Received: 12 March 2023 | Revised: 3 July 2023 | Accepted: 6 September 2023

IET Renewable Power Generation

DOI: 10.1049/rpg2.12853

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Adversarial false data injection attacks on deep learning-based short-term wind speed forecasting

Lei Yang[1] | Gaoshen Liang[2] | Yanrong Yang[3] | Jiaqi Ruan[4] | Peipei Yu[5] | Chao Yang[3]

[1]School of Foreign Languages and Business, Shenzhen Polytechnic, Shenzhen, China

[2]School of Information Technology, Beijing Normal University, Zhuhai, Zhuhai, China

[3]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China

[4]Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

[5]Department of Electrical and Computer Engineering, University of Macau, Macau, China

Correspondence
Gaoshen Liang, School of Information Technology, Beijing Normal University, Zhuhai 519087, Zhuhai, China.
Email: gaoshen.liang2000@gmail.com

Abstract

Developing accurate wind speed forecasting methods is indispensable to integrating wind energy into smart grids. However, current state-of-the-art wind speed forecasting methods are almost data-driven deep learning models, which may incur potential adversarial cyber-attacks. To this end, this paper proposes an adversarial false data injection attack tactic to investigate such a cyber threat. First, targeting the deep learning-based short-term wind speed forecasting model, an optimization model is constructed to obtain the optimally false data that should be injected into the forecasting model input so as to expand the prediction deviation as much as possible. Then, as the optimization model is non-differentiable, a particle swarm optimization-based method is developed to solve the optimization problem, in which the near-optimal solution is able to be explored, directing the false data that should be injected. At last, numerical studies of the proposed attack tactic are conducted on different-hour ahead wind speed forecasting models, revealing the feasibility and effectiveness.

1 | INTRODUCTION

In recent years, climate change has drawn special attention due to frequently witnessed extreme natural disasters [1, 2]. This has sparked a strong collective aspiration among humans to create a sustainable future [3]. The combustion of fossil fuels, which dates back to the industrial revolution, has been identified as the primary source of greenhouse gas emissions. These emissions are recognized as the principal cause of global warming and resultant climate change. To achieve a low-carbon energy mix, renewable energy sources have garnered substantial attention in recent years [4]. Since renewable energy does not generate emissions or contribute to climate change, it is an excellent alternative to fossil fuels. The adoption of renewable energy, by decreasing the usage of fossil fuels [5], assumes a crucial role in reducing the impact of climate change and ensuring a sustainable future for future generations [6].

For different types of renewable energy resources, the use of wind energy has proliferated in recent years, and many countries have started investing heavily in this form of energy production [7]. However, unlike solar energy with periodic patterns, wind energy exhibits strong uncertainties due to the weather system's chaotic nature. The weather-dependent characteristic of wind energy and its strong variability and intermittence [8] could bring great challenges to the smart grid operation, including voltage and frequency instability [3, 9]. The instability of wind energy could cause issues in the energy supply-demand balance, leading to a potential failure of the grid's control system [10]. At the extreme end, smart grids may experience system collapse or even blackouts [11].

To address the challenges posed by wind energy's variable nature, various short-term wind speed forecasting (WSF) methods have been developed [12]. With an accurate wind speed prediction, wind farm decisions and controls can be properly formulated and implemented so that the smart grid stability can be influenced by the wind power generated as small as possible [13]. This would ensure optimal energy production and utilization, as well as grid stability and resilience, thus guaranteeing uninterrupted energy supply to consumers. While traditional forecasting methods may not be able to capture the complex patterns and relationships present in the voluminous wind speed data, deep learning (DL) technology has been shown to be highly effective in establishing WSF models [14]. DL methods can learn nonlinear and complex features from data and address data uncertainty, making them highly suitable for WSF [15]. For example, the deep belief network was utilized for deterministic and probabilistic WSF [16]. The long short-term memory neural network was adopted to improve the prediction performance of wind speed [17]. The interval deep generative neural network was designed to capture unsupervised temporal features from wind speed data [18].

The DL-based WSF model may be vulnerable to cyber threats. The introduction of information communication technology (ICT) into the smart grid has exposed various cyber assets to malicious adversaries [19], who may manipulate the forecasting input data during transmission to impair the model's performance. Because the DL model is a black-box model, operators receive the prediction result by directly inputting the data without any intermediate examination. As a result, the operator may make improper decisions based on the predicted results from the false input, which can lead to severe consequences for the smart grid.

Several studies have examined adversarial attack techniques in learning-based applications within the smart grid domain. Ref. [20] proposed two algorithms to generate adversarial input data to evaluate the security issues of load forecasting procedures in power system operations. Ref. [21] designed a methodological framework to explore the vulnerability of machine learning-based inertia forecasting models, with a special focus on data integrity attacks that are able to significantly increase the system operation cost. Ref. [22] proposed a generative adversarial network-based adversarial data injection attack method against data-drive stability assessment in power systems. Although these references have studied adversarial attacks on different learning-based applications in smart grids, it is essential to highlight a key distinction between the adversarial attack on these applications and on WSF. Unlike other applications, the WSF model does not necessitate the adversary's knowledge of the operational state of smart grids for launching adversarial attacks. This unique characteristic exposes the potential for significant prediction errors that can detrimentally impact smart grid operations, in contrast to other applications that do require such knowledge.

Despite the potential risks, there has been insufficient research into the cyber threat on the DL-based WSF model in the existing literature. Given the significance of the smart grid to modern society, it is critical to explore and address these vulnerabilities to ensure the resilience of the smart grid. With the rapid development of cyber technologies, it is essential to continuously evaluate and enhance the security of the DL-based WSF model against potential cyber threats. Further studies in this area are urgently needed to develop effective countermeasures that can safeguard the smart grid from cyber attacks.

In order to examine the resilience of well-trained DL models for short-term WSF, this paper suggests a novel method of attacking the DL model by injecting false data, known as adversarial false data injection attacks, which is implemented in the practical application stage. The primary objective of this paper is to demonstrate the weaknesses of DL models in WSF when exposed to this type of attack. As a result of this research, several significant contributions have been made, which are outlined below.

1. An adversarial false data injection attack method is proposed targeting the DL-based short-term WSF model, in which an optimization model is constructed to obtain the optimally false data that should be injected into the model input data so as to expand the prediction error as much as possible.
2. A particle swarm optimization (PSO)-based method is designed to solve the proposed optimization model. As conventional solvers are unable to solve the optimization model containing the DL model, which is sealed and thus non-differentiable, the PSO algorithm is a promising alternative and effective to obtain the solution of the proposed optimization model.
3. The attack performance and effect are assessed comprehensively and extensively on the realistic dataset. The simulation results suggest that only needing to modify small values in the input data, it is able to output a large deviated prediction, which jeopardizes the resilience of smart grids with high renewable energy penetration.

