEFCom2012 data, classic Huber's robust method and our modified version are compared under various attacks in a simulation study. With an appropriate tuning parameter, the proposed method has a much lower mean absolute percentage error than the classic Huber's method when the percentage of the attacked data is high or when the magnitude of the attack is high.

The rest of this paper is organized as follows: Section II reviews Tao's vanilla benchmark forecasting model [21] and describes two attack models that will be used. Section III presents classic Huber's robust method and our modified version based on the quantiles of the absolute residuals. A simulation study to compare the performance of different forecasting methods under different attack models is reported in Section IV. Section V concludes with a discussion and proposals for future works.

This work was supported in part by the National Science Foundation under Grants ECCS-1611095, CNS-1647209 and ECCS-1831811, in part by the Department of Energy Cybersecurity for Energy Delivery Systems (CEDS), and in part by the Office of the Provost, University of Connecticut.

J. Jiao and J. Yan are with the Department of Statistics, University of Connecticut, Storrs, CT 06269, USA.

Z. Tang and P. Zhang are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA.

M. Yue is with Sustainable Energy Technologies Department, Brookhaven National Laboratory, Upton, NY 11973, USA.

C. Chen is with Energy Systems Division, Argonne National Laboratory, Lemont, IL 60439, USA.

978-1-7281-1981-6/19/$31.00 ©2019 IEEE

II. Load Forecasting Model and Attack Model

A. Load Forecasting Model

Load forecasting models and tools are well-developed and are already widely used by various utilities. The GEFCom2012 data used in this study covers a time period from January 2004 to June 2008 and contains hourly load data from 20 power stations, as well as hourly weather data from 11 weather stations across the New England area. In this study, the total load from 20 power stations and the average of 11 weather stations' data were used. A discussion of choosing the weather stations is presented in [22].

The vanilla benchmark model proposed by [21] was used to produce benchmark scores for GEFCom2012 [23]. This is a multiple linear regression model that uses calendar variables and weather data as covariates:

$$\begin{aligned} y_t =& \beta_0 + \beta_1 L_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t \\ &+ \beta_5 W_t H_t + f(T_t) + \varepsilon_t \end{aligned}$$

where y_t is load at time point t, L_t is trend term which has a linear relationship with time index. T_t is temperature, ε_t is normal error distributed as $N(0, \sigma_\varepsilon^2)$, and M_t, W_t and H_t are the dummy variables showing the month-of-the-year, day-of-the-week and hour-of-the-day. For the sake of simplicity, only one variable name is used to represent dummy variables, but one needs to know that there should be more covariates.

B. Attack Models

One way to influence load forecasting is to change the model parameters' estimation by altering the history load data used to build the model. Two widely used attack templates — random attack and ramping attack — are used here.

Random attack randomly chooses p percent of data and alters them using a positive scale parameter:

$$y_{t,a} = (1 + s\%) y_t \tag{1}$$

where $y_{t,a}$ is attacked data, y_t is true data, $s \sim N(\mu, \sigma^2)$.

Ramping attack consists of many single attack periods. One single attack period is determined by a starting attack point and a length:

$$y_{t,a} = [1 + \lambda_R (t - t_s)] y_t, \qquad t_s < t < \frac{t_s + t_e}{2} \tag{2}$$

$$y_{t,a} = (1 + \lambda_R (t_e - t)] y_t, \qquad \frac{t_s + t_e}{2} < t < t_e \tag{3}$$

where $y_{t,a}$ and y_t are attacked and real data, respectively, λ_R is a scale parameter, t_s and t_e are, respectively, the starting and ending time points of one single attack period, and $l = t_e - t_s$ is the length. Attackers will choose many of these length l periods and make the total attacked data points p percent of the whole data. Simulations can be done in the following fashion: divide whole data into $[N/l]$ pieces, where N is the length of whole data, then randomly choose $[pN/l]$ pieces to attack, and bracket "$[x]$" means the integer part of a number x.

