Applied Energy 353 (2024) 122168

Contents lists available at ScienceDirect

Applied Energy

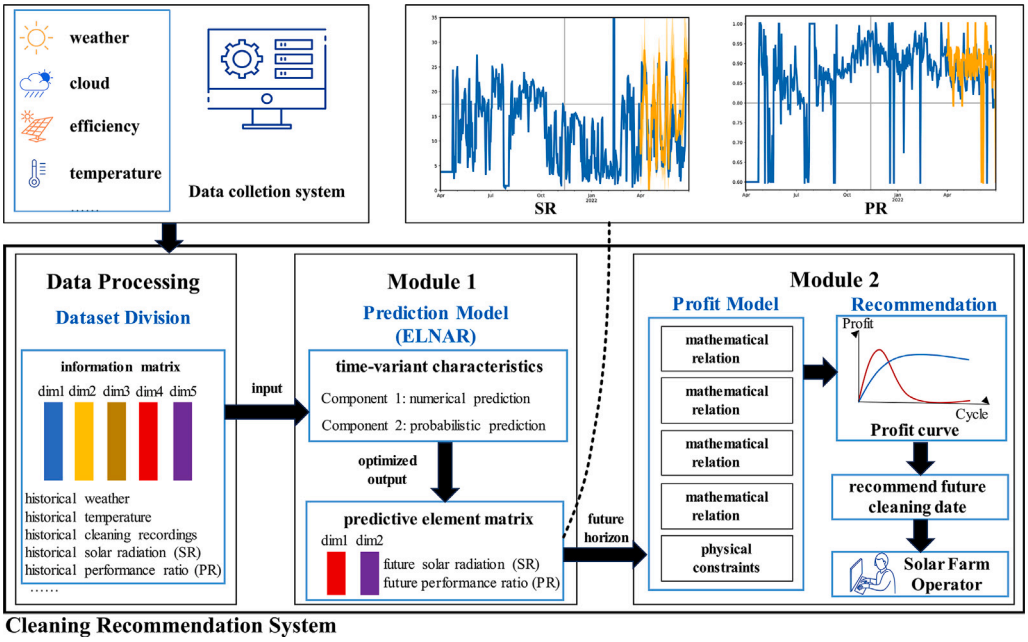
journal homepage: www.elsevier.com/locate/apenergy

A rolling-horizon cleaning recommendation system for dust removal of industrial PV panels

Chao Zhang [a,*], Yunfeng Ma [a,b], Zengqiang Mi [b], Fan Yang [c], Long Zhang [a]

[a] Department of Electrical and Electronic Engineering, University of Manchester, UK
[b] Department of Electrical and Electronic Engineering, North China Electric Power University, China
[c] Goldwind Science &, Technology Co., Ltd., China

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Renewable energy
Dust removal
PV panel
Rolling-horizon
Nonlinear autoregressive model
Recommendation system

ABSTRACT

Periodic cleaning of photovoltaic (PV) panels, such as every three months, is a common industry practice. However, this fixed period may not be optimal for maximizing the profit of a PV power generation system, due to numerous time-variant influencing factors, such as weather, temperature, etc. To increase the overall profit of a solar farm, it is highly desirable to have a flexible cleaning schedule that considers time-variant influencing factors. For this requirement, a rolling-horizon cleaning recommendation system is presented in this paper. Within this cleaning recommendation system, a prediction model and profit model are proposed. The prediction model, called the ensemble long-term and nonlinear autoregressive, can provide a time-variant future horizon by analyzing and compressing the time-variant characteristics in historical information. The profit model based on mathematical constraints, can process time-variant future horizon output from prediction model to

* Corresponding author.
E-mail address: chao.zhang-3@manchester.ac.uk (C. Zhang).

https://doi.org/10.1016/j.apenergy.2023.122168
Received 25 April 2023; Received in revised form 21 September 2023; Accepted 23 October 2023
Available online 31 October 2023

generate a flexible optimized recommendation for cleaning schedule. The effectiveness of the proposed system is validated in real farms and all data used in this paper is collected from real world. The two case studies in experiments show that the profit improvement can reach up to 6% and 30%, respectively.

Nomenclature

Abbreviations

AR	AutoRegressive
BAL	Bayesian Augmented Lagrange
ELNAR	Ensemble Long-term and Nonlinear Autoregressive
LSTM	Long Short-Term Memory
NARX	Nonlinear Autoregressive Model with Input
PR	Performance Ratio
PV	PhotoVoltaic
RNN	Recurrent Neural Network
SR	Solar Radiation
SVM	Support Vector Machine