The remainder of the paper is organized as follows. Section 2 presents the proposed adversarial false data injection attack strategy, including introducing the DL-based short-term WSF model and modeling the adversarial attack method. Section 3 elaborates on the PSO-based optimization-solving scheme and gives the overall framework. Section 4 demonstrates and discusses the case studies on different hour-ahead WSF models. At last, Section 5 summarizes the paper.

2 | THE PROPOSED ADVERSARIAL FALSE DATA INJECTION ATTACK METHOD

2.1 | Deep learning-based short-term wind speed forecasting model

The precise forecasting of wind speed is vital for ensuring the efficient operation of smart grids [23]. Although weather systems are inherently unpredictable, it is feasible to generate accurate short-term predictions, covering a few hours or days, for wind speed. This capability is crucial for maximizing the production of wind energy and incorporating it into the grid, as well as achieving a more efficient balance between energy supply and demand. Ultimately, this results in reduced operational expenses and enhanced energy efficiency.

Artificial intelligence (AI) has become an essential tool in various industries, including the energy sector. In particular, deep learning (DL), a subset of AI, has shown impressive potential in predicting short-term wind speeds, which is crucial in ensuring efficient and reliable wind energy production. One of the reasons why DL is effective in wind speed forecasting (WSF) is its ability to extract complex and non-linear features from data, resulting in higher accuracy and lower errors [24].

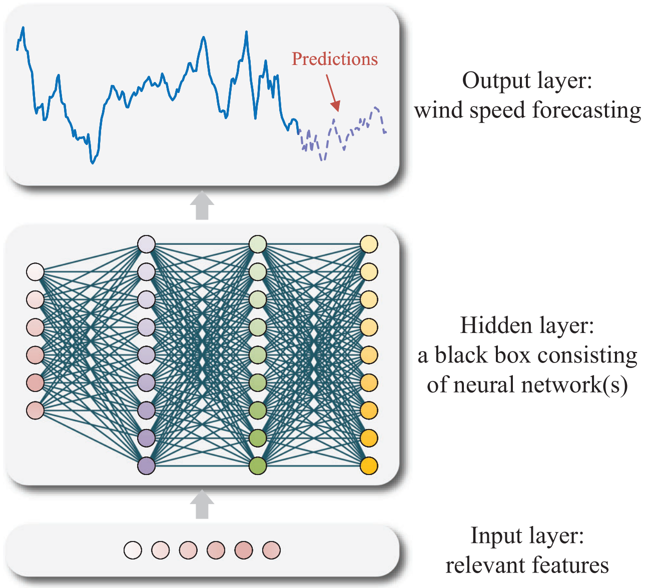

FIGURE 1 Diagram of the deep learning-based wind speed forecasting model.

A typical DL-based WSF model consists of three components: the input layer, hidden layer, and output layer, as shown in Figure 1. The input layer plays a crucial role in incorporating essential features such as historical wind speed data, temperature, pressure, humidity, and wind direction that contribute to improved accuracy in WSF. The hidden layer, on the other hand, refers to the neural network layers between the input layer and the output layer. Its primary purpose is to learn intricate relationships between the inputs and outputs. These hidden layers are called "hidden" because their outputs are not directly observable [25]. Various hidden layer architectures are used in WSF, including the multi-layer perceptron, deep belief network, and long short-term memory network, among others.

Lastly, the output layer takes the processed input from the hidden layer and generates a single value or multiple values that represent the forecasted wind speed in the short-term future. Common DL-based short-term WSF models include 1-h, 2-h, 3-h, 4-h, 8-h, 12-h, and 24-h ahead forecasting. The model can be trained by using the mean absolute error (MAE) or the mean square error (MAE) as the loss function, which measures the difference between the actual and predicted wind speeds, as follows,

$$\mathcal{L}_\theta = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}^{(i)} - y^{(i)}| \tag{1}$$

$$\mathcal{L}_\theta = \frac{1}{N}\sum_{i=1}^{N} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 \tag{2}$$

where $\mathcal{L}_\theta$ represents the loss function under the DL parameters θ; N is the total number of samples; $y^{(i)}$ and $\hat{y}^{(i)}$ are the output true value and prediction of the ith sample, respectively.

Once the DL model is established and well-trained, it can be used for prediction. Mathematically, a DL-based h-hours ahead wind speed point forecasting model can be formulated as follows,

$$\hat{y}_{t+h} = \mathcal{F}(\mathcal{X}_t) \tag{3}$$

where $\hat{y}_{t+h}$ is the predicted wind speed after h hours at time t; $\mathcal{X}_t$ represents a set of model input including features available at time t; $\mathcal{F}(\cdot)$ denotes the well-trained DL model for WSF.

2.2 | Adversarial false data injection attack model

The application of ICT facilitates convenient and efficient interaction between different sectors in the smart grid [26]. However, this increased connectivity also exposes the smart grid's digitized assets to cyber threats [27]. In the context of the WSF model, meteorological data is used as input. To obtain the meteorological data, the outer weather forecast system due to its professionality is integrated with the WSF model. In general, an interface must be requisitioned for transmitting the meteorological data from the weather forecast system to the WSF model. However, during the data transmission process, a capable adversary may compromise the meteorological data by injecting false data, which could adversely affect the performance of the WSF model.

To investigate this problem, this paper proposes an adversarial false data injection attack method that aims to worsen the prediction performance of the WSF model as large as possible. This attack method is achieved by constructing an optimization model to suggest the optimal false data that should be injected into the WSF model's input, as follows. Equation (4) is the objective function that maximizes the prediction error. Equations (5) and (6) are constraints to obtain the predicted wind speed under normal and attacked scenarios. Equation (7) is a constraint to quantify the capability for the attack, in which the adversary can inject false data only within certain limits.

$$\Delta\mathcal{X}_t^{a*} = \arg\max_{\Delta\mathcal{X}_t^a} \quad \left|\hat{y}_{t+h} - \hat{y}_{t+h}^a\right| \tag{4}$$

$$\hat{y}_{t+h} = \mathcal{F}(\mathcal{X}_t) \tag{5}$$

$$\hat{y}_{t+h}^a = \mathcal{F}(\mathcal{X}_t + \Delta\mathcal{X}_t^a) \tag{6}$$

$$\left|\Delta\mathcal{X}_t^a\right| \le \Delta\bar{\mathcal{X}}_t^a \tag{7}$$

where $\Delta\mathcal{X}_t^{a*}$ represents the optimal false data for the attack; $\Delta\mathcal{X}_t^a$ and $\hat{y}_{t+h}^a$ are the false data injected and the wind speed predicted under the false data, respectively; $\Delta\bar{\mathcal{X}}_t^a$ is an upper bound for the false injection $\Delta\mathcal{X}_t^a$.

