Dissecting Renewable Uncertainty via Deconstructive Analysis-based Data Valuation

Yanzhi Wang[1], *Student Member, IEEE,* Jie Song[1], *Senior Member, IEEE,*

***Abstract*—Integrating renewable energy sources into power systems is crucial to lower carbon emissions, yet the resulting uncertainty presents challenges to network reliability. In the era of digital energy, big data models help reduce uncertainty by identifying data patterns for accurate predictions. Yet, the efficacy of data-driven models is constrained by the scarcity of high-quality training data, underscoring the significance of identifying and selecting datasets of superior quality. This paper proposes a novel data valuation framework based on deep reinforcement learning for the analysis and decomposition of uncertainty in renewable energy datasets. Our framework merges meteorological and power uncertainty through predictive tasks, training a neural network to uncover the intrinsic relationships between data features and their value via a sampling-feedback mechanism. By incorporating policy gradient and other optimization techniques, we enhance the algorithm's stability and efficiency, supplemented by comparative experiments for validation. We tested our valuation approach using 2017-2018 aggregate wind related data from Yunnan province for power forecasting. The results demonstrate that our proposed data value approach effectively enhances the quality of the dataset, leading to a proportional improvement of 7.69% in prediction accuracy.**

***Index Terms*—Renewable uncertainty, data valuation, data quality, reinforcement learning, policy gradient, wind power forecasting, meteorological feature.**

I. Introduction

Renewable energy, as a clean and sustainable energy source utilized in power systems, plays a vital role in significantly reducing carbon emissions [1], [2]. Due to the deviations of natural climatic conditions, the uncertainty of renewable energy poses challenges to the safety and reliability of power system operation when it is integrated into power systems [3]. To address it, the integration of digital technologies optimizes the efficiency and life cycle of renewable energy usage through data-driven and knowledge-based analysis methods, providing foundational support in mitigating the uncertainties through prediction and modeling [4], [5]. However, the practical application faces challenges due to the scarcity of high-quality training data, which not only escalates the expenses associated with data storage and processing but also undermines the performance of models [6]. In the presence of low-quality data, these models struggle to discern relevant numerical patterns, leading to the acquisition of complex and irrelevant rules that limit their effectiveness [7]. Hence, it becomes crucial to identify and select data sets of higher quality to ensure accurate forecasting and minimize the uncertainty associated with renewable energy.

The uncertainty of renewable resources corresponds to the ambiguity of numerical patterns within power data sets, as well as their inherent unpredictability [8]. Therefore, learning to select the subset of high-quality data from the redundant data set is, in essence, a method of deconstructing renewable uncertainty. It involves comprehending the underlying causes of uncertainty within the data set through data-driven knowledge, uncovering potential numerical patterns based on this understanding, and ultimately providing valuable guidance for networks scheduling.

Building upon this foundation, proposing data-based selection strategies to enhance presence of high-quality data becomes highly constructive for power system. The contemporary research of data quality valuation encompasses three primary categories. The first category employs non-supervisory indicators as a basis for defining data quality. In the traditional field of data science, scholars employ a multidimensional framework to assess the value of a dataset, encompassing key metrics such as *Completeness*, *Accuracy*, *Timeliness*, *Consistency*, and other relevant dimensions [9]. By examining these dimensions collectively, researchers gain a nuanced understanding of the dataset's reliability, relevance, and effectiveness across a spectrum of criteria. In big data energy, some scholars utilize Shannon entropy and the non-noise ratio as metrics to assess the quality of photovoltaic-related data [10]. They have discovered an exponential relationship between the quality of the data and the accuracy of predictions. The second category centers around the design of data markets, utilizing transactions as a reflection of the value of the data. Scholars emphasize the contribution of data in reducing uncertainty two-settlement market system that incorporate load demand to define the value of data in energy transactions [11]. Other scholars adopting existing auction mechanism to renewable energy forecasting data market [12]. In this framework, data's capacity for generating more accurate predictions contributes to higher revenue, particularly through collaborative sharing, elucidating the rationale for the data's value. However, these two type of methods focus on identifying and estimating the quality of the datasets but do not fundamentally deconstruct and analyze the inherent patterns of uncertainty within them. The former underscores the correlation of characteristics at the statistical level of the datasets in influencing prediction accuracy, without initiating the analysis inside dataset from the specific characteristics of renewable energy. In contrast, the latter treats uncertainty

This work was supported by National Natural Science Foundation of China (72131001). *(Corresponding author: Jie Song).*

Yanzhi Wang and Jie Song are with the Department of Industrial Engineering and Management, Peking University, Beijing 100871, China yanzhiwang@pku.edu.cn; songjie@coe.pku.edu.cn

as an economic factor within the market but not emphasize its relation to the distributional characteristics inherent in the dataset specific to renewable energy.

The third category places a spotlight on the identification of noisy data, particularly prevalent in the realm of computational science. Scholars have applied the concept of utility allocation from the cooperative games theory to assign values to data, which has been proven to be highly effective in diagnosing mislabeled samples [13]. To mitigate the computational cost of algorithms in large samples and complex models, scholars have embraced the concept of meta-learning, deriving the value of data from the output of deep neural networks [14]. However, none of these algorithms can handle the uncertainty inherent in renewable energy sources, specifically, the data structures originating from the continuous time dimension, challenging the efficiency of the valuation framework.

In this paper, we propose a novel data valuation framework based on deep reinforcement learning to achieve data dimensionality reduction of renewable uncertainty. We effectively integrate meteorological uncertainty and power uncertainty through the design of forecasting tasks. We introduce a neural network that learns the value patterns underlying the entire data set through continuous exploration and awareness of the distribution of high-value data sets, finally guides us to distinguish high-quality data.

The major contribution of this paper are as follows:

1) We propose a theoretical framework based on deep reinforcement learning that utilizes deconstructive analysis of uncertainty scenarios. Utilizing the sampling-feedback training mechanism, the framework progressively mines the most valuable segments of the original data set. Additionally, as a significant factor, we incorporate uncertainty in meteorological characteristics into the quality assessment of renewable energy data, thereby enhancing the comprehensiveness and instructiveness of our deconstructive analysis and numerical results.
2) We utilize the policy gradient as the foundational algorithm for implementing the framework and incorporate multi-level optimization techniques to address the instability of real data performance. This approach ensures a more stable and secure screening of the optimal high-quality data set.
3) The comparative experimental results demonstrate that our algorithm, along with the proposed improvements, significantly enhances the stability of data value computation. The effectiveness of our data valuation algorithm has been confirmed in the scenario of renewable energy forecasting, as the extraction of higher-quality data significantly reduces uncertainty and demonstrates more versatility.

II. Meteorology-aided Renewable Forecasting

A. Selection of Meteorological Features

To address the renewable energy prediction problem, we meticulously considered the nature of the time series and formulated a continuous data pattern that incorporates meteorological features and power features using a sliced representation, and use it in the prediction of subsequent power. In our approach, we consider multidimensional Numerical Weather Prediction (NWP) as valuable features that can be acquired promptly and utilized as a reliable predictor. We extract both *daily* and *hourly* data, with the specific metrics outlined in Table 1. It considered the various meteorological dimensions that can impact wind power generation, representing the aspects represented by the features mentioned in Fig. 1 as well. It's worth noting that "PAR" is an abbreviation for *Photosynthetically Active Radiation*.