Variables and Symbols

$\mathbb{A}$	Project capacity
$\mathbb{C}$	Power cost
$\mathbb{C}_r$	Cleaning recordings
$\mathbb{F}_1$	SR value in cleaning recommendation system
$\mathbb{F}_2$	PR value in cleaning recommendation system
$\mathbb{I}$	Power earnings
$\mathbb{P}_c$	Cleaning price per unit capacity
$\mathbb{P}_e$	Electricity price
$\mathbb{T}$	Temperature
$\mathbb{W}$	Weather information
$\mathbb{W}_c$	Cloudiness label value
$\mathbb{W}_{fg}$	Fog label value
$\mathbb{W}_{he}$	Haze label value
$\mathbb{W}_{hl}$	Hail label value
$\mathbb{W}_{hn}$	Heavy rain label value
$\mathbb{W}_{ln}$	Light rain label value
$\mathbb{W}_{mn}$	Moderate rain label value
$\mathbb{W}_{se}$	Sunshine label value
$\mathbb{W}_{sm}$	Sandstorm label value
$\mathbb{W}_{sw}$	Snow label value
$\mathbb{X}$	Theoretical power generation
$\mathbf{o}$	Output from optimized prediction
$\mathbf{y}$	Output from numerical prediction
$\mathbf{z}$	Output from probabilistic prediction

1. Introduction

Solar power, as one of the most promising options, enjoys non-pollution, sustainability, and low cost [1]. The annual installation of photovoltaic (PV) panels has increased significantly worldwide in the past decade [2]. Prior studies have indicated that dust accumulation on the cover glass of PV modules significantly reduces efficiency value of PV energy conversion [3]. Huge levels of dust particles are often found in arid areas [4], so a typical reduction in output power due to dusty panels is around 10%, but in some hostile environments, such as sandy and dry regions, this reduction can reach 30%–45% [5]. This physical truth indicates that cleaning the PV panel to remove dust is very crucial for some places.

In industry, the cleaning schedule for PV panels is primarily determined based on engineers' experience [6]. The experience-oriented decisions may not be the ideal solutions for the operation of solar farms and cannot guarantee the maximal profit for a solar farm. In recent years, some research on cleaning frequency of dust on the PV modules has been conducted [7–9]. These methods ignore the time-varying nature of the influencing factors to obtain a fixed period. However, this fixed period may not be optimal for maximizing the profit of a PV power generation system. Besides, it is inevitable that some unexpected events would occur during the regular operation of solar farms, and the routine execution of fixed cleaning schedule would be interrupted.

Therefore, it is necessary to provide a flexible cleaning schedule which could take into account the time-variant characteristics of the influencing factors, such as weather, temperature, etc. To realize this target, this paper focuses on a cleaning recommendation system to generate the flexible cleaning schedule. The importance of prediction and economics in planning are highlighted in [10,11]. Inspired by this, the prediction model and the profit model would be the significant parts of our proposed cleaning recommendation system.

In terms of the prediction methods for predicting solar energy and modeling profits, the prevailing approaches put into applications can be grouped into two categories: statistical method and deep learning method [12].

The statistical method refers to those approaches utilizing the characteristics of mathematical statistics. For example, Amato et al. [13] employed Markov chain and Fourier analysis to describe and simulate daily solar irradiance, considering the state transition probability among different timestamps. Reikard and Gordon [14] utilized autoregressive moving average to make the high-resolution prediction for solar power by comparing several time series forecasting methods. Prema and Uma Rao [15] used decomposition enhanced moving average model to predict solar power, improving the forecast comprehensiveness by considering seasonal and trend patterns. Li et al. [16] created a hybrid enhanced multi-verse optimizer method to optimize the support vector machine (SVM) for forecasting PV output, improving the inner structures of traditional SVM method. However, the shortcoming of statistical method is that long-term prediction accuracy cannot be ensured by preferring to describe the local variable relationships affecting output in complex energy systems.

The deep learning method utilizes the neural networks to learn the inner characteristics of time series, depending on strong computing resources. For instance, Zheng et al. [17] used long short-term memory (LSTM) model with particle swarm optimization to predict solar power, which can help the transitional LSTM method determine the best hyperparameters intelligently. Marzouq et al. [18] proposed an evolutionary artificial neural network (ANN) to predict solar irradiance, which can use the development of the forecasting history and ANN architecture to produce multiple models for various time horizons up to six hours in advance. Qing and Niu [19] applied adaptive LSTM networks into the prediction of hourly day-ahead solar irradiance, considering the relationship between the subsequent hours on the same day. Leonardo and Zaher [20] proposed a hybrid intelligent method to build robust predictive models for daily span solar irradiation, which can be utilized as an auxiliary tool for predicting solar irradiation on the stations. However, the drawback of deep learning method is that short-term predictions cannot be well made by primarily learning the global characteristics hidden in complex energy systems.

Additionally, a suitable profit model, also another crucial functional module in our proposed cleaning recommendation system, can generate reasonable scheduling results. Among the existing papers [21–23], most

of these profit modeling methods are intricate and idealistic as numerous influencing factors in their constructs are challenging to obtain in real-world settings, which results in difficult industrial deployment.

Hence, a novel ensemble long-term and nonlinear autoregressive (ELNAR) model is proposed as the prediction model in the recommendation system. A mathematical profit model that primarily takes into account two measurable factors, namely solar radiation (SR) and performance ratio (PR). The SR and PR predictions generated by the proposed prediction model, ELNAR, can be utilized directly within the profit model, thereby creating a seamless connection between the two models.