Ideally, a bigger upper bound value for the modification in input data means that the adversary is able to cause a larger prediction deviation. However, although the adversary has the ability to inject false data within the limits, any increased attack resource (i.e. the value of injected false data) imposes extra attack costs (i.e. the possibility of being detected). To reduce the attack cost, the previous objective function (4) can be improved by introducing the minimization of attack resources used in the meantime, as follows,

$$\Delta\mathcal{X}_t^{a*} = \arg\max_{\Delta\mathcal{X}_t^a} \quad \left|\hat{y}_{t+h} - \hat{y}^a_{t+h}\right| - \alpha||\Delta\mathcal{X}_t^a||_1 \tag{8}$$

where α is a weight to balance the attack resources being used.

On the other hand, sometimes there are anomaly detection algorithms to check with the transmitted input data, which are based on parameter statistics such as the 3 sigma criterion and the box plot principle [28]. Once the falsified data is unable to bypass these algorithms and is detected, the launched cyberattack will be unsuccessful and ineffective. To further avoid such detection, Equation (7) can be enhanced by reasonable scaling, as follows,

$$(1-\beta)\mathcal{X}_t \leq \Delta\mathcal{X}_t^a \leq (1+\beta)\mathcal{X}_t \tag{9}$$

where β is a positive value for scaling. For example, β could be 0.05 or 0.1 for achieving a 5% or 10% maximum modification degree according to the initial value.

To accomplish the proposed attack method, the adversary has three management strategies to compromise the meteorological data: pre-transmission, during-transmission, and post-transmission. The pre-transmission strategy involves the manipulation or concealment of meteorological data using predetermined values within the external weather forecast system [29]. Consequently, the transmission interface is supplied with false data prior to the commencement of data transmission. The during-transmission strategy encompasses actions taken to compromise meteorological data while it is being transmitted [30]. This may include activities such as eavesdropping, interception, and tampering with data in the communication channel. The post-transmission strategy aims to modify the memory block where transmitted data is stored [31]. After the data has been transmitted from the external weather forecast system, it is initially retained in a device and subsequently utilized as input for the WSF model. Consequently, the adversary must erase the true meteorological data within the memory block and substitute it with fabricated data, thereby ensuring the compromise of the WSF input.

It is noteworthy that the proposed adversarial false data injection attack method is based on the white-box scenario, wherein the adversary possesses complete knowledge of the operator-own WSF model. However, acquiring such detailed information about the operator-own WSF model presents considerable challenges in real-world settings. As an alternative, the adversary may consider employing a black-box attack scenario. In this approach, the adversary first specifies the wind power plant and gathers relevant data. Then, based on the collected data, a set of appropriate DL models is selected and trained specifically for the WSF purpose. Following the training phase of all selected WSF models, the best-performing model is determined, serving as a surrogate model for the subsequent attack optimization problem. This surrogate model will replace the operator-own WSF model, facilitating the generation of adversarial false data.

3 | PARTICLE SWARM OPTIMIZATION-BASED ATTACK MODEL SOLVING SCHEME

The proposed attack optimization model is a complex model that includes the DL model. However, conventional optimization solvers are unavailable for this model, making it difficult to obtain an optimized solution. The absence of conventional optimization solvers stems from the complexity of the model, which requires derivative information that is not easily obtainable. To overcome this difficulty, a PSO-based method is designed to fulfill the solution of the attack optimization model.

The PSO algorithm is an optimization technique that imitates the social behaviour of bird flocks or fish schools. Unlike gradient-based optimization algorithms, the PSO algorithm does not require derivative information, which makes it suitable for solving the proposed attack optimization problem where the gradient of the objective function is difficult to obtain. Instead, PSO utilizes a population of particles that iteratively move in search of the optimal solution to the problem. Each particle's movement is guided by its own best solution and the best solution found by the entire population. This approach enables the PSO algorithm to converge towards the global optimal solution of the problem efficiently.

PSO is an ideal method to solve the proposed attack optimization problem since the gradient of the objective function is challenging to obtain, and multiple local minima may exist in the problem. The PSO algorithm's capability to find the global optimal solution makes it well-suited for the attack optimization problem, ensuring that the best possible solution is achieved for the model. The proposed PSO-based method provides a robust and efficient solution for the attack optimization model, which would have been unachievable using conventional optimization solvers. Basically, for the initial meteorological data $\mathcal{X}_t$, the PSO works as follows,

$$\begin{aligned} v_{p,k}^{(i+1)} &= v_{p,k}^{(i)} + c_1 r_1 \left(x_{p,k}^{\text{best},(i)} - x_{p,k}^{(i)}\right) \\ &\quad + c_2 r_2 \left(x_k^{\text{best},(i)} - x_{p,k}^{(i)}\right) \end{aligned} \tag{10}$$

$$x_{p,k}^{(i+1)} = x_{p,k}^{(i)} + v_{p,k}^{(i+1)} \tag{11}$$

where the superscript (i) denotes the current iteration; the subscript $p = 1, \ldots, N$ represents the pth particle for a total of N particles; the subscript k indicates the kth feature of the model input $\mathcal{X}_t$; v and x are the current speed and position of the particle, respectively; c_1 and c_2 are learning factors; r_1 and r_2 are random values generated from the interval (0,1); $x_{p,k}^{\text{best},(i)}$ denotes the best value of the kth feature for the pth particle and $x_k^{\text{best},(i)}$ is the global best value crossing all particles.

Based on Equations (10) and (11), the model input for pth particle in the ith iteration (denoted as $\mathcal{X}_p^{(i)}$) can be updated. Initially, $\mathcal{X}_p^{(0)} = \mathcal{X}_t$, and the speeds are set with small values for updating. However, if the position $x_{p,k}^{(i+1)}$ is updated out-of-bound, its value should be set as the corresponding boundary, as follows,

$$x_{p,k}^{(i+1)} = \begin{cases} x_k^{\text{upper}}, & x_{p,k}^{(i+1)} \ge x_k^{\text{upper}} \\ x_k^{\text{lower}}, & x_{p,k}^{(i+1)} \le x_k^{\text{lower}} \end{cases} \tag{12}$$

where x_k^{upper} and x_k^{lower} are the upper and lower bounds for the kth feature of $\mathcal{X}_t$ according to the optimization constraint (7) or (9).