III. Robust Regression

A. Classic Huber's Method

Huber's robust regression is a classic way to improve the robustness of a linear regression model against abnormal training data [24]. The basic idea is to use the weighted least square method to downweight potential abnormal data points iteratively until the estimation converges. One of the key parts of this process is the weight function used in every iteration, denoted as ψ function.

The algorithm is provided as follows:

1) Fit regular linear regression model using least square method, get the residual $r_t = y_t - \hat{y}_t$.
2) Calculate the robust estimation for error standard deviation σ_ε:

$$\hat{\sigma}_\varepsilon = \mathrm{MAR}/0.6745 \tag{4}$$

where MAR is the median of absolute residuals, and standardize the residual using $\hat{\sigma}_\varepsilon$:

$$e_t = r_t / \hat{\sigma}_\varepsilon \tag{5}$$

3) Plug in the ψ function using the scaled residual e_t:

$$\psi(e) = \begin{cases} 1 & |e| \leqslant k \\ k/|e| & |e| > k \end{cases} \tag{6}$$

to get the weights: $w_t = \psi(e_t)$. k in this equation is a fixed threshold. Use these weights to fit the same model using weighted least square method, and also get the residuals.
4) Repeat steps 2)–3) until the parameter estimates converge.

In order to maximize efficiency in the normal case, threshold k is usually chosen to be 1.345 such that, under the normal assumption, this algorithm will produce 95% efficiency but offer protection against outliers.

B. Modified Huber's Method

Based on Huber's idea, a modified version of the iteration algorithm is proposed here. Instead of first standardizing the residuals and then using a fixed threshold, a quantile of the residuals is used as a threshold in every iteration step. This way, the robust estimation of error stand deviation can be avoided, and, with the updated threshold, the percentage of data to be downweighted can be controlled at a fixed level determined by the quantile.

First, a fixed percentage $\tilde{p}$ needs to be set up for downweighting. This means we always downweight those data points whose residuals are among the largest $\tilde{p}$ percent. In every iteration step, the threshold $q_{\tilde{p}}$ is defined as the upper $\tilde{p}$ quantile of $|r_t|$, which is the absolute value of the residuals:

$$q_{\tilde{p}} = |r|_{([(1-\tilde{p})N])} \tag{7}$$

where N is the total number of observations. $|r|_{(i)}$ is the order statistics of absolute residuals such that $|r|_{(1)} \leqslant |r|_{(2)} \leqslant \cdots \leqslant |r|_{(N)}$. Then the new weight function is:

$$\psi_{\tilde{p}}(r) = \begin{cases} 1 & |r| \leqslant q_{\tilde{p}} \\ k/|r| & |r| > q_{\tilde{p}} \end{cases} \tag{8}$$

Table I
MAPE BASED ON PREDICTIONS OF VALIDATION DATASET.

Attack Type	LS	Huber's	Modified Huber's
random	0.195	0.165	0.078
ramping	0.233	0.100	0.067

C. Real Data Analysis

Here, two years of data were used as training data (2004 2005), and one year (2006) was used as validation data. Attacks were simulated and imposed onto the load data on the training dataset. After fitting the model, mean absolute percentage error (MAPE) was calculated on the validation dataset using Equation (9), which is a generally-used criteria for measuring prediction accuracy. The smaller the MAPE, the better the fit. The input weather variables used in the forecasting step were real weather data, instead of weather forecasting data.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{y_t} \tag{9}$$

This paper discusses the following three methods: least square, classic Huber's method and a modified version of Huber's method. The least square method is a standard fitting method used for linear regression models. It doesn't have robustness against attacks and was used as a benchmark method to show the robustness of the other two methods. Classic Huber's method and the modified Huber's method are both robust methods but with different weight functions. Actually, the least square method can also be viewed as a special case of the modified Huber's method. When $\tilde{p} = 0$, i.e, no data points gets downweighted, the modified Huber's method becomes the least square method. Both in this section and the simulation section, discussion and comparison are conducted among these three methods.