The original contributions of this paper are as follows: (1) A cleaning recommendation system is presented to recommend reasonable cleaning schedule for solar farm operators. The composition of this system is clear, including model construction for prediction and profit, and it is convenient for industrial deployment. (2) A novel long-term and nonlinear prediction model, namely ELNAR, is proposed to provide a time-variant future horizon by analyzing and compressing the time-variant characteristics in historical information. This model combines the advantages of precise short-term description in statistical methods and accurate long-term forecasting in deep learning methods. (3) A mathematical profit model is proposed to process time-variant future horizon output from the prediction model to generate a profit curve, where the two measurable time-variant factors SR and PR are regarded as the future horizon, and flexible optimized recommendations for cleaning schedules could be obtained according to the profit curve. This practice could reduce the complexity of profit model construction. (4) This cleaning recommendation system is tested effective in real farms and all data used in this paper is collected from real world. To the best of the authors' knowledge, this is presumably the first work to develop a cleaning recommendation system for solar farms.

The rest of this paper is organized as follows. Section 2 models the effect of multiple parameters on PV module performance. In Section 3, the theoretical process of the cleaning recommendation system, related to the prediction model, and the profit model are explained in detail. Then Section 4 is dedicated to the analysis of the forecasting performance and expected profit results. Section 5 offers some detailed analysis and discussion for this system from the PV side. Finally, Section 6 concludes the current work.

2. Effect of multiple parameters on PV module performance

Various factors impact the performance of PV modules [24], resulting in time-varying changes in their output. This section primarily discusses transmittance information from three different aspects: (1) the effect of material parameters, primarily the cover glass and encapsulant; (2) the effect of electric parameters of solar cells; (3) the effect of PV tilt angle on SR; (4) the effect of dust accumulation on power output. To account for the combined effects of these aspects, corresponding scaling coefficients are introduced.

2.1. Effect of material parameters on PV module power output

The main components of a PV panel include cover glass, the encapsulant, frame, solar cells and so on. This subsection analyzes the effect of cover glass and encapsulant on PV module performance. The material information of the PV modules used in this study is given in Appendix A. Table A.6.

The effect of dust accumulation on PV power output varies among different types of cover glass. Ref. [25] discusses the potential of using anti-reflective (AR) coated glasses in minimizing dust fouling, and compares the superiority against the traditional tempered glass. Ref. [25] analyses multiple kinds of PV module cover glass. Here we focus on two of them, i.e., without and with AR-coated glass. Fig. 1 presents how the power output changes along the number of weeks of exposure to the outside environment, where the solid curve is the experimental results and the dashed lines are the fitted curves $\kappa_{glass}(t)$. Hence, the effect of cover glass material could be represented as a time-varying coefficient on PR:

$$\kappa_{glass} = \begin{cases} f_{AR}(t), \text{if cover type} = \text{AR} \\ f_{Tempered}(t), \text{if cover type} = \text{Tempered} \end{cases} \quad (1)$$

Encapsulant materials play a crucial role in the performance and lifespan of PV panels. Different encapsulant materials possess distinct characteristics. Ethylene vinyl acetate (EVA), a copolymer, was commonly used in the majority of PV modules. In Ref. [26], various EVA encapsulant materials were tested under 42 UV suns at temperatures ranging from 80 °C to 95 °C. Fig. 2 illustrates that transmittance coefficient κ_{eva} degrades over exposure time. Notably, there is a turning point in transmittance for EVA material after a certain duration of UV exposure. The PV farms selected for this study have relatively short operational periods, and we assume that the EVA material performs well, with a transmittance coefficient of approximately $\kappa_{eva} = 0.87$.

In terms of frame design, Ref. [27] introduced an approach for a comprehensive improvement of PV module frames and analyzed the influence of the width of the front frame overlap on the normalized power of PV modules. The simulated and measured results are presented in Fig. 3. It can be observed that the normalized power decreases slightly as the width of the front frame overlap increases. This influence on the normalized power is introduced as a frame scaling coefficient κ_{frame}. Compared to a frameless structure, the normalized power decreases to approximately 0.99 when the width is around 11 mm–12 mm. Additionally, the frame material also affects the power output by influencing the PV module temperature. While this impact is not analyzed in detail here, it will be comprehensively considered in the method outlined in Section 3 when addressing time-varying factors.

2.2. Effect of electric parameters on PV module performance

The fill factor (FF) is an important parameter for assessing the output performance of PV modules. A larger fill factor indicates higher conversion efficiency and higher quality in PV panels. Typically, the fill factor (FF) falls within the range of 60% to 85%, with its value determined by the materials and device structure of the solar cell. Hence, we choose FF as the scaling coefficient κ_{ff} describing the electric performance, which could be expressed as

$$\kappa_{ff} = FF = \frac{P_{\text{MPP}}}{I_{\text{sc}} V_{\text{oc}}} = \frac{V_{\text{MPP}} I_{\text{MPP}}}{I_{\text{sc}} V_{\text{oc}}} \quad (2)$$

where P_{MPP} is the nominal maximum power of the solar cell. V_{MPP} is the voltage at the maximum power. I_{MPP} is the current at the maximum power. I_{sc} is the short current. V_{oc} is the open voltage. The electric information of the PV modules this study is given in Appendix A. Table A.7.