The comparison of true and modified model inputs will determine the updates of the particle optimum $\mathcal{X}_p^{\text{best},(i)}$ and global optimum $\mathcal{X}^{\text{best},(i)}$, as follows,

$$\mathcal{X}_p^{\text{best},(i+1)} = \begin{cases} \mathcal{X}_t^{(i+1)}, & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{(i+1)}\right)| > \\ & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{\text{best},(i)}\right)| \\ \mathcal{X}_p^{\text{best},(i)}, & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{(i+1)}\right)| \le \\ & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{\text{best},(i)}\right)| \end{cases} \tag{13}$$

$$\mathcal{X}^{\text{best},(i+1)} = \begin{cases} \mathcal{X}_t^{(i+1)}, & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{(i+1)}\right)| > \\ & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}^{\text{best},(i)}\right)| \\ \mathcal{X}^{\text{best},(i)}, & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}_p^{(i+1)}\right)| \le \\ & |\mathcal{F}(\mathcal{X}_t) - \mathcal{F}\left(\mathcal{X}^{\text{best},(i)}\right)| \end{cases} \tag{14}$$

Initially, $\mathcal{X}^{\text{best},(0)} = \mathcal{X}_t$ and $\mathcal{X}_p^{\text{best},(0)} = \mathcal{X}_t$. After the finish of Equations (13) and (14), it goes to the next iteration. The iteration process stops if the predetermined maximum number for iterations (denoted as τ) reaches or there is no global optimum update for $\mathcal{X}^{\text{best},(i)}$ in a default number of past iterations. Therefore, the optimal injected value for the attack optimization model will be obtained as follows,

$$\Delta\mathcal{X}_t^{a*} = \mathcal{X}^{\text{best}} - \mathcal{X}_t \tag{15}$$

where $\mathcal{X}^{\text{best}}$ is the final updated value for $\mathcal{X}^{\text{best},(i)}$. A complete process for the PSO-based attack optimization solving method is illustrated in Figure 2.

4 | NUMERICAL STUDIES

In this section, the performance of the proposed adversarial false data injection attack has been tested on DL-based different hours ahead WSF models.

4.1 | Set up

The WSF model selected is a state-of-the-art DL model in which the hidden part consists of two LSTM layers, one attention layer, and three fully-connected layers for feature extraction and nonlinear learning. The LSTM network is efficient in dealing with time series data (the model input containing available wind speeds at the past 4 h); the attention mechanism is able further to extract the most valuable features for deep learning; the fully-connected layer is used for feature extraction and dimension conversion [32]. The mean square error loss function and the adaptive moment estimation solver are adopted during the model training process. The dataset from Kaggle is employed for WSF [33]. Specifically, the dataset is split into training and test sets with the 8:2 ratio. The simulation environments are simulated by Python 3.7 on a PC with Intel(R) Core(TM) i9-10900 CPU @ 2.80 GHz, 64.0 GB RAM, and a GPU of NVIDIA GeForce RTX 2060.

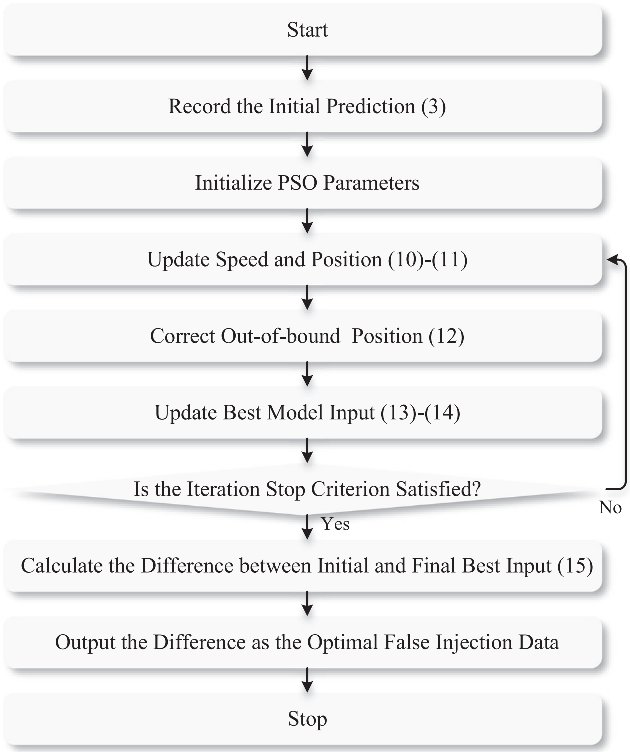
Start
Record the Initial Prediction (3)
Initialize PSO Parameters
Update Speed and Position (10)-(11)
Correct Out-of-bound Position (12)
Update Best Model Input (13)-(14)
Is the Iteration Stop Criterion Satisfied?
No
Yes
Calculate the Difference between Initial and Final Best Input (15)
Output the Difference as the Optimal False Injection Data
Stop

FIGURE 2 Diagram of the PSO-based attack optimization solving method.

4.2 | Case studies

As wind resources vary dramatically in seasons, the proposed attack method is launched to different seasons for assessment. First, the statistics of performances of different-hour ahead wind forecasting models in seasons are shown in Table 1, in which the mean absolute error (MEA) and root mean square error (RMSE) metrics are used to quantify the performance. Obviously, all three forecasting scenarios have low errors in all seasons. The reason is that the WSF model employs the advanced attention mechanism, wherein the most temporal correlations can be identified and extracted. The attention mechanism enables the WSF model to focus on specific parts of the input data that are more relevant to the output and use them

TABLE 1 Performance statistics of different-hour ahead wind speed forecasting models.

Season	Error (m/s)	1-h ahead	2-h ahead	3-h ahead
Spring	MAE	0.2075	0.2608	0.2638
	RMSE	0.3685	0.4085	0.4271
Summer	MAE	0.2035	0.2671	0.2676
	RMSE	0.3622	0.4325	0.4281
Autumn	MAE	0.2010	0.2610	0.2581
	RMSE	0.3559	0.4061	0.4119
Winter	MAE	0.1980	0.2574	0.2648
	RMSE	0.3371	0.3951	0.4115

to make more accurate predictions. Moreover, the model performance improves slightly as the prediction time horizon becomes shorter. It is rational because the weather system is chaotic. The wind speed is highly sensitive to various meteorological conditions, and small changes in these conditions can lead to significant differences in the future. Therefore, the wind speed variation in the far future would be more uncertain compared with the near future.