In order to give a visualized comparison of the different methods' performances, plots with load data (true and attacked) and estimated load under certain parameter settings are shown on Figure 1. The parameters were chosen to be attacked data proportion $p = 0.3$ for both attack templates; $s \sim N(50, 0)$ for random attack and $l = 50, \lambda_R = 0.05$ for ramping attack. For the fitting process, $k = 1.345$ was used in Huber's method and $\tilde{p} = 50\%$ for the modified Huber's method. In addition, the numeric criteria MAPE of these three methods were also displayed in Table I.

Both the classic and the modified versions of Huber's method show robustness against attacks when compared to the simpler least square method since they provide more accurate estimates and smaller MAPE on the validation dataset, which means more accurate predictions. Here, the reason that the modified Huber's method performs better than the classic Huber's method is that the modified Huber's method downweights 50% of data points (from the parameter setting). But, for the classic Huber's method, threshold $k = 1.345$ corresponds to the upper 9% quantiles of standard normal, so it will downweight about 18% of data points. This is not sufficient since the real percentage of attacked data is 30%.

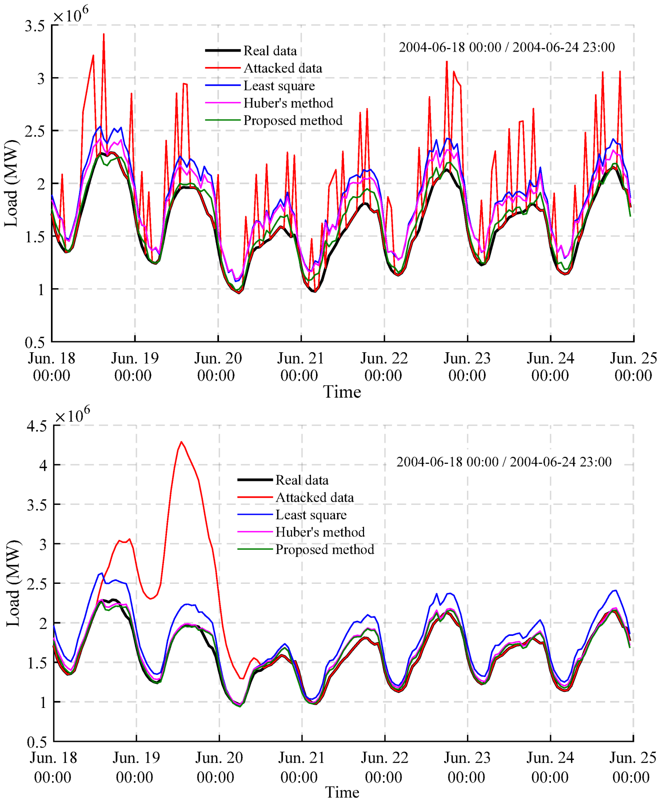

Figure 1. Load data fitting under a random attack (top) and ramping attack (bottom). Models were built based on whole training dataset (year 2004-2005), plot shows one week period within training dataset.

IV. SIMULATION STUDY

Since both the random and ramping attack templates display randomness, it's not convincing to compare them using only one scenario output. In this section, results from the simulation study are displayed. Just as before, data from the years 2004 to 2005 were used as training data, while the next whole year was used as validation data. Attacks were simulated and imposed onto the training dataset, and MAPE were calculated based on the validation dataset.

For each setting investigated, experiments need to be repeated 100 times and MAPE should be calculated as the average of these 100 repetitions' MAPE. This means that, for each setting, all parameters should be exactly the same, but only attacks are regenerated across every repetition. Since both random and ramping attacks randomly choose a proportion of the data to attack, attacks won't happen at the same location for every repetition even though the same parameter settings are used. In this fashion, randomness can be averaged out.