2.3. Effect of PV panel tilt angle on SR efficiency

The appropriate PV panel tilt angle is significant for improving the SR of PV panels. Due to variations in the geographical locations where PV panels are installed, the best tilt angle may be different. In general, the best tilt angle for PV panels is set to the latitude of the location. However, in practical engineering applications, actual tilt angles are often uniformly set between 15 degrees and 20 degrees. This results in a deviation between the actual PV panel tilt angle δ and best PV panel tilt angle δ^*, the schematic of effect of PV panel tilt angle on the solar radiation is shown in Fig. 4. In our cases, to analyze the effect of tilt angle on the SR, we use an approximation method to estimate the SR loss due to the deviation of δ, which is expressed as:

$$\kappa_{angle} = sin\left(\frac{\pi}{2} - (\delta - \delta^*)\right) \quad (3)$$

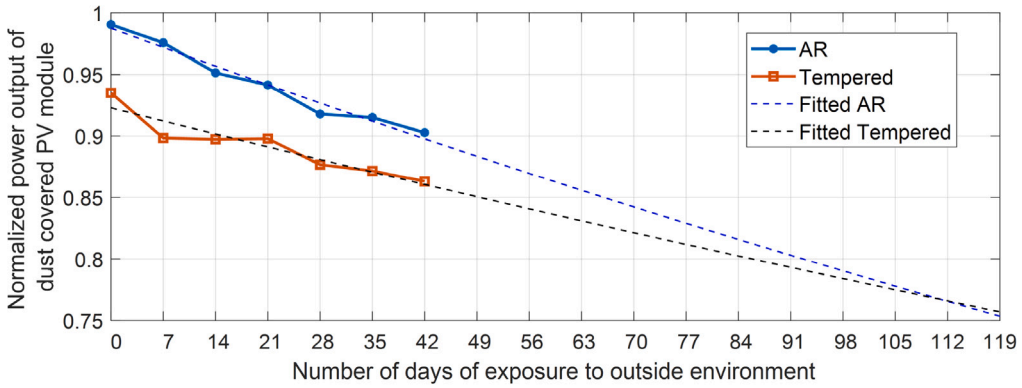


Fig. 1. The effect of cover glass on the dust covered PV module.

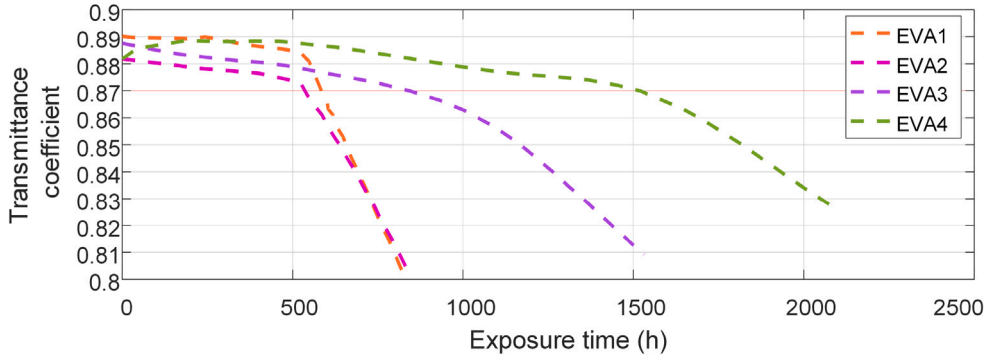


Fig. 2. Transmittance coefficient for different EVA materials with exposure to 42 UV suns of radiation.

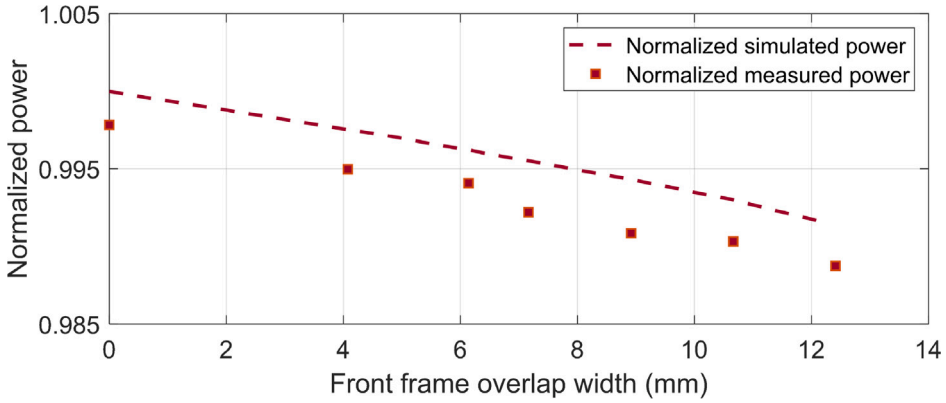


Fig. 3. Normalized measured (square points) and simulated (line) module power for different front frame overlap widths.