The attack effects are presented in Figures 3–5, wherein random two continuous days are selected for visualization. In the attack optimization problem, the parameter α in the objective function (8) is set as 0.1, and the parameter β in the constraint (9) is set as 0.05. Clearly, from these figures, the normal prediction almost overlaps with its true value. However, when the proposed adversarial false data injection attack is involved, the prediction after being injected with false data would deviate from its true value to a great extent. Furthermore, such prediction deviations are more noticeable at larger wind speeds. The reason lies in the constraint (9), in which the injected false data in the model input can reach 5% of the true value. If the initial input is able to cause a large prediction value in the wind speed, it means that the latest available wind speed in the model input would not be small. By injecting false data into these large input values in a collaborative manner, the compromised input results in a large change in the prediction. Consequently, the prediction deviation caused by attacking prediction in small wind speed values is not as significant as that caused by large values.

Interestingly, in a few attack cases, the predictions between normal and attacked are very close. Since the PSO algorithm does not resort on the gradient principle for optimization, sometimes it is sensitive to the initialization of its parameters. The initial population of particles and their positions can significantly impact the algorithm's ability to find an optimal solution. In the context of adversarial attacks, this means that the PSO algorithm may not always be able to find the optimal false data to cause a large attack effect, even if such false data exists. This is because the algorithm may get stuck in a local minimum or fail to explore the search space effectively due to poor initialization.

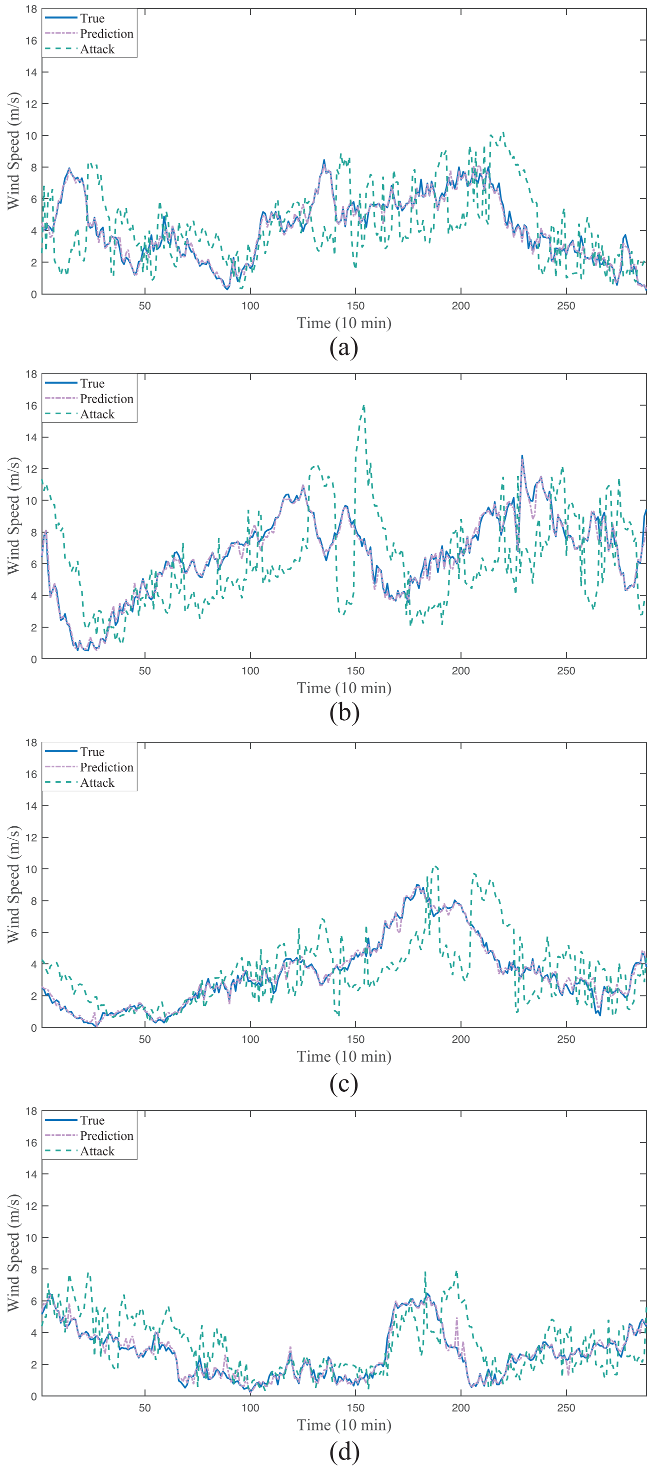

FIGURE 3 Comparison of normal and attacked scenarios for 1-h ahead wind speed forecasting in (a) spring, (b) summer, (c) autumn, and (d) winter.

To quantify the attack effects on different hour-ahead WSF models, statistics of the average prediction deviation (percentage change) are provided in Table 2. From these statistics, it is found that after launching the proposed attack, the average deviation percentages incurred for all models are quite large. Moreover, there exists a trend that if the average wind speed across these attack cases is slow, the average deviation percentage will be large. The reason is that the slow wind speed means a small denominator for the deviation percentage. Hence, even though the deviated wind speed after launching the attack is