A. Robustness When Percentage of Attacked Data Changes

For both random and ramping attack templates, an attacker can choose a percentage of the data to attack, which is unknown to system operators. In order to investigate how

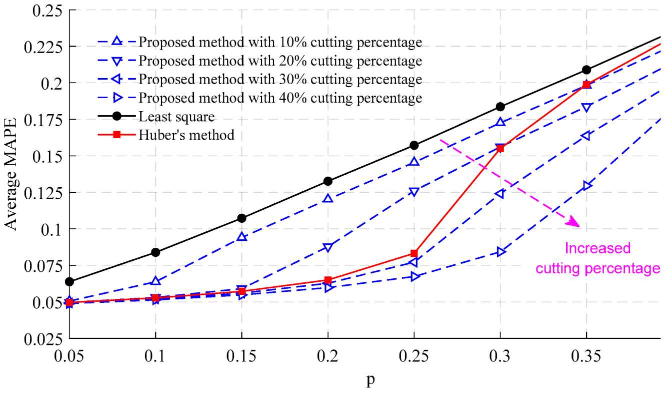

Figure 2. Average MAPE under random attack with $s \sim N(50, 0)$. The black line corresponds with simple least square methods, the red line shows the Huber's method result, and the blue lines are results from the modified Huber's method with different cutting quantiles.

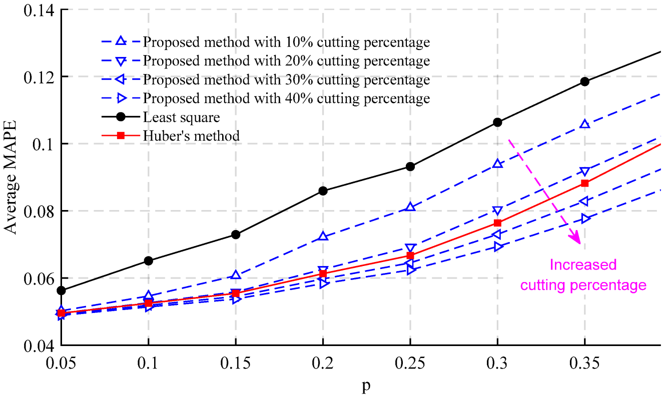

Figure 3. Average MAPE under ramp attack with $l = 50$ and $\lambda = 0.02$. The black line corresponds with the simple least square methods, the red line shows the Huber's method result, and the blue lines are results from modified Huber's method with different cutting quantiles.

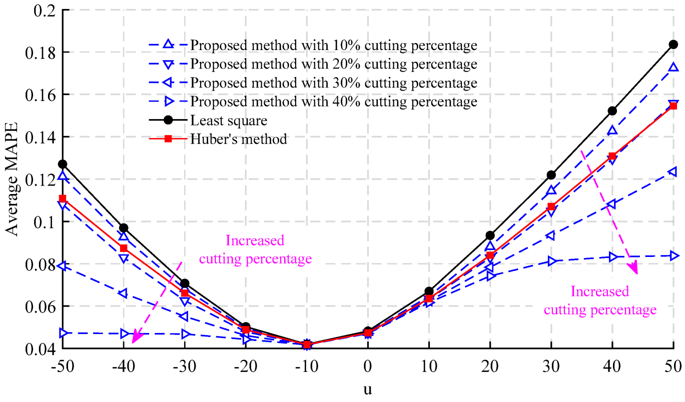

Figure 4. Average MAPE under random attack with $s \sim N(\mu, 0)$ and $p = 0.3$. The black line corresponds with the simple least square methods, the red line shows the Huber's method result, and the blue lines are the results from the modified Huber's method with different cutting quantiles.