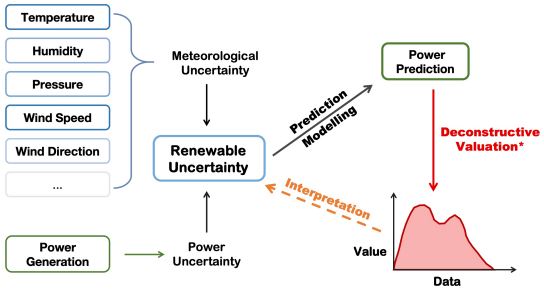

Fig. 1: The framework of deconstruct the uncertainty from power and NWPs

TABLE I: Meteorological Data Utilization with Various Features

Meteorological Feature	Hourly	Daily
Temperature at 2 Meters	√	√
Temperature at 2 Meters Maximum		√
Temperature at 2 Meters Minimum		√
Specific Humidity at 2 Meters	√	√
Relative Humidity at 2 Meters	√	√
Surface Pressure	√	√
Wind Speed at 10 Meters	√	
Wind Direction at 10 Meters	√	
Precipitation		√
Clear Sky Surface PAR	√	
All Sky Surface PAR	√	

Indeed, meteorological features are typically obtained from meteorological stations, offering data at the county level. However, when our forecasts target provincial wind power output, it necessitates the extraction of features from all meteorological stations within a specific geographic area. In order to establish a basis for the average meteorological features of the entire province, we first perform clustering of the meteorological features at the county level, and followed by arithmetic averaging. We integrate all the meteorological features of each county into a one-dimensional vector, and k-means clustering algorithm is employed to identify featured counties. Given a set of meteorological vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, where n for the number of county-level meteorological features in the province and each vector is a d-dimensional daily and hourly features, k-means clustering aims to partition the n vectors into sets $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$. Formally, the objective is to find:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \quad (1)$$

where $\boldsymbol{\mu}_i$ is the mean of vectors in S_i, as follow:

$$\boldsymbol{\mu_i} = \frac{1}{|S_i|} \sum_{\mathbf{x} \in S_i} \mathbf{x} \tag{2}$$

After obtaining the clustering centers, our process involves selecting counties with the closest Euclidean distance to the feature vectors as representative counties. Subsequently, we compute the average meteorological feature values of these k counties. These averaged values then function as the meteorological feature inputs for the entire province.

B. Design of Forecasting Tasks

Our forecasting model has been specifically designed to effectively incorporate both the hourly and day-level meteorological data. In our forecasting model, we specifically utilize the daily NWPs and the previous 48 hours of power data as fixed inputs to predict the power generation for the next 24 hours. Each hourly NWPs is then employed to independently forecast the corresponding 24 hours of power data for the day, as shown in Fig. 2. This methodology allows us to harness the temporal relationship between weather patterns and power generation. Additionally, it enables the seamless integration of weather uncertainty into power uncertainty, ensuring a comprehensive and reliable assessment of the overall uncertainty.

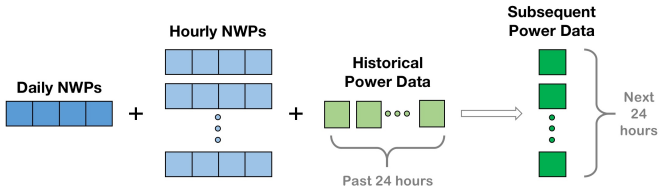

Fig. 2: The Design of Renewable Power Prediction incorporate NWPs

We have selected Mean Absolute Percentage Error (MAPE) as the metric to accurately measure the accuracy of our prediction model. However, due to the fluctuation and instability of wind power and the proportional calculation in MAPE, we aim to mitigate the impact of small perturbations in cases of low generation on our overall judgment of the prediction. Therefore, we employ an adjusted Mean Absolute Percentage Error (MAPE) as follows:

$$MAPE = \frac{100\%}{n * 24} \sum_{i=1}^{n} \sum_{t=1}^{24} \left| \frac{\hat{y}_{i,t} - y_{i,t}}{\max\left(y_{i,t}, \epsilon_y\right)} \right| \tag{3}$$

where $y_{i,t}$ means real wind power generation on day i hour t and $\hat{y}_{i,t}$ represent the predicted value. The ϵ_y is the lower bound that count in the metric, which is generally chose a relatively small value.

C. Light Gradient Boosting Machine

To enhance the fitting of wind power time series and incorporate the influence of multi-scale meteorological features in our forecast, we have employed an improved gradient boosting tree called Light Gradient Boosting Machine (LightGBM) as our forecasting model [15]. It allows us to capture the intricate relationships and dependencies within the multidimensional data, leading to more precise wind power predictions.

LightGBM can be seen as an enhanced algorithm compared to Gradient Boosting Decision Tree, offering substantial improvements in terms of training speed and memory consumption. Assuming that the predicted value from features vector $\mathbf{x}_i$ can be written as $\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i)$, where each f_k is a regression tree. The objective for f_k to minimize is as follows:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(k-1)} + f_k(\mathbf{x}_i)\right) + \Omega(f_k) \tag{4}$$

where $\hat{y}_i^{(k-1)}$ represent the prediction value under the previous $k-1$ regression trees, and the metric Ω penalizes the complexity of the tree model [16]. LightGBM improves the efficiency of data sampling and sorting, which significantly reduces complexity and makes it well-suited for handling large-scale data during training computations. Furthermore, LightGBM has been thoroughly verified to effectively capture patterns in time series data and handle a large number of features with high accuracy [17]. In line with our specific renewable energy prediction design, we have trained a total of 24 separate models. Each model is trained to predict wind power at a specific hour, ranging from 1h to 24h, allowing us to make accurate predictions for each hour of the day.

III. Data Quality Valuation

A. Reinforcement Learning Framework

We have represented the deconstructive analysis-based data quality valuation framework visually in the form of Fig. 3. In the initial step, all features of the data are inputted into a quality evaluator composed of deep learning network. Based on the classification output, the evaluator forms high-quality and low-quality data sets. Subsequently, the high-quality data set is selected for uncertainty testing within renewable prediction scenarios. Upon satisfying the boundary conditions, the data set can be utilized as smart data extracted from the original data set for better forecasting in distribution network (DN) application. This sampling process based on data quality aims to minimize the uncertainty of renewable energy, resulting in a more reliable data set that aligns with the given conditions. However, if the boundary conditions are not met, it implies that the high-quality data set obtained through the current evaluator selection still exhibits significant uncertainty and is biased compared to the actual high-quality data. In such cases, the current evaluation is fed back to the evaluator as an unsuccessful signal. This prompts the data quality evaluator to iteratively readjust the parameter weights through self-learning and repeat the "sampling-feedback" process. This iterative process continues until the uncertainty validation condition is met, ensuring that the high-quality data set ultimately satisfies the desired level of uncertainty.