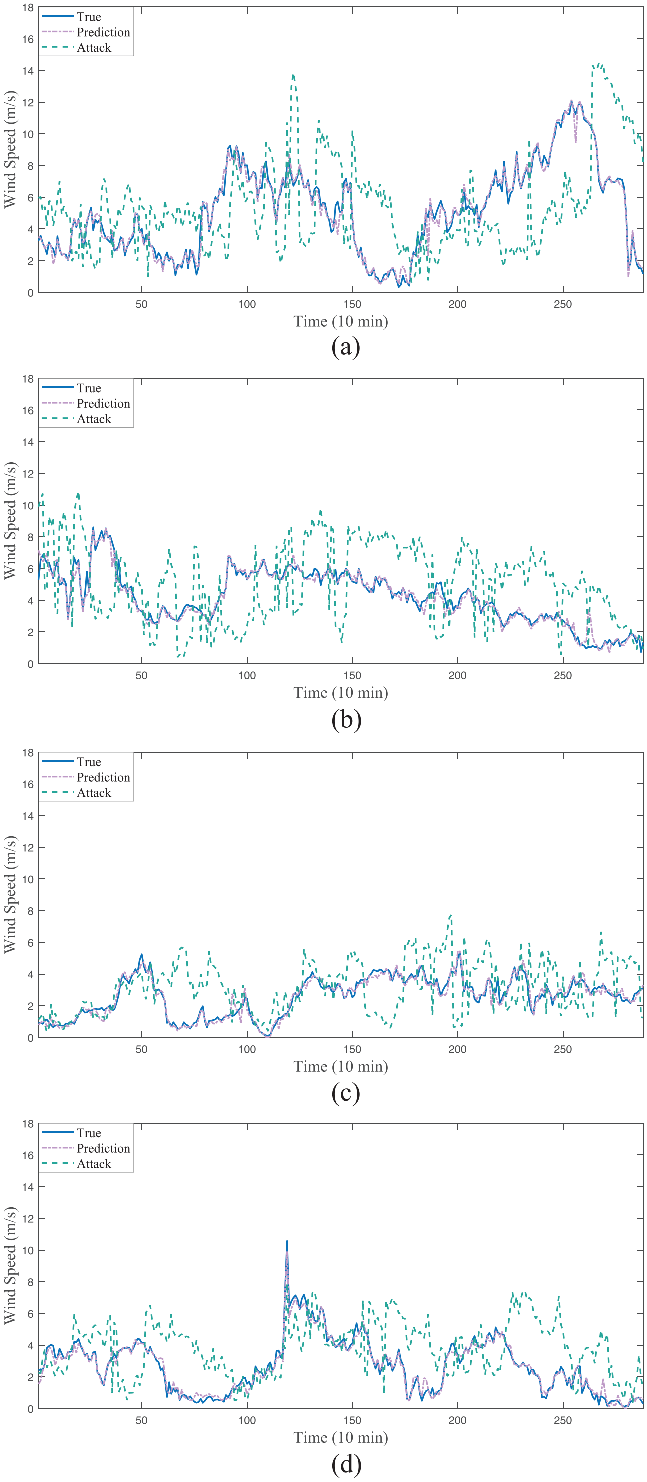

FIGURE 4 Comparison of normal and attacked scenarios for 2-h ahead wind speed forecasting in (a) spring, (b) summer, (c) autumn, and (d) winter.

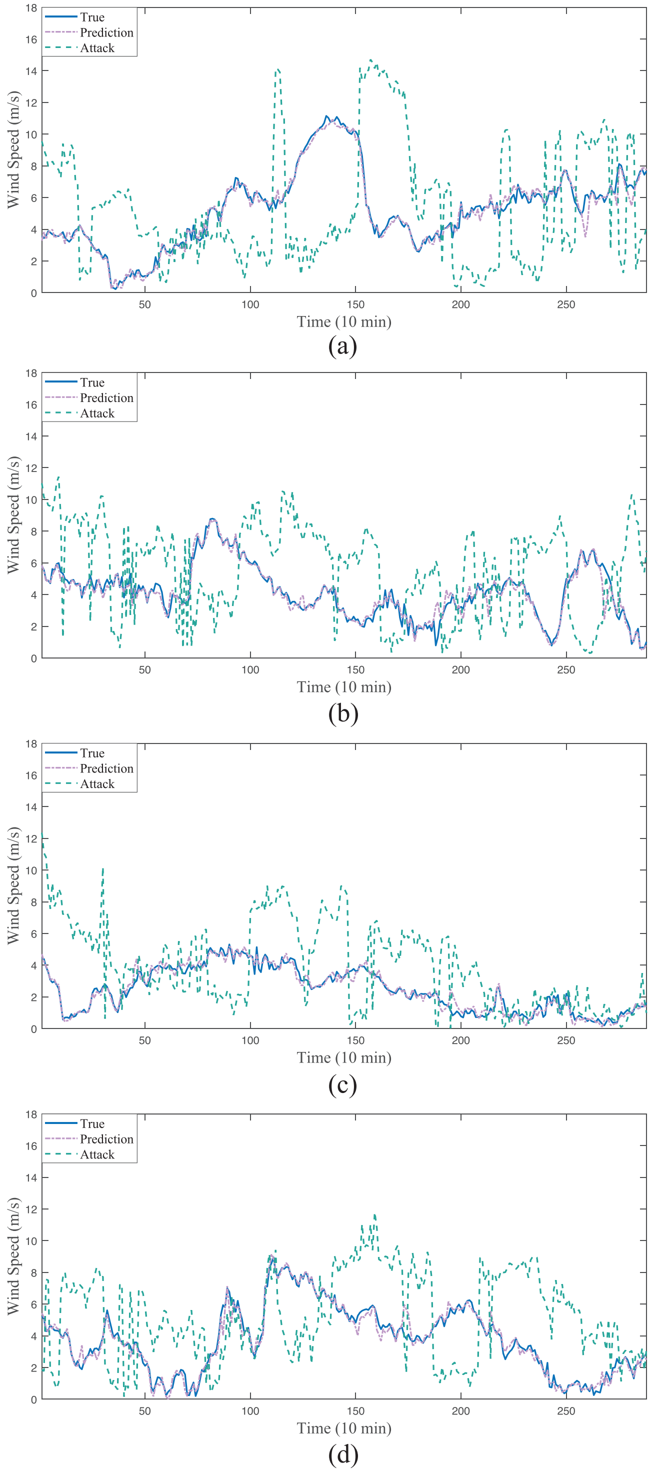

FIGURE 5 Comparison of normal and attacked scenarios for 3-h ahead wind speed forecasting in (a) spring, (b) summer, (c) autumn, and (d) winter.

larger under the initially fast wind speed (which means a large denominator), the resulting deviation percentage would not be such significant compared with the initially slow wind speed.

The average wind speed and the average wind speed deviation incurred by the attack are compared in Table 3. The table suggests that in the same WSF model scenario, there is a correlation between wind speed and wind speed deviation, with slower wind speeds resulting in smaller deviations and faster wind speeds leading to larger deviations. This correlation may reflect the fact that the proposed attack method has a more significant impact on periods of high wind resource availability when wind power can be generated at its maximum potential. Further-

TABLE 2 Statistics of average percentage prediction deviation in different-hour ahead wind speed forecasting models.

Model	Average prediction deviation			
	Spring	Summer	Autumn	Winter
1-h ahead	68.53%	62.61%	98.14%	69.03%
2-h ahead	105.00%	79.84%	121.31%	180.63%
3-h ahead	130.03%	123.64%	139.73%	265.25%

TABLE 3 Statistics of average and deviation of wind speed in different-hour ahead wind speed forecasting models.

Season	Wind speed (m/s)	1-h ahead	2-h ahead	3-h ahead
Spring	Average	4.1809	5.1851	5.2460
	Deviation	2.2274	3.4218	4.3300
Summer	Average	6.5428	4.2042	4.1485
	Deviation	3.0229	2.7542	3.6501
Autumn	Average	3.3890	2.5935	2.4503
	Deviation	1.7249	1.6021	2.4365
Winter	Average	2.6851	2.7443	3.8132
	Deviation	1.3599	2.0168	3.3807

TABLE 4 Statistics of average attack resource utilization.