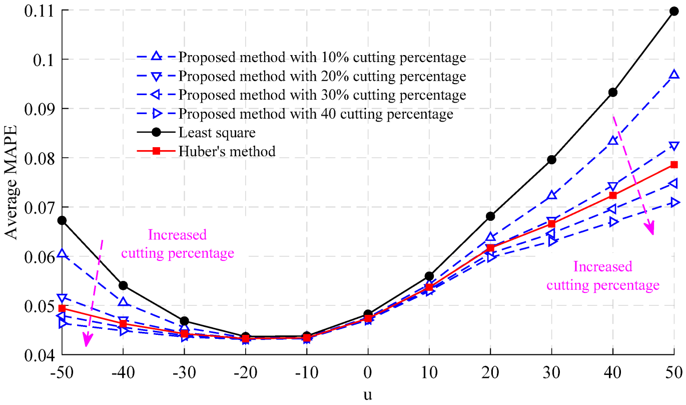

Figure 5. Average MAPE under ramp attack with $l = 50$ and $\lambda = \mu/2500$. The black line corresponds with the simple least square methods, the red line shows the Huber's method result, and the blue lines are the results from the modified Huber's method with different cutting quantiles.

the performance of different methods change along with the attacked percentage, p was set to increase from $5\% \sim 40\%$ with a step of 5%. $k = 1.345$ for Huber's method and $\tilde{p}$ (cutting percentage) changed from 0.1 to 0.4 by step 0.1 in the modified version of Huber's method.

For random attack, $s \sim N(50, 0)$. The length of a single ramping attack period is $l = 50$ and, in order to make the attack magnitude comparable, $\lambda_R = 0.02$. Results are shown in Figures 2 and 3:

These plots show that classic Huber's method has some robustness when the amount of attacked data is below some reasonable level. If the downweight percentage is 40%, the modified Huber's method will present better robustness even when the proportion of attacked data points increases. It is natural that these methods will be less robust if more data points were attacked. But the modified Huber method's robustness decrease is slower than that of the classic Huber's method. Here for both random and ramping attacks, they'll always make the data larger than the true value. So, as along as the attack exists, the estimated load will tend to be larger than the true value, even for the data points that were not attacked. Since the weight function will downweight those large residual data points using a decreasing function instead of simply throwing them away, it would be better if more data points were downweighted. That's why when p is small, the robust methods still perform better than the least square method.

B. Robustness When Attack Scale Changes

In this section, attack proportion is fixed at $p = 0.3$. The relationship between robustness and attack scale was investigated, that is s for random attack and λ_R for ramping attack (length for single ramping attack period was still set to be 50). In order to make these two attack templates comparable, $\lambda_R = \mu/2500$, $s \sim N(\mu, 0)$ were used. Results are shown in Figure 4 and 5:

The robustness of the classic and modified Huber's methods are more obvious when the attack scales become larger, since outliers will have larger influence. Besides, these two methods still show robustness when the attacks make the load data smaller than the true value (when μ is negative), and they won't perform worse when there is no attack ($\mu = 0$). The robustness of the classic and modified Huber's methods is less obvious than it is in the case when μ is positive. This is because when μ is negative, the difference between attacked data and real data is smaller than in the case of positive μ.

V. Discussion

Robust statistical methods offer a solution to cyberattack-resilient load forecasting in a linear regression setting. The classic Huber's weight defined with the normal distribution as a benchmark does not perform well when the distribution of the model error differs from the normal distribution by large. Our modified version of Huber's method downweights observations based on the percentile of the absolute value of the residuals, which makes the method insensitive to the distribution of the residuals caused by various attack models. The simulation study using the GEFCom2012 dataset suggests that the proposed method provides a significant improvement on forecasting accuracy over the classic Huber's method when the proportion of the attacked data is high with a large scale change.

The proposed method needs a prespecified percentile in the absolute value of the residuals beyond which the downweight starts. This is similar in all robust regression methods; the classic Huber's method needs to prespecify a robust estimate of the error term's standard deviation and uses a quantity corresponding to the 90th percentile of the standard normal distribution. For the classic Huber's method, a comparison among different versions of robust standard deviation estimator would be of interest. For the proposed modified Huber's approach, an adaptive approach which first estimates the proportion of the attacked data and then uses the estimated proportion to specify the approach is under investigation. The idea from adaptive least trimmed square method [25] can be adapted and extended to the maximum trimmed likelihood method so that it is applicable for generalized linear models to deal with non-linear situations.