The sampling-feedback mechanism within this framework can be facilitated with the assistance of reinforcement learning theory. In this context, the data quality evaluator functions as the *Agent*, while the uncertainty validation outcome serves

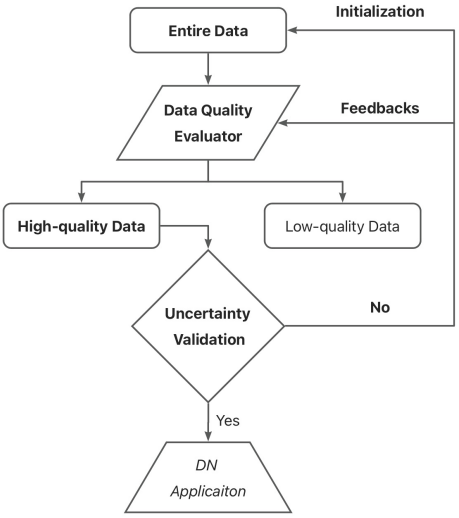

Fig. 3: Sampling-feedback mechanism in data quality assessment

as the *Reward* provided by the iteratively trained prediction model, which can be interpreted as the *Environment*. The agent's strategy for selecting high-quality data is continuously updated *Action* based on this feedback, enabling a dynamic and adaptive approach to data quality assessment.

B. Problem Formulation

In this section, our objective is to rationalize the modeling of reinforcement learning, with the ultimate goal of expanding the capabilities of computing data values through the integration of deep learning techniques. Without loss of generality, we denote the entire training data set as $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N} \sim \mathcal{P}$ where $\mathbf{x}_i \in \mathcal{X}$ is a g-dimensional input vector, containing meteorological features and power sequence which already shown in Fig. 2 and $\mathcal{P}$ symbolize the origin data distribution along with renewable uncertainty. Besides, $\mathbf{y}_i \in \mathcal{Y}$ is a k-dimensional output vector, which is generally treated as the forecast for a future period of time (24 hours). To account for the distinction between the training and test datasets, we introduce a disjoint testing dataset denoted as $\mathcal{D}^t = \{(\mathbf{x}_j^t, \mathbf{y}_j^t)\}_{j=1}^{M} \sim \mathcal{P}^t$. In this context, the target distribution $\mathcal{P}^t$ is not necessarily identical to the training distribution $\mathcal{P}$. Our target distribution $\mathcal{P}^t$ is default to be deterministic, serving as a reference for correcting and evaluating the uncertainty present in distribution $\mathcal{P}$.

The framework of the Data Quality Valuation is trained as shown in Fig. 4. The whole training process involves two functions: the Data Quality Evaluator (DQE, h_ϕ) and the Renewable Prediction model ($f_\theta = \sum_{k=1}^{n} f_k$). The h_ϕ is a deep learning network with parameters ϕ, designed to capture intricate patterns in data values. Its role is to assess and filter the data to ensure high quality, enhancing the overall model performance. The $h_\phi : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ is optimized to output the weight that determine the probability of being sampled as high-quality data. These weights can be interpreted as data values, representing the probability of individual data points trained for the better prediction model f_θ, finally to meet the condition. In essence, the higher the data weight $h_\phi(\mathbf{x}_i, \mathbf{y}_i)$, the more likely the corresponding data point $(\mathbf{x}_i, \mathbf{y}_i)$ is to be selected for training the final model f_θ. This reflects the higher quality and significance of the data, as it has a better chance to contribute to the training of the final task model, which good enough to meet the uncertainty condition. Conversely, lower data weights suggest that the data points are less likely to be chosen, indicating their lower value and limited impact on the optimal model's training process. By assigning these data values, we can effectively identify and prioritize high-quality data, leading to improved performance of the final predictive model. To provide a more intuitive understanding, we can view the sampling process as a binomial filter $S : [0, 1] \to \{0, 1\}$. This filter determines whether a data point represented by $(\mathbf{x}_i, \mathbf{y}_i)$ is selected (value of 1) or not selected (value of 0) for training the predictor model. The formulate of the corresponding optimization problem in the Reinforcement Learning is as follows:

$$\min_{h_\phi} \underset{(\mathbf{x}^t, \mathbf{y}^t) \sim P^t}{E} \left[\mathcal{L}_h \left(f_\theta \left(\mathbf{x}^t \right), \mathbf{y}^t \right) \right] \tag{5}$$

$$\text{s.t. } f_\theta = \arg\min_{f} \underset{(\mathbf{x}, \mathbf{y}) \sim P}{E} \left[S(h_\phi \left(\mathbf{x}, \mathbf{y} \right)) \mathcal{L}_f (f(\mathbf{x}), \mathbf{y}) \right] \tag{6}$$

where $\mathcal{L}_f$ is the loss function calculate during the training of renewable prediction model, while $\mathcal{L}_h$ symbolize the validation loss between the predicted and testing data.

In this framework, the training data set $\mathcal{D}$ is the constant *state*, and the selection of high-quality data set $S(h_\phi \left(\mathbf{x}, \mathbf{y} \right))$ is the *action* made by DQE (*agent*). Finally, for each iteration, the loss is represented as $\mathcal{L}_h \left(f_\theta \left(\mathbf{x}^t \right), \mathbf{y}^t \right)$, which serve as the *reward* reinforced the training of the DQE. In flowchart Fig. 4, we use prediction accuracy as the uncertainty validation and set a convergence exit condition. Due to the requirement of the Back Propagation (BP) algorithm for function differentiability during DQE network training, the function $S(h_\phi)$ with discrete output values of $\{0, 1\}$ does not allow for regular training. To handle non-differentiable problems, we employ the *Policy Gradient* algorithm in reinforcement learning [18]. The nature of the Policy Gradient algorithm allows for converting the gradient computation from the original target to the logarithm of the probability of the current action. This transformation helps to circumvent the problem of non-differentiability in terms of the *reward* concerning the *agent* parameters. We simply assume the high-quality data set filter S is based on binomial distribution, so the probability that the selection vector $\mathbf{s}$ is selected based on $h_\phi(\mathcal{D})$ is that $\pi_\phi(\mathcal{D}, \mathbf{s}) = Prob \left(\mathbf{s} = [S(h_\phi \left(\mathbf{x}_i, \mathbf{y}_i \right))]_{i=1 \ldots N} \right) = \prod_{i=1}^{N} \left[h_\phi \left(\mathbf{x}_i, \mathbf{y}_i \right)^{s_i} \cdot \left(1 - h_\phi \left(\mathbf{x}_i, \mathbf{y}_i \right) \right)^{1 - s_i} \right]$, where s_i is the component of the vector $\mathbf{s}$ on data point i. As a result, the single step reward based on the action $\mathbf{s}$ is defined as $l(\phi)$, where:

$$\begin{aligned} & l(\phi) \\ &= \underset{(\mathbf{x}^t, \mathbf{y}^t) \sim P^t, \mathbf{s} \sim \pi_\phi(\mathcal{D}, \cdot)}{E} \left[\mathcal{L}_h \left(f_\theta \left(\mathbf{x}^t \right), \mathbf{y}^t \right) \right] \\ &= \int \sum_{\mathbf{s} \in [0,1]^N} \pi_\phi(\mathcal{D}, \mathbf{s}) \cdot \left[\mathcal{L}_h \left(f_\theta \left(\mathbf{x}^t \right), \mathbf{y}^t \right) \right] dP^t \left(\mathbf{x}^t, \mathbf{y}^t \right) \end{aligned} \tag{7}$$