2.4. Effect of dust accumulation on transmittance

To quantify the impact of dust accumulation on the transmittance of PV modules, Ref. [28] proposed a mathematical model to predict the effect of dust accumulation based on the shape of dust particles, primarily modeling two shapes: spherical and cubic. The transmittance coefficient can be modeled as:

$$\frac{\tau}{\tau_0} = \begin{cases} \kappa_c = \exp\left(-\frac{0.6\sqrt{2}M}{4\rho\pi r}\right), 0 \leqslant i \leqslant 90 \\ \kappa_s = \exp\left(-\frac{0.2M}{\rho\sqrt{\pi r}}\right), n_1 = n_2 = 0 \end{cases} \tag{4}$$

where r is the mean radius of deposition dust, τ is the transmittance after dust accumulation and τ_0 is the theoretical transmittance. M is equivalent to the dust-deposited density, ρ is the density of a single dust particle. i is the refraction angle. n_1 and n_2 are the refractive index of air and refractive index of the transparent device, respectively. κ_c represents the transmittance ratio when the dust particle shape is cubic, and κ_s represents the transmittance ratio when the dust particle shape is spherical. Fig. 5 shows how the transmittance ratio changes with varying dust deposition density, using radii of 50 μm and 75 μm as examples. Since the specific shapes of dust particles were not measured in this study, the mean values of κ_c and κ_s are adopted, which is

$$\kappa_{dust} = 0.5(\kappa_c + \kappa_s) \tag{5}$$

3. Cleaning recommendation system for PV panels

3.1. Generic structure

This section provides a generic structure of the proposed recommendation system that guides the realization of the recommendation for

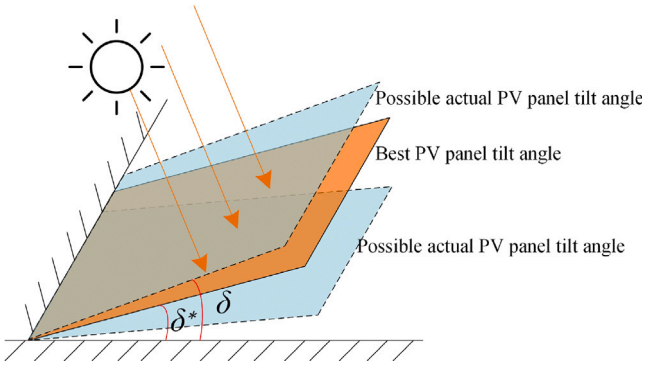


Fig. 4. Schematic of PV tilt angle deviation.

cleaning schedule of PV panels. As can be seen in Fig. 6, it consists of data processing, module 1 and module 2. The prediction model in module 1, namely ELNAR, uses the historical information in information matrix to forecast the future horizon. The future horizon, i.e., the output of the prediction model, uses future SR and PR in the process of implementation. Subsequently, the future horizon can be processed by the profit model in module 2, which utilizes multiple f under some constraints to generate a profit curve. The recommendation for flexible cleaning schedules can be ultimately provided according to the profit curve.

Fig. 7 shows a rolling-horizon process for the proposed cleaning recommendation system, which rolls with each day. The rolling interval could be set according to the requirements of solar farm operators. Within each horizon, profits from various cleaning cycles would be summarized as an ultimate profit curve. The flexible cleaning cycle could be obtained by analyzing profit curves and the next recommended cleaning date can be inferred by the given cycle.

3.2. Proposed prediction model ELNAR

The prediction model, playing an important role in the cleaning recommendation system, influences the recommendation for cleaning schedule of PV panels. The inner details of the prediction model is presented in Fig. 6 module 1. Firstly, the dataset division would be the preliminary work, and the information matrix could be prepared for prediction. Subsequently, the numerical component 1 and probabilistic component 2, representing short-term statistical and long-term deep learning method, would process data independently to generate intermediate results. Finally, the intermediate results from two components are optimized and fused to obtain the predictive element matrix that would be regarded as future horizon to enter the profit model for recommendation. The two predictive components and their working principles can be concretely explained in the following text.

3.2.1. Predictive component 1: numerical prediction

The historical observation and future prediction time range are referred to as $[1,\ t_0-1]$ and $[t_0,\ T]$, respectively. The vectors of historical observation and future values of time series can be denoted as follows:

$$\mathbf{y}_{1:t_0-1} = \left[y(1), y(2), \ldots, y(t_0-1)\right] \tag{6}$$

$$\mathbf{y}_{t_0:T} = \left[y(t_0), y(t_0+1), \ldots, y(T)\right] \tag{7}$$

The output value y at time t can be named as $y(t)$. To construct the model, t here should be within the range $[1,\ t_0-1]$ to estimate the inner parameters of the model. Covariates are those input features that do not depend on time and are always multiple in this paper, so we can define the value of covariates i at time t as $x^i(t)$. Finally, a nonlinear autoregressive model with multiple covariates x can be represented as follows [29]:

$$\begin{aligned} y(t) &= f\left(y(t-1), \ldots, y\left(t-n_y\right), x^0(t), \ldots, x^0\left(t-n_{x^0}\right), x^1(t), \ldots, x^1\left(t-n_{x^1}\right)\right. \\ &\quad \left., \ldots, x^m(t), \ldots, x^m\left(t-n_{x^m}\right)\right) + \xi(t) \\ &= f(\mathbf{u}(t)) + \xi(t) \end{aligned} \tag{8}$$

where the output value y at time t can be mapped by the former output value and multiple covariates before time t. The covariates can be $\{x^0, x^1, \ldots, x^m\}$, and their total number should be determined by the requirements of the real application. n_y indicates the maximum lag of output. $\{n_{x^0}, n_{x^1}, \ldots, n_{x^m}\}$ denotes the maximum lag of covariates, respectively. $\xi(t)$ denotes the error at time t.