One limitation of our work is that we only considered two attack templates: random attack and ramping attack. The performance of the proposed method and the classic Huber's method under other attack templates merits further investigation. Another limitation is that we only investigated the linear regression load forecasting model, which is used in long-term forecasting. Time series models which are better suited for short-term forecasting, such as the dynamic model [19] based on Tao's vanilla benchmark model, are worth studying.

References

[1] Y.-Y. Hsu and K.-L. Ho, "Fuzzy expert systems: An application to short-term load forecasting," in *Generation, Transmission and Distribution*, vol. 139, no. 6. IET, 1992, pp. 471–477.

[2] B.-J. Chen, M.-W. Chang *et al.*, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE transactions on power systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[3] M. Easley, L. Haney, J. Paul, K. Fowler, and H. Wu, "Deep neural networks for short-term load forecasting in ERCOT system," in *Texas Power and Energy Conference (TPEC)*. IEEE, 2018, pp. 1–6.

[4] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1535–1547, 1990.

[5] W. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Transactions on Power Apparatus and Systems*, no. 2, pp. 900–911, 1971.

[6] J.-F. Chen, W.-M. Wang, and C.-M. Huang, "Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting," *Electric Power Systems Research*, vol. 34, no. 3, pp. 187–196, 1995.

[7] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914 – 938, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207015001508

[8] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 730–737, 2017.

[9] J. W. Taylor and R. Buizza, "Using weather ensemble predictions in electricity demand forecasting," *International Journal of Forecasting*, vol. 19, no. 1, pp. 57–70, 2003.

[10] K. Siwek, S. Osowski, and R. Szupiluk, "Ensemble neural network approach for accurate load forecasting in a power system," *International Journal of Applied Mathematics and Computer Science*, vol. 19, no. 2, pp. 303–315, 2009.

[11] G. T. Ribeiro, M. C. Gritti, H. V. H. Ayala, V. C. Mariani, and L. dos Santos Coelho, "Short-term load forecasting using wavenet ensemble approaches," in *Neural Networks (IJCNN), 2016 International Joint Conference on*. IEEE, 2016, pp. 727–734.

[12] A. Jain and N. Shivakumar, "Power system tracking and dynamic state estimation," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES*. IEEE, 2009, pp. 1–8.

[13] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, 2011.

[14] J. Zhao, G. Zhang, M. La Scala, Z. Y. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1580–1590, 2017.

[15] B. G. Amidan, T. Ferryman, and S. K. Cooley, "Data outlier detection using the chebyshev theorem," *IEEE Aerospace Conference*, pp. 3814–3819, 2005.

[16] J. Yang and J. Stenzel, "Historical load curve correction for short-term load forecasting," *Internal Power Engineering Conference*, pp. 1–40, 01 2005.

[17] M. Yue, "An integrated anomaly detection method for load forecasting data under cyberattacks," *IEEE Power and Energy Society General Meeting*, pp. 1–5, 2017.

[18] J. Xie and T. Hong, "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1012–1016, 2016.

[19] J. Luo, T. Hong, and M. Yue, "Real-time anomaly detection for very short-term load forecasting," *Journal of Modern Power Systems and Clean Energy*, pp. 1–9, 2018.

[20] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," 2016.

[21] T. Hong, *Short Term Electric Load Forecasting*. North Carolina State University, 2010.

[22] T. Hong, P. Wang, and L. White, "Weather station selection for electric load forecasting," *International Journal of Forecasting*, vol. 31, no. 2, pp. 286–295, 2015.

[23] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," 2014.

[24] P. J. Huber, "Robust regression: Asymptotics, conjectures and Monte Carlo," *The Annals of Statistics*, pp. 799–821, 1973.

[25] R. Bacher, F. Chatelain, and O. Michel, "An adaptive robust regression method: Application to galaxy spectrum baseline estimation," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, March 2016, pp. 4423–4427.