The gradient of $l(\phi)$ incorporate with Policy Gradient is as follows:

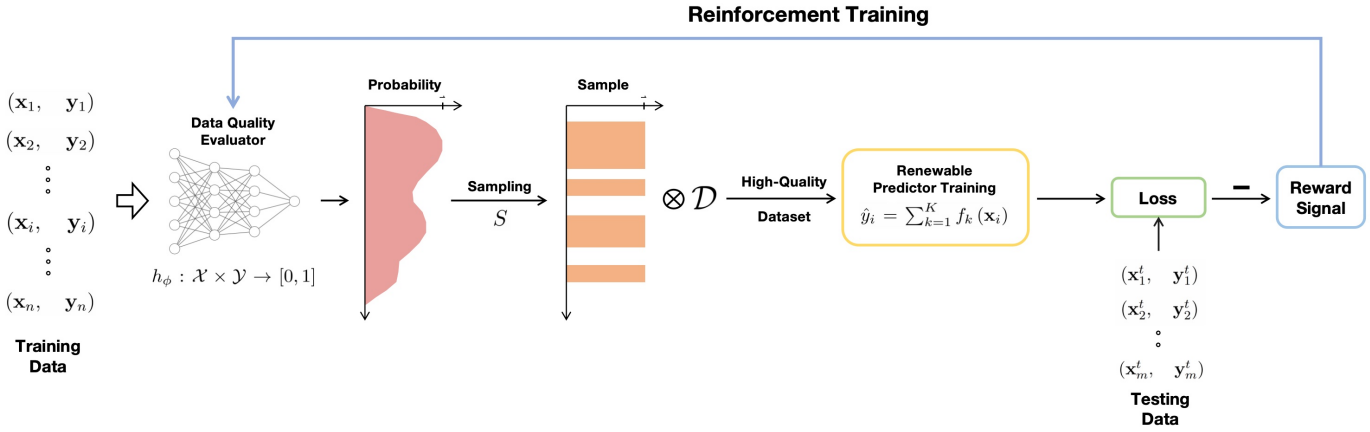

Fig. 4: The Framework of Data Quality Valuation

$$
\begin{aligned}
&\nabla_\phi l(\phi) \\
&= \int \sum_{\mathbf{s}\in[0,1]^N} \nabla_\phi \pi_\phi(\mathcal{D},\mathbf{s}) \cdot \left[\mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] dP^t\left(\mathbf{x}^t,\mathbf{y}^t\right) \\
&= \int \sum_{\mathbf{s}\in[0,1]^N} \frac{\nabla_\phi \pi_\phi(\mathcal{D},\mathbf{s})}{\pi_\phi(\mathcal{D},\mathbf{s})} \pi_\phi(\mathcal{D},\mathbf{s}) \cdot \left[\mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] dP^t \\
&= \underset{\substack{\left(\mathbf{x}^t,\mathbf{y}^t\right)\sim P^t \\ \mathbf{s}\sim\pi_\phi(\mathcal{D},\cdot)}}{E} \left[\mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] \nabla_\phi \log\left(\pi_\phi(\mathcal{D},\mathbf{s})\right)
\end{aligned} \tag{8}
$$

Since the DQE output $h_\phi(\mathbf{x},\mathbf{y})$ for each data point $(\mathbf{x}_i,\mathbf{y}_i)$ remains continuously differentiable with respect to the action $\mathbf{s}$, the gradient for the log probability can be computed, allowing for the iteration of the neural network parameters using the BP algorithm. By applying the convergence condition, the optimal DQE can iteratively identify the value of each data point and subsequently filter the high-quality data, enhancing the performance of the prediction task and contributes to the overall success of the assessment analysis.

C. Efficient Valuation on Deconstruct Uncertainty

The essence of our proposed framework centers on translating the complex estimation of data quality into the training of a DQE network. This process is tantamount to embedding the interpretation of uncertainty introduced by data quality directly into this data valuation model. As the well-trained DQE model processes the data, consisting of wind power and pertinent meteorological features, it generates a value index. Consequently, the DQE model has the capability to decipher the quality distribution within internal data, thereby deconstructing the uncertainty introduced by the dataset.

Regarding the validity of our proposed framework, it can be scrutinized from two perspectives. First, in contrast to employing traditional fixed statistical metrics, e.g., entropy, for measuring data ambiguity to explain uncertainty, we adopt a supervised learning model with a data-driven loss function L_h optimized by the DQE network. Importantly, the selection of the uncertainty metric is flexible, and in this paper, we opt for using the MAPE of the prediction results to directly represent uncertainty. Simultaneously, we introduce reinforcement learning and leverage policy gradient techniques to address the non-differentiability issue arising from dataset sampling, which ensures the iterability and unbiasedness of training. Consequently, we are able to directly utilize uncertainty in renewable energy scenarios to assign values to dataset quality in a more direct, flexible and robust manner.

Another validity stems from the algorithm. In traditional data value assignment method, like the Shapley Value (SV), it often requires traversing all potential combinations, as illustrated below:

$$
Q(i) = \sum_{S\subseteq D\setminus\{i\}} \frac{|S|!(N-|S|-1)!}{N!}\left(V(S\cup\{i\}) - V(S)\right) \tag{9}
$$

where $Q(i)$ is the value of data i, D is the full set of data, $|S|$ equals to the amount of data in subset S, and V is the utility defined by data subset. The time complexity of SV, involving the traversal of the entire subset to compute the average editorial utility, is exponential and becomes particularly large as the volume of data increases. Moreover, the outcomes are not directly pertinent to the acquisition of a subset of high-value data. The advancement in our approach lies in reorienting the optimization goal straightly towards the acquisition of high-quality values, specifically learning the optimal selection strategy $S\left(h_\phi(\mathbf{x},\mathbf{y})\right)$. Leveraging the framework of deep reinforcement learning, we employ gradients to iteratively navigate through batches of samples, facilitating the exploration and identification of higher-value subsets. This proves to be efficient in spaces with a large number of subset samples (2^N). Given the stochastic nature of the policy gradient, we cannot ascertain global optimality in the final data value ranking. Nevertheless, we can affirm that the identification of high-quality subsets will result in substantial improvements in uncertainty reduction within the specified time constraints. This strengthens the deconstructive analysis of uncertainty.

D. The Network Design of Data Quality Evaluator

We have opted for a regionally separated fully-connected neural network to assess the value of our defined dataset.