Model	Average attack resource utilization		
	$\beta = 1\%$	$\beta = 5\%$	$\beta = 10\%$
1-h ahead	0.97%	4.33%	7.87%
2-h ahead	0.97%	4.38%	7.62%
3-h ahead	0.98%	4.32%	7.58%

more, the wind power is proportional to the cube of wind speed, which means that periods of fast wind speeds are more vulnerable to the proposed attack strategy, as even a small change in wind speed can significantly impact the amount of power generated. Overall, these findings suggest that the proposed attack method could significantly impact the efficiency and reliability of wind power generation, particularly during periods of high wind resource availability.

To further investigate the influence of attack resource utilization and its subsequent impact, we conducted a comparative analysis of different values for the parameter β, including 0.01, 0.05, and 0.10. These values correspond to 1%, 5%, and 10% of input data deviation ranges, respectively. For each WSF model, a random selection of 100 samples from the test set is used for assessment. Their average attack resource utilization and impact on prediction deviation are showcased in Tables 4 and 5. Notably, the employed attack resources remain below their maximum limits, indicating the practicality of the proposed attack resource minimization outlined in Equation (8). Specifically, a smaller β value leads to increased attack resource utilization, whereas looser restrictions on attack deviation ranges (i.e. larger β values) result in greater savings in attack resource utilization.

TABLE 5 Statistics of attack impact on average prediction deviation.

Model	Average prediction deviation		
	$\beta = 1\%$	$\beta = 5\%$	$\beta = 10\%$
1-h ahead	12.82%	66.47%	98.49%
2-h ahead	15.90%	75.92%	136.61%
3-h ahead	35.77%	113.67%	188.95%

The impact of the β value on prediction deviation is also evident in Table 5. A smaller β value corresponds to a reduced prediction deviation. When the adversary possesses the capability to inject larger falsified data, the proposed attack demonstrates the potential to induce substantial deviations in the WSF model. Furthermore, different WSF models exhibit varying degrees of resilience against the proposed attack. WSF models with shorter prediction times tend to be slightly less affected by the attack, indicating a higher resilience, while those with longer prediction times exhibit the opposite trend. Nevertheless, even under these circumstances, the proposed attack reveals the inherent vulnerability of WSF models. A mere 1% restriction on the input data deviation range can lead to prediction deviations exceeding 10%. Notably, for the 3-h ahead WSF model, the attack can result in prediction deviations surpassing 30%.

To summarize, the proposed attack method is evaluated comprehensively and extensively on different WSF models. The results show that the proposed attack method significantly impacts the performance of WSF models, especially during periods of high wind speeds where prediction deviations are more significant. This vulnerability could be further exploited by malicious attackers to cause more catastrophic impacts on the integration of wind energy into smart grids.

5 | CONCLUSION

The use of data-driven methods has proven to be effective in developing accurate short-term water supply forecasting (WSF) models. However, the current advanced WSF models primarily rely on deep learning (DL) technology, which makes them vulnerable to adversarial attacks due to their black-box nature. In this paper, we investigate the susceptibility of DL-based short-term WSF models to adversarial attacks. To do so, we propose a novel adversarial false data injection attack method that aims to degrade the forecasting performance of the model. This is achieved by building an optimization model that suggests optimal false data to be injected into the model input. To effectively optimize the attack model, we design a particle swarm optimization (PSO) based method that can explore the optimal solution, overcoming the non-differentiable problem in the attack optimization model. We then conduct comprehensive experiments of the proposed attack method on DL-based 1-h, 2-h, and 3-h ahead WSF models. The numerical results demonstrate the severe attack effects, where a small strength but highly coordinated attack may bring an extremely large degree of prediction deviation for initially well-performed WSF models. In summary, this study highlights the vulnerability of DL-based short-term WSF models to adversarial attacks and proposes a novel attack method that can effectively degrade the forecasting performance of these models. Moreover, it calls for more attention to study such vulnerability and developresilient countermeasures.

AUTHOR CONTRIBUTIONS

Yang Lei: Conceptualization, investigation, methodology, writing - original draft. Gaoshen Liang: Data curation, resources, software, writing - review and editing. Yanrong Yang: Resources, software, validation. Jiaqi Ruan: Formal analysis, funding acquisition, supervision, writing - review and editing. Peipei Yu: Formal analysis, investigation, visualization. Chao Yang: Investigation, project administration, supervision.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jiaqi Ruan https://orcid.org/0000-0003-2584-0738
Chao Yang https://orcid.org/0000-0002-3952-5299