Suppose the model input is $\mathbf{u}(t) = \left[y(t-1), \ldots, y\left(t-n_y\right), x^0(t), \ldots, x^0\left(t-n_{x^0}\right), x^1(t), \ldots, x^1\left(t-n_{x^1}\right), \ldots, x^m(t), \ldots, x^m\left(t-n_{x^m}\right)\right]$. The general nonlinear autoregressive model with input (NARX) can be represented as a linear combination of the nonlinear functions with the polynomial format:

$$y(t) = \sum_{i=1}^{M} p_i(\mathbf{u}(t))\Theta(i) + \xi(t) \tag{9}$$

The formula (9) can be further expressed in the matrix form:

$$\mathbf{y} = \boldsymbol{P\Theta} + \boldsymbol{\xi} \tag{10}$$

where vector $\mathbf{y}$ denotes the output value, $\mathbf{y} = [y(1), y(2), \ldots, y(N)]^T$. The $\mathbf{y}$ here only considers time range within $[1, t_0-1]$, namely $N \le t_0-1$, to determine suitable inner parameters constructing the model. $\boldsymbol{\Theta} = [\Theta(1), \Theta(2), \ldots, \Theta(M)]^{\mathrm{T}}$ represents the parameter being estimated. The $\boldsymbol{\xi}$ demonstrates the stochastic error during the process of model approximation, $\boldsymbol{\xi} = [\xi(1), \xi(2), \ldots, \xi(N)]^T$.

Recently, sparse Bayesian learning [30] has been developed under the Bayesian framework, which can calculate the solutions iteratively. The detailed solutions could be seen in Appendix B.

Then, obtained parameters Θ in NARX model can be used to make the future prediction $\mathbf{y}_{t_0:T}$. Assume the predictive output calculated with estimated Θ is named as $\widetilde{y}$. As shown in Fig. 8, the Bayesian augmented Lagrange (BAL) algorithm, as significant sub-module, can provide the adjusted parameters Θ to predictive output $\widetilde{y}$, so that the future prediction $\mathbf{y}_{t_0:T}$ can be inferred subsequently.

3.2.2. Predictive component 2: probabilistic prediction

The historical observation and future prediction time range are referred to as $[1,\ t_0-1]$ and $[t_0,\ T]$, respectively. Historical observation and future values of time series can be denoted as follows:

$$\mathbf{z}_{1:t_0-1} = \left[z(1), z(2), \ldots, z(t_0-1)\right] \tag{11}$$

$$\mathbf{z}_{t_0:T} = \left[z(t_0), z(t_0+1), \ldots, z(T)\right] \tag{12}$$

Referring to the method Deep AR [31], the purpose here is to obtain the condition distribution, namely the distribution of z_t as the value at time t, as follows:

$$P\left(\mathbf{z}_{t_0:T} \mid \mathbf{z}_{1:t_0-1}, \mathbf{X}_{1:T}\right) \tag{13}$$

where $\mathbf{X}_{1:T} = \left[x^0(1), x^0(2), \ldots, x^0(T); x^1(1), x^1(2), \ldots, x^1(T); \ldots; x^n(1), x^n(2), \ldots, x^n(T)\right] = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(T)]$ are the $n+1$ covariates that are known from the history to future all the time. The $x^i(t)$ represents covariate i at time t and $\mathbf{x}(t)$ indicates all the covariates at time t.

Unlike the traditional prediction model getting predictive value directly, the primary idea of this model is to obtain a predictive likelihood, which can be subsequently sampled to get predictive value. Firstly, the model distribution is denoted as $Q_{\Theta'}\left(\mathbf{z}_{t_0:T} \mid \mathbf{z}_{1:t_0-1}, \mathbf{X}_{1:T}\right)$,

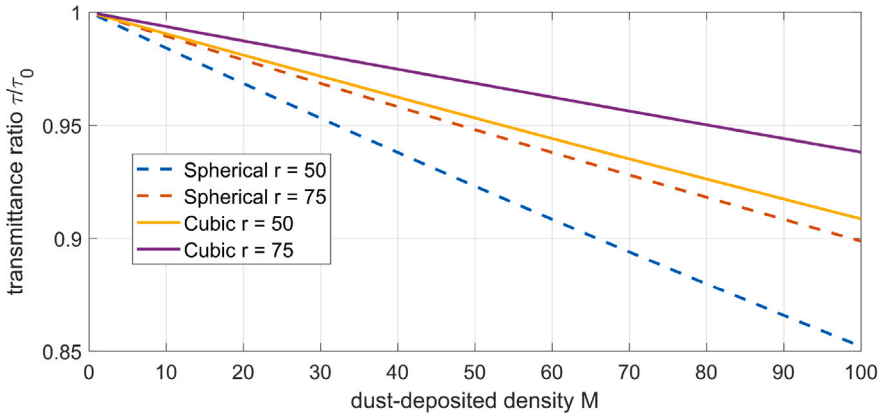


Fig. 5. The effect of mean particle size on transmittance.