Specifically, in the initial stage of the network, we partition the input data $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, including daily NPWs, hourly NWPs, past 48 hours power, and next 24 hours power, separately. Afterward, we reintegrate these partitions in the later stages of the network. The final output of the network utilizes the *sigmoid* function to generate continuous values between 0 and 1, representing the value of the data, shown in Fig. 5.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

This approach effectively segregates the characteristics of different dimensions while also reflecting the importance of data integration.

In our experiments, we observed the presence of outliers and noises in the wind power data. These anomalies often pose challenges for the DQE, leading to difficulties in fitting the underlying patterns of data features and values, oscillate during the training period at the beginning. To expedite convergence, we undertake the base model $\tilde{f}_v$ which trained on the testing dataset before DQE training. This trained model is then employed as a parameter to calculate the deviation of wind power data relative to $\tilde{f}_v$, named as Error Correction (EC), serving as one of the reference indices for the output value of the DQE.

$$EC\left(\mathbf{x}_i, \mathbf{y}_i\right) = \frac{\mathbf{y}_i - \tilde{f}_v\left(\mathbf{x}_i\right)}{\mathbf{y}_i + \tilde{\epsilon}} \tag{11}$$

where $\tilde{\epsilon}$ is a small value avoid EC tends to infinity. It is imperative to emphasize the rationale behind this approach lies in the consistency of parameters within $\tilde{f}_v$ throughout the training process. The primary objective of DQE model training is to scrutinize the uncertainty distribution within the training set dataset by evaluating predictive performance on the test set during validation. The error correction furnishes the distance of the input data relative to the predictive patterns observed in the test dataset, utilizing pretraining-like parameters as a potential reference for the DQE for potentially identifying the outliers and noise. Importantly, it does not impact the modeling of the inputs and outputs of the DQE, nor does it contradict DQE's ability to gain insights into D^t in Reinforcement learning. In summary, the DQE produces 0-1 values as data values by taking into account both the meteorological features and the wind power sequences of the data as inputs, and the EC mechanism plays a role in potentially managing outliers and noise within the data during the training process.

IV. Mechanism for Algorithm Optimization

A. Ensurement of Exploratory

Combining the actual numerical testing, we find that $h_\phi\left(\mathbf{x}_i, \mathbf{y}_i\right)$ is likely to be near 0 or 1 for all data in $\mathcal{D}$. The former means that the data are all worthless and the datasets will be randomly selected in each round, which is the protection mechanism we design to avoid learning interruption. While the latter means that the datasets are all very valuable and cannot be discarded, which caused by the small set of data changing bring little effect of the whole performance.

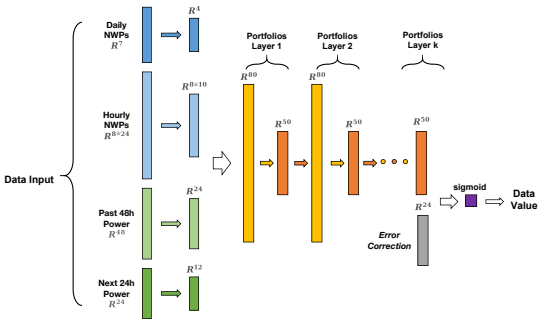

Fig. 5: The Network Design of Data Quality Evaluator

This convergence to a *local optimum* is due to the lack of exploration of potential strategies. In RL framework, we achieve exploration through a probability-based stochastic strategy, which favors diversity and avoids overly deterministic strategies. By incorporating randomness and uncertainty in our exploration process, we can effectively explore various possibilities and discover more diverse solutions. This stochastic approach enhances the model's ability to adapt and learn from different scenarios, leading to improved performance and better handling of uncertainties in the predictive tasks. We then design penalty objective function $l_{penalty} = l(\phi) + p(\phi; \sigma, h)$ to increase the likelihood that the learner will jump out of the *local optimum* and continue exploring:

$$\begin{aligned} p(\phi; \sigma, h) =& \sigma \cdot \max\left(\sum_{i=1}^{N} h_\phi\left(\mathbf{x}_i, \mathbf{y}_i\right) - N \cdot h, 0\right) + \\ & \sigma \cdot \max\left(1 - \sum_{i=1}^{N} h_\phi\left(\mathbf{x}_i, \mathbf{y}_i\right) - N \cdot h, 0\right) \end{aligned} \tag{12}$$

where σ is the penalty factor and h is the threshold near 1. When the value of the data is excessively concentrated around certain pole of $\mathbf{0}$ or $\mathbf{1} \in R^N$, the penalty term activates on a factor σ scale. This mechanism guides the DQE away from local optima by imposing a substantial loss for persistent exploration. In doing so, the DQE is encouraged to explore alternative solutions more extensively, preventing it from becoming overly focused on a limited range of data values and facilitating better convergence towards *global optima*.

B. Improvement of Sample Efficiency and Stability

As an on-policy algorithm, the policy gradient is prone to instability. This instability can arise from issues such as inadequately determined training step size. If the step size is too large, the learned policy may oscillate and fail to converge; if it is too small, the training process may become lengthy and computationally inefficient, which is highly undesirable for calculating data values [19].

To overcome these challenges and improve the algorithm, we introduce importance sampling. By implementing importance sampling, we enhance the efficiency of sample usage, effectively transforming the policy gradient into an off-policy algorithm. This modification addresses the instability issues and allows for more stable and efficient training, making the

algorithm more suitable for calculating data values in our context. As the result, the DQE learn from the samples that come from another networks $\tilde{\phi}$, and the objective function l can be re-written as follows:

$$\begin{aligned} l &= \int \sum_{\mathbf{s}\in[0,1]^N} \pi_{\phi}(\mathcal{D},\mathbf{s}) \cdot \left[\mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] dP^t\left(\mathbf{x}^t,\mathbf{y}^t\right) \\ &= \int \sum_{\mathbf{s}\in[0,1]^N} \pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s}) \frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \cdot \left[\mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] dP^t \\ &= \underset{\substack{(\mathbf{x}^t,\mathbf{y}^t)\sim P^t \\ \mathbf{s}\sim\pi_{\tilde{\phi}}(\mathcal{D},\cdot)}}{E} \left[\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right)\right] \end{aligned} \tag{13}$$

However, this enhancement comes at a cost. The use of off-policy algorithm increases the variance of the estimation, resulting in the following:

$$\begin{aligned} &\underset{\mathbf{s}\sim\pi_{\tilde{\phi}}(\mathcal{D},\cdot)}{\mathrm{Var}} \left[\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \mathcal{L}_h\right] \\ &= \underset{\mathbf{s}\sim\pi_{\tilde{\phi}}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \mathcal{L}_h\right]^2 - \left(\underset{\mathbf{s}\sim\pi_{\tilde{\phi}}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \mathcal{L}_h\right]\right)^2 \\ &= \underset{\mathbf{s}\sim\pi_{\phi}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})} \mathcal{L}_h^2\right] - \left(\underset{\mathbf{s}\sim\pi_{\phi}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\mathcal{L}_h\right]\right)^2 \end{aligned} \tag{14}$$

For the online-policy, the expression for its variance is:

$$\underset{\mathbf{s}\sim\pi_{\phi}(\mathcal{D},\cdot)}{\mathrm{Var}} \left[\mathcal{L}_h\right] = \underset{\mathbf{s}\sim\pi_{\phi}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\mathcal{L}_h^2\right] - \left(\underset{\mathbf{s}\sim\pi_{\phi}(\mathcal{D},\cdot)}{\mathrm{E}} \left[\mathcal{L}_h\right]\right)^2 \tag{15}$$

The second term remains the same, while the first term differs due to the distribution gap between the two policies ϕ and $\tilde{\phi}$. If the gap between the two distributions is considerable, it may result in a significant deviation from the expected value of the final objective function obtained [20].