REFERENCES

1. Yan, R., Masood, N.-A., Kumar Saha, T., Bai, F., Gu, H.: The anatomy of the 2016 south australia blackout: a catastrophic event in a high renewable network. IEEE Trans. Power Syst. 33(5), 5374–5388 (2018). doi: https://doi.org/10.1109/TPWRS.2018.2820150
2. Bialek, J.: What does the GB power outage on 9 August 2019 tell us about the current state of decarbonised power systems? Energy Policy 146, 111821 (2020). doi: https://doi.org/10.1016/j.enpol.2020.111821
3. Liang, J., Tang, W.: Ultra-short-term spatiotemporal forecasting of renewable resources: an attention temporal convolutional network based approach. IEEE Trans. Smart Grid 13(5), 3798–3812 (2022). doi: https://doi.org/10.1109/TSG.2022.3175451
4. Panwar, N.L., Kaushik, S.C., Kothari, S.: Role of renewable energy sources in environmental protection: a review. Renewable Sustainable Energy Rev. 15(3), 1513–1524 (2011). doi: https://doi.org/10.1016/j.rser.2010.11.037
5. Yu, P., Zhang, H., Song, Y., Hui, H., Chen, G.: District cooling system control for providing operating reserve based on safe deep reinforcement learning. IEEE Trans. Power Syst. (2023). doi: https://doi.org/10.1109/TPWRS.2023.3237888
6. Papadis, E., Tsatsaronis, G.: Challenges in the decarbonization of the energy sector. Energy 205, 118025 (2020). doi: https://doi.org/10.1016/j.energy.2020.118025
7. Bera, A., Nguyen, N., Chalamala, B., Mitra, J.: Quantification of storage required for preserving frequency security in wind-integrated systems. IET Renewable Power Gener. 17(9), 2366–2378 (2023). doi: https://doi.org/10.1049/rpg2.12765
8. Peter, J.: How does climate change affect electricity system planning and optimal allocation of variable renewable energy? Appl. Energy 252, 113397 (2019). doi: https://doi.org/10.1016/j.apenergy.2019.113397
9. Yu, P., Zhang, H., Song, Y., Hui, H., Huang, C.: Frequency regulation capacity offering of district cooling system: an intrinsic-motivated reinforcement learning method. IEEE Trans. Smart Grid 14(4), 2762–2773 (2022). doi: https://doi.org/10.1109/TSG.2022.3220732
10. Ruan, J., Liu, G., Qiu, J., Liang, G., Zhao, J., He, B., Wen, F.: Time-varying price elasticity of demand estimation for demand-side smart dynamic pricing. Appl. Energy 322, 119520 (2022). doi: https://doi.org/10.1016/j.apenergy.2022.119520
11. Chen, C., Liang, H., Zhai, X., Zhang, J., Liu, S., Lin, Z., Yang, L.: Review of restoration technology for renewable-dominated electric power systems. Energy Convers. Econ. 3(5), 287–303 (2022). doi: https://doi.org/10.1049/enc2.12064
12. Bazionis, I.K., Karafotis, P.A., Georgilakis, P.S.: A review of short-term wind power probabilistic forecasting and a taxonomy focused on input data. IET Renewable Power Gener. 16(1), 77–91 (2022). doi: https://doi.org/10.1049/rpg2.12330
13. Han, L., Li, M., Wang, X., Lu, P.: Wind power forecast based on broad learning system and simplified long short term memory network. IET Renewable Power Gener. 16(16), 3614–3628 (2022). doi: https://doi.org/10.1049/rpg2.12588
14. He, C., Chen, Q., Fang, X., Zhou, Y., Fu, R., Jin, W.: Wind speed forecasting in fishing harbor anchorage using a novel deep convolutional neural network. Front. Earth Sci. 9, 731803 (2021)
15. Wang, H., Ruan, J., Ma, Z., Zhou, B., Fu, X., Cao, G.: Deep learning aided interval state prediction for improving cyber security in energy internet. Energy 174, 1292–1304 (2019)
16. Wang, H.Z., Wang, G.B., Li, G.Q., Peng, J.C., Liu, Y.T.: Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl. Energy 182, 80–93 (2016). doi: https://doi.org/10.1016/j.apenergy.2016.08.108
17. Shahid, F., Zameer, A., Mehmood, A., Raja, M.A.Z.: A novel wavenets long short term memory paradigm for wind power prediction. Appl. Energy 269, 115098 (2020). doi: https://doi.org/10.1016/j.apenergy.2020.115098
18. Khodayar, M., Wang, J., Manthouri, M.: Interval deep generative neural network for wind speed forecasting. IEEE Trans. Smart Grid 10(4), 3974–3989 (2019). doi: https://doi.org/10.1109/TSG.2018.2847223
19. Ruan, J., Fan, G., Zhu, Y., Liang, G., Zhao, J., Wen, F., Dong, Z.Y.: Super-resolution perception assisted spatiotemporal graph deep learning against false data injection attacks in smart grid. IEEE Trans. Smart Grid (2023). doi: https://doi.org/10.1109/TSG.2023.3241268
20. Chen, Y., Tan, Y., Zhang, B.: Exploiting vulnerabilities of load forecasting through adversarial attacks. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems, pp. 1–11. ACM, Press (2019)
21. Chen, Y., Sun, M., Chu, Z., Camal, S., Kariniotakis, G., Teng, F.: Vulnerability and impact of machine learning-based inertia forecasting under cost-oriented data integrity attack. IEEE Trans. Smart Grid 14(3), 2275–2287 (2023). doi: https://doi.org/10.1109/TSG.2022.3207517
22. Liu, Z., Wang, Q., Ye, Y., Tang, Y.: A GAN-based data injection attack method on data-driven strategies in power systems. IEEE Trans. Smart Grid 13(4), 3203–3213 (2022). doi: https://doi.org/10.1109/TSG.2022.3159842
23. Ruan, J., Yang, C., Wang, Q., Wang, S., Liang, G., Zhao, J., Qiu, J.: Assessment of spatiotemporally coordinated cyberattacks on renewable energy forecasting in smart energy system. Appl. Energy 347, 121470 (2023). doi: https://doi.org/10.1016/j.apenergy.2023.121470
24. Wang, H., Ruan, J., Wang, G., Zhou, B., Liu, Y., Fu, X., Peng, J.: Deep learning-based interval state estimation of AC smart grids against sparse cyber attacks. IEEE Trans. Ind. Inf. 14(11), 4766–4778 (2018). doi: https://doi.org/10.1109/TII.2018.2804669
25. Ruan, J., Liang, G., Zhao, J., Zhao, H., Qiu, J., Wen, F., Dong, Z.Y.: Deep learning for cybersecurity in smart grids: review and perspectives. Energy Convers. Econ. 4(4), 233–251 (2023). doi: https://doi.org/10.1049/enc2.12091
26. Aziz, S., Irshad, M., Haider, S.A., Wu, J., Deng, D.N., Ahmad, S.: Protection of a smart grid with the detection of cyber- malware attacks using efficient and novel machine learning models. Front. Energy Res. 10, 964305 (2022)
27. Ruan, J., Liang, G., Zhao, J., Qiu, J., Dong, Z.Y.: An inertia-based data recovery scheme for false data injection attack. IEEE Trans. Ind. Inf. 18(11), 7814–7823 (2022). doi: https://doi.org/10.1109/TII.2022.3146859

28. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. 41(3), 15:1–15:58 (2009). doi: https://doi.org/10.1145/1541880.1541882
29. Hassan, M.U., Rehmani, M.H., Chen, J.: Differential privacy techniques for cyber physical systems: a survey. IEEE Commun. Surv. Tut. 22(1), 746–789 (2020). doi: https://doi.org/10.1109/COMST.2019.2944748
30. Pang, Z.-H., Fan, L.-Z., Sun, J., Liu, K., Liu, G.-P.: Detection of stealthy false data injection attacks against networked control systems via active data modification. Inf. Sci. 546, 192–205 (2021). doi: https://doi.org/10.1016/j.ins.2020.06.074
31. Sharma, A., Kaur, P.: Tamper-proof multitenant data storage using blockchain. Peer-to-Peer Netw. Appl. 16(1), 431–449 (2023). doi: https://doi.org/10.1007/s12083-022-01410-8
32. Ruan, J., Liu, W., Zhao, J., Liang, G., Yang, C., Wen, F.: Data-driven electricity retail pricing strategy for demand response. Autom. Electr. Power Syst. 47(7), 133–141 (2023). doi: https://doi.org/10.7500/AEPS20220824007
33. Renewable Energy Dataset. https://github.com/ruanjiaqi01/Renewable-Energy-Forecasting-Cyberthreats/ (2023). Accessed 10 Sep 2022

How to cite this article: Yang, L., Liang, G., Yang, Y., Ruan, J., Yu, P., Yang, C.: Adversarial false data injection attacks on deep learning-based short-term wind speed forecasting. IET Renew. Power Gener. 18, 1370–1379 (2024). https://doi.org/10.1049/rpg2.12853