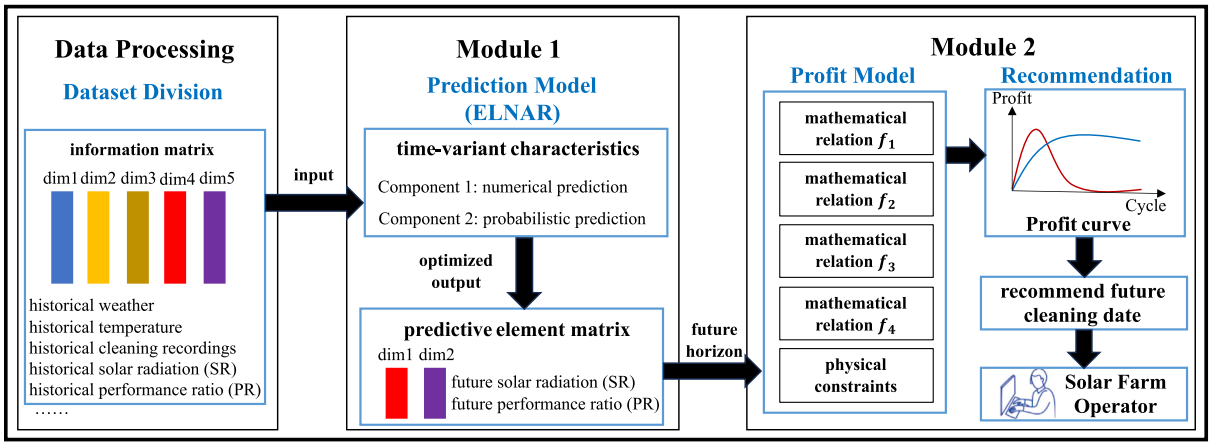


Fig. 6. Generic structure of the proposed cleaning recommendation system.

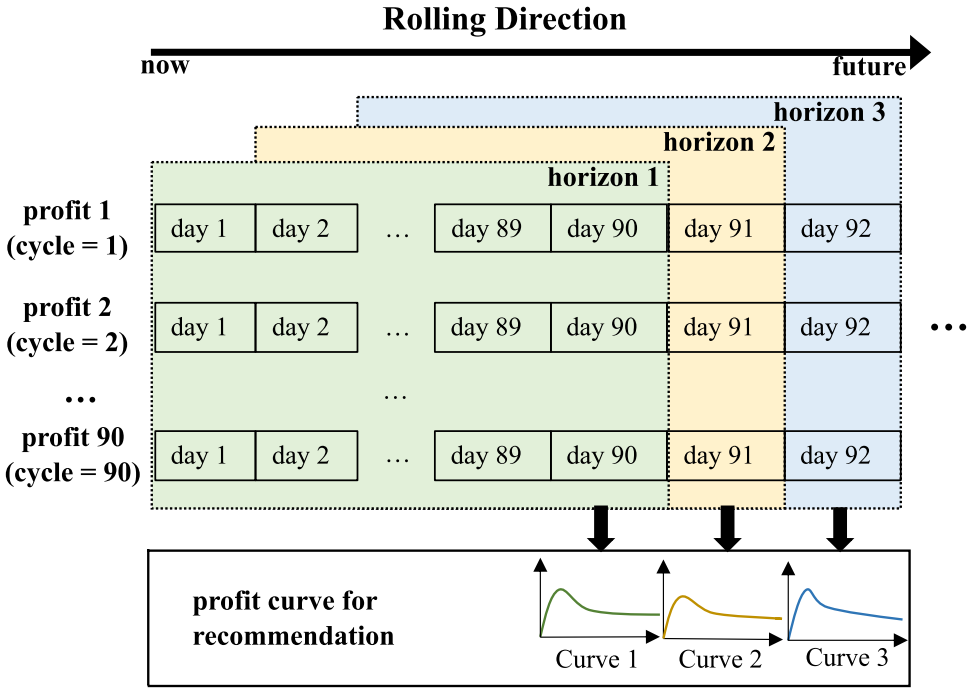


Fig. 7. Diagram of rolling-horizon process for cleaning recommendation system with daily rolling operation.