Considering that off-policy sampling increases the variance and restricts updates to new policies, we mitigate these effects by introducing the *clipped function*. This function limits the difference between old and new policies, allowing us to maintain a more controlled and less sensitive policy gradient, even with larger step sizes, as follows:

$$clip(x,a,b) = min(max(x,a),b) \tag{16}$$

The use of the clipped function helps strike a balance between stability and efficiency in our policy gradient algorithm [21]. In summary, we reformulate the objective in off-policy and clipped function as l_{clip}:

$$\begin{aligned} &l_{clip} \\ &= \underset{\substack{(\mathbf{x}^t,\mathbf{y}^t)\sim P^t \\ \mathbf{s}\sim\pi_{\tilde{\phi}}(\mathcal{D},\cdot)}}{E} clip\left(\frac{\pi_{\phi}(\mathcal{D},\mathbf{s})}{\pi_{\tilde{\phi}}(\mathcal{D},\mathbf{s})}, 1-\epsilon, 1+\epsilon\right) \cdot \mathcal{L}_h\left(f_\theta\left(\mathbf{x}^t\right),\mathbf{y}^t\right) \end{aligned} \tag{17}$$

where ϵ is a small constant that restricts the probability ratio of the same action under different policies, thereby ensuring that the gradient used for network updates remains appropriately balanced and neither too large nor too small. In our experiments, we aim to maintain the sample network close to the DQE network. To achieve this, we update the sample network with the DQE network assignment every c times. This approach helps to stabilize and align the two networks, facilitating more consistent and reliable training outcomes. By ensuring that the moments of the sample network are in proximity to the DQE network, we enhance the performance and convergence of our algorithm during the training process. We combine these two components with the original framework to form the following Penalty Clipped Data Valuation (PCDV) algorithm with baseline δ for increasing stability.

Algorithm 1 Penalty Clipped Data Valuation (PCDV)

Input: training dataset $\mathcal{D}$, loss function $\mathcal{L}_{f,h}$, learning rate β, penalty factor σ, value threshold h, clipping threshold ϵ, network update cycle c

Output: learner h_ϕ, data value $h_\phi(\mathcal{D})$ for $i = 1, 2...N$

Initialize: parameter ϕ

for $m \in 1, 2...M$ **do**
  if $c \bigm| m$ **then**:
    $\tilde{\phi} \leftarrow \phi$
  for $i \in 1, 2...N$ **do**
    Sample a filter vector $s_i \leftarrow S\left(h_\phi\left(\mathbf{x}_i, \mathbf{y}_i\right)\right)$
  Train the Renewable Predictor:
    $f_\theta = \arg\min \quad \mathbf{s}^T \mathcal{L}_f(f(\mathbf{x}), \mathbf{y})$
  Update the Learner parameter:
    $\phi \leftarrow \phi + \beta \cdot [l_{clip} - \delta] \cdot \nabla_\phi \log\left(\pi_\phi(\mathcal{D}, \mathbf{s})\right)$
    $+\beta \cdot \nabla_\phi p(\phi; \sigma, h)$
  Update the baseline:
    $\delta \leftarrow \frac{m-1}{m}\delta + \frac{1}{m}\left[l_{clip} + p(\phi; \sigma, h)\right]$
  if convergence **then**:
    break
return h_ϕ, $h_\phi(\mathcal{D})$ for $i = 1, 2...N$

V. Numerical Result

Given the considerable cost associated with parameter tuning in reinforcement learning, we adopted the relevant parameters and initialization settings from [21], [22]. In particular, we set $c = 3$, $h = 0.9$ and $\epsilon = 0.2$. We conducted an uncertainty analysis using the 2017-2018 hourly aggregate wind power data from Yunnan Province [23]. Concerning meteorological features, we leveraged NASA's wind-related data for 136 counties in Yunnan Province, spanning the hourly readings throughout 2017-2018. This dataset was employed to underpin the prediction of wind power and comprehensively account for sources of uncertainty.

A. Meteorological Feature Clustering

To investigate the impact of varying the number of clustering centers (k) on capturing averaged weather features in the Yunnan province, we performed a sensitivity analysis based on temperature distribution for values of 1, 3, and 5, as illustrated in Fig. 6. Our analysis indicates that varying values of k do not yield significant changes in the distribution of temperature features. Consequently, to ensure consistency and rationale in our approach, we have chosen to set $k = 3$ for clustering 136 counties in Yunnan Province based on meteorological features.

The resulting clusters were then averaged to summarize the features sequence of Yunnan Province.

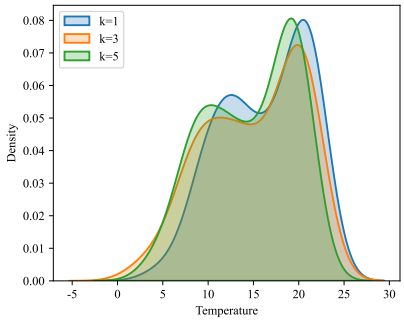

Fig. 6: Comparison of varied clustering centers on temperature distribution

The results, representing the clustering of counties using latitude and longitude, are depicted in Fig. 7. We observe a lack of clear geographic patterns in the meteorological clustering centers, indicating inconsistencies with the actual geographic clustering centers. This highlights the disparity between the meteorological and geographic distributions in Yunnan Province, underscoring the necessity for feature-based clustering.

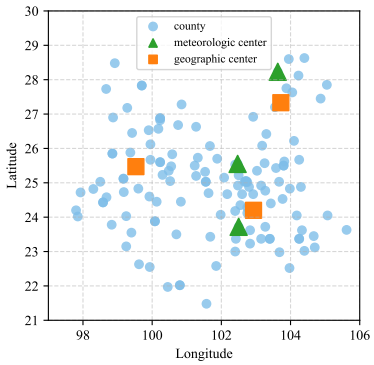

Fig. 7: The clustering results of 136 counties in Yunnan Province

B. Prediction Model Selection

Following the averaging of clustering centers, we acquire all the requisite input and output features for the prediction task. Our intention is to utilize the superior-performing prediction model as a benchmark for the identification of data value in our proposed PCDV algorithm. Given that the results of the predictive model in data valuation serve as a reward for each iteration, and considering our emphasis on obtaining a more accurate assessment of data value rather than aiming for high precision under a fixed dataset, our specific request is for the prediction model to converge rapidly and output its performance accurately and robustly within a reasonable time-frame. We select LASSO, Support Vector Regression (SVR), Multilayer Perceptron (MLP), RandomForest (RF), Gradient Boosting Decision Tree (GBDT), and LightGBM as candidate predictive models, and conduct a comparative analysis using cross-validation based on MAPE to assess their predictive performance, the results are in Table II. Our analysis revealed that LightGBM demonstrates a notably inferior error (30.55%) in estimating wind power output compared to other predictive models. Consequently, we have opted for LightGBM as the preferred prediction model for our study.

TABLE II: Prediction Model Comparison

Model	Performance (MAPE)
LASSO	73.19%
SVR	66.30%
MLP	60.77%
RF	48.39%
GBDT	50.56%
LightGBM	**30.55%**

C. Validations and Comparisons of Data Valuation

The proposed PCDV algorithm is repeatedly tested in multiple experiments. In each experiment, optimal data values are obtained after the training process of DQE, serving as a index for the probability of data in the best quality dataset. To evaluate the algorithm's effectiveness, we employ a method wherein we sort the calculated data values and progressively eliminate a portion of data from the lowest/highest, symbolized by "rlow"/"rhigh". Subsequently, we retrain the prediction model using a subset of the retained data and assess the prediction accuracy (MAPE) on the test set. To underscore the algorithm's superiority, we conduct a comparative analysis with data Shapley Value (SV) [13] and Shannon Entropy (SE) with Parzen window estimation method [10]. Given that entropy is interpreted as a statistical characterization of the dataset, we employ a leave-one-out approach to compute the value for individual data. Results are illustrated in Fig. 8.

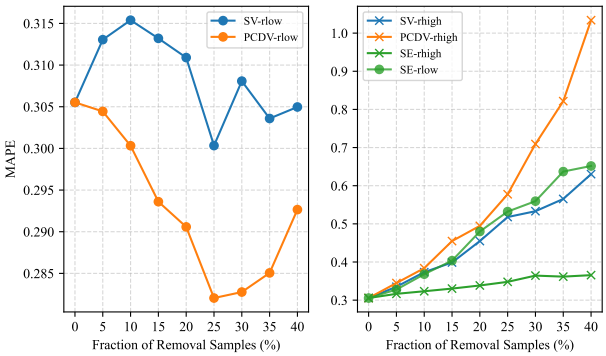

Fig. 8: The numerical outcomes of data removal experiments

Under the PCDV algorithm, we observed that the removal of low-value data effectively reduces MAPE, especially when the removal amount is within 25% with the impact intensifies with increased data removal. At the 25% removal mark, the quality of the retained data subset peaks, resulting in a controlled MAPE of 28.20%, accompanied by a proportional 7.69% reduction in uncertainty. However, as data removal surpasses this

threshold, the value of the removed data is increasing. Some feature data loss leads to a decline in the model's generality, causing a rebound and rise in MAPE, while reflecting the rise in prediction uncertainty caused by the dataset. Conversely, the trend in removing high values is notably more consistent: the MAPE of the predictive model increases with the growing amount of removed data, and this rise tends to be more rapid and steep. This pattern emerges because data uncertainty is influenced by a blend of data quality and data quantity. The gradual intensification of large-scale feature data missingness with increased removal contributes to heightened uncertainty in the dataset.

In contrast, SE's assessment of the quality of datasets with large-scale features tends to be significantly biased. The values defined by entropy cannot not effectively reduce uncertainty by enhancing data quality; instead, they introduce disorder in the ordering of values. This limitation arises because Entropy-based data value does not consider the impact of predictive modeling on the data, and its ability to extract complex predictive laws is constrained when dealing with datasets with large dimensions. Besides, the removal pattern of SV is not as pronounced, and the maximum MAPE improvement achieved by removing low-value data remains at 25%. This reduction in uncertainty is nearly by 1.70%, which is considerably smaller compared to PCDV. Simultaneously, the MAPE increase observed in PCDV when removing high-value data is significantly larger than that in SV. These more interpretable results and the more accurate enhancement in prediction performance affirm that PCDV possesses an advantage in evaluating data quality.

TABLE III: Performance differences under removal of valued dataset (ΔMAPE)

Fraction	CDV	PDV	PCDV
5%	**4.52%**	3.72%	4.04%
15%	13.43%	14.07%	**16.12%**
25%	25.60%	26.69%	**29.55%**
35%	44.21%	42.84%	**53.66%**

On the other hand, we also conduct ablation tests on the improvements, by removing the penalty (CDV) and clipped function (PDV) from the PCDV algorithm. To quantify the differences more effectively, we utilize the MAPE difference (ΔMAPE) between the removal of equal amounts of high-value data and low-value data as an evaluation index. The specific results are presented in Table III. As the amount of removed data increases, we observe that the penalty (P) and clipped (C) mechanisms in PCDV play a more pronounced role, which enables the algorithm to more accurately capture the quality patterns present in the dataset.

D. Data Quality in Wind Power Forecasting

To illustrate the influence of our data valuation on wind power forecast performance while parsing uncertainty, we present the actual and predicted outputs for wind power curves. The predicted outputs encompass results obtained by training the model with the full dataset without data filtering (r_0%), as well as predictions after removing 25% of low-value data (rlow_25%) and 25% of high-value data (rhigh_25%), as depicted in Fig. 9. In the two presented prediction scenarios, we observe that the impact introduced by data filtering is negligible within the initial 8 hours. This can be attributed to the time series nature of wind power, which exhibits some inertia during next short period, resulting in a minimal impact from different quality prediction models on prediction accuracy. Nevertheless, as time progresses, the role of numerical inertia on wind power output diminishes, and the significance of forecasting models trained using historical wind and meteorological data becomes more pronounced. We observe that the variance in forecast performance among different quality of datasets becomes substantial in the later stages of wind power forecasting, particularly around the 18-hour mark. The removal of low-quality datasets considerably diminishes prediction bias compared to no data filtering, whereas the exclusion of high-value data markedly amplifies the range of bias, deviating significantly from the actual output. This serves as an intuitive demonstration of the efficacy of our uncertainty deconstructive analysis within the context of our valuation paradigm. The significance of data value becomes distinctly evident during periods characterized by heightened uncertainty in wind power forecasts.

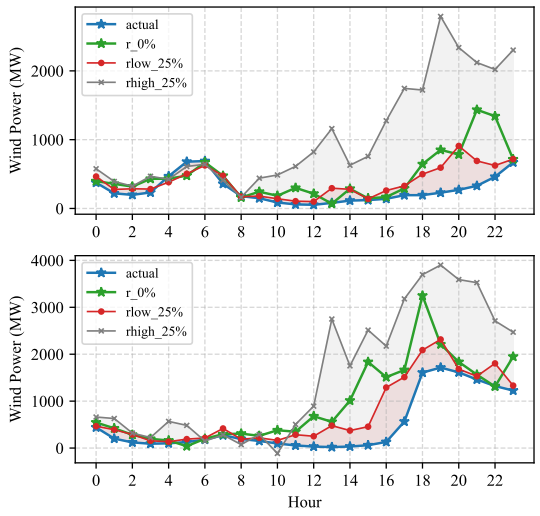

Fig. 9: Actual and predicted output of wind power

Furthermore, to intuitively showcase the data quality evaluator's performance through the feature distribution of high/low value datasets, we segment the data set into 8 equal intervals based on data values, arranged from high to low. We examined the disparities in the temperature and humidity distributions across these datasets, and the specific results are depicted in Fig. 10. Each point represents a data subset, and the color reflects the magnitude of averaged data value of each subset. We observe a discernible trend wherein higher-value data subsets exhibit greater variance in their features distribution. This phenomenon can be attributed to the increased dispersion of features within a dataset, encompassing a broader range of crucial characteristic samples. Consequently, higher-value

data subsets tend to be more advantageous for general model training. In contrast, datasets with more concentrated features cover a relative narrower span of important characteristic samples, resulting in a weaker average marginal effect on model training.

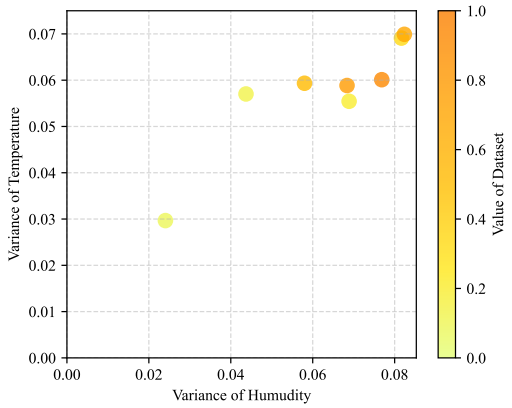

Fig. 10: Variance of features in datasets of different value

VI. Conclusion

In this paper, we present an uncertainty evaluation framework that analysis meteorological and energy data in renewable scenarios. Leveraging the technical framework of reinforcement learning, we conduct the algorithm theoretically and propose enhancement strategies to improve its efficiency and stability. Using the 2017-2018 aggregate wind power generation data from Yunnan Province and hourly meteorological data from NASA, we successfully extract high-quality dataset by assessing the value of each wind uncertainty-related data point based on forecasting performance. We validate the effectiveness of our data quality valuation framework and the PCDV algorithm in wind power prediction enhancement through comparative and removal experiments. Furthermore, we conduct an in-depth analysis of the role identified high-quality data plays in reducing wind power uncertainty and its correlation with wind-related features' patterns. In conclusion, this data quality assessment paradigm can serve as a guiding framework for data-driven power system forecasting and subsequent decision-making applications.

In contemplating future research directions, two crucial points warrant in-depth investigation: 1. Exploring the incorporation of applications for smart grid decision-making. 2. Investigating the generalizability of the methods to a broader range of data dimensions.

References

[1] A. Olabi and M. A. Abdelkareem, "Renewable energy and climate change," *Renewable and Sustainable Energy Reviews*, vol. 158, p. 112111, 2022.

[2] Z. Wu, J. Wang, M. Zhou, Q. Xia, C.-W. Tan, and G. Li, "Incentivizing frequency provision of power-to-hydrogen toward grid resiliency enhancement," *IEEE Transactions on Industrial Informatics*, 2022.

[3] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1322–1330, 2015.

[4] T. Adefarati and R. C. Bansal, "Reliability, economic and environmental analysis of a microgrid system in the presence of renewable energy resources," *Applied energy*, vol. 236, pp. 1089–1114, 2019.

[5] J. Wang, F. Gao, Y. Zhou, Q. Guo, C.-W. Tan, J. Song, and Y. Wang, "Data sharing in energy systems," *Advances in Applied Energy*, vol. 10, p. 100132, 2023.

[6] S. M. B. C. D. R. Taleb, I., "Big data quality framework: a holistic approach to continuous quality management," *Journal of Big Data*, vol. 8, 2021.

[7] N. Gupta, S. Mujumdar, H. Patel, S. Masuda, N. Panwar, S. Bandyopadhyay, S. Mehta, S. Guttula, S. Afzal, R. Sharma Mittal, *et al.*, "Data quality for machine learning tasks," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 4040–4041.

[8] 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.

[9] C. Cichy and S. Rass, "An overview of data quality frameworks," *IEEE Access*, vol. 7, pp. 24 634–24 648, 2019.

[10] M. Yu, J. Wang, J. Yan, L. Chen, Y. Yu, G. Li, and M. Zhou, "Pricing information in smart grids: A quality-based data valuation paradigm," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3735–3747, 2022.

[11] B. Wang, Q. Guo, T. Yang, L. Xu, and H. Sun, "Data valuation for decision-making with uncertainty in energy transactions: A case of the two-settlement market system," *Applied Energy*, vol. 288, p. 116643, 2021.

[12] C. Goncalves, P. Pinson, and R. J. Bessa, "Towards data markets in renewable energy forecasting," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 533–542, 2020.

[13] A. Ghorbani and J. Zou, "Data shapley: Equitable valuation of data for machine learning," in *International conference on machine learning*. PMLR, 2019, pp. 2242–2251.

[14] J. Yoon, S. Arik, and T. Pfister, "Data valuation using reinforcement learning," in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, H. D. III and A. Singh, Eds., vol. 119. PMLR, 13–18 Jul 2020, pp. 10 842–10 851.

[15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[16] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[17] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on lightgbm," *Finance Research Letters*, vol. 32, p. 101084, 2020.

[18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[19] T. Zhao, H. Hachiya, G. Niu, and M. Sugiyama, "Analysis and improvement of policy gradient estimation," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[20] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International conference on machine learning*. PMLR, 2015, pp. 1889–1897.

[21] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[22] L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, and A. Madry, "Implementation matters in deep policy gradients: A case study on ppo and trpo," *arXiv preprint arXiv:2005.12729*, 2020.

[23] X. Lu, M. B. McElroy, W. Peng, S. Liu, C. P. Nielsen, and H. Wang, "Challenges faced by china compared with the us in developing wind power," *Nature Energy*, vol. 1, no. 6, pp. 1–6, 2016.

Yanzhi Wang received his B.S. degree in Theoretical and Applied Mechanics from Peking University, Beijing, China, in 2022, where he is pursuing his Ph.D. degree with the Department of Industrial Engineering and Management.

His research interests include uncertainty estimation, data-driven optimization and data interpretation under complex systems.

Jie Song (Senior Member, IEEE) received her B.S. degree in Applied Mathematics from Peking University, Beijing, China, in 2004 and her Ph.D. degree in Management Science and Engineering from Tsinghua University, Beijing, in 2010. She is currently a full professor (with tenure) in the Department of Industrial Engineering and Management at Peking University and honored as the Chang Jiang Scholar by China's Ministry of Education.

Her research interests include stochastic simulation and optimization, and online algorithm design, with applications in resource allocation of complex service systems.