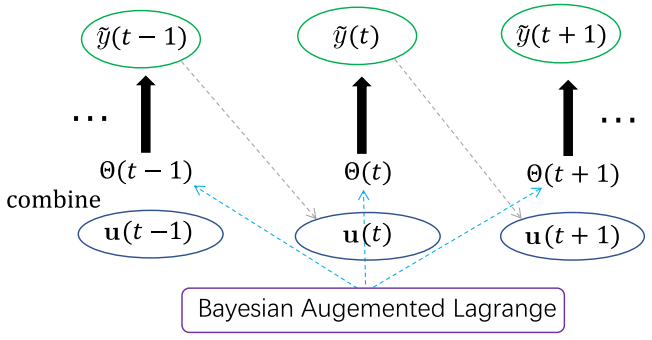


Fig. 8. Summary of the NARX model with Bayesian augmented Lagrange analysis.

consisting of many likelihood factors:

$$\begin{aligned} &Q_{\Theta'}\left(\mathbf{z}_{t_0:T} \mid \mathbf{z}_{1:t_0-1}, \mathbf{X}_{1:T}\right) \\ &= \prod_{t=t_0}^{T} Q_{\Theta'}\left(z_t \mid \mathbf{z}_{1:t-1}, \mathbf{X}_{1:T}\right) \\ &= \prod_{t=t_0}^{T} \ell\left(z_t \mid \theta\left(h_t, \Theta'\right)\right) \end{aligned} \tag{14}$$

where $h(t) = g\left(h(t-1), z(t-1), \mathbf{x}(t), \Theta'\right)$, which is the output of a neural network with mapping function $g(\cdot)$. The neural network in this paper can be multi-layer recurrent neural network (RNN), such as LSTM or GRU. Θ' is the weight parameters in the neural networks.

This paper uses the Gaussian distribution as the expected likelihood, ℓ in (14) can be represented as ℓ_{G}, with intrinsic parameters $\theta = (\mu, \sigma)$:

$$\ell_{\mathrm{G}}(z \mid \mu, \sigma) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-(z-\mu)^2 / \left(2\sigma^2\right)\right) \tag{15}$$

where the mean and standard deviation can be updated exactly as $\mu(h(t)) = w_\mu h(t) + b_\mu$ and $\sigma(h(t)) = \log\left(1 + \exp\left(w_\sigma h(t) + b_\sigma\right)\right)$. $h(t)$ represents the output of RNN networks. w_μ and w_σ are the weight for calculation of μ and σ respectively. b_μ and b_σ are the bias adding to μ and σ respectively.

Note that the θ is the parameters of Gaussian distribution, including mean and variance. Θ' is the parameter of neural networks, covering weight parameters which are need to be updated through iterations. When the parameter Θ' is determined, the joint samples $\mathbf{z}_{t_0:T} \sim Q_{\Theta'}\left(\mathbf{z}_{t_0:T} \mid \mathbf{z}_{1:t_0-1}, \mathbf{X}_{1:T}\right)$ can be obtained subsequently, and then the future prediction series $\mathbf{z}_{t_0:T}$ can be acquired by sampling. As can be seen in Fig. 9, the neural networks h here just output a likelihood ℓ, and the real predictive values can be obtained by sampling.

3.2.3. Optimized predictive output

In this section, $\mathbf{y}_{t_0:T}$ in (7) and $\mathbf{z}_{t_0:T}$ in (12) will construct an ensemble and optimized output $\mathbf{o}_{t_0:T}$, where $\mathbf{o}_{t_0:T} = \left[o(t_0), o(t_0+1), \ldots, o(T)\right]$. Referring to the idea of kalman filter [32,33], it can be assumed that the current state value $o(t)$ depends on the previous value $o(t-1)$ and correlated variable $s(t)$. $o'(t)$ can be regarded as the observed value that is obtained from observation model. The $s(t)$ can be the temperature or cleaning recordings in this paper, which should be the significant factors influencing the current state $o(t)$. The ideal system can be defined as follows:

$$o(t) = Ao(t-1) + Bs(t) + w(t) \tag{16}$$

$$o'(t) = Ho(t) + v(t) \tag{17}$$

where w and v are the normally distributed noise with 0 mean value. The covariance matrixes of w and v are Q and R, respectively. A is the transferring matrix, which can be simplified as identity matrix here. B is the correlated matrix that control the input weight of correlated variable. H is the measurement matrix and the elements in H can range from 0 to 1.

The predicted value $y(t)$ in $\mathbf{y}_{t_0:T}$ will be adjusted by observed value $z(t)$ in $\mathbf{z}_{t_0:T}$, and the final optimized value $\hat{o}(t)$ can output as the superior prediction, described in (19). The complete adjusted process is shown below, including the updating of $\hat{o}(t)$, $P(t)$, and $P'(t)$.

$$\begin{aligned} y(t) &= A\hat{o}(t-1) + Bs(t) \\ P'(t) &= AP(t-1)A^T + Q \end{aligned} \tag{18}$$

$$\begin{aligned} \hat{o}'(t) &= o'(t) - Hy(t) = z(t) - Hy(t) \\ K(t) &= P'(t)H^T\left(HP'(t)H^T + R\right)^{-1} \\ \hat{o}(t) &= y(t) + K(t)\hat{o}'(t) \end{aligned} \tag{19}$$

$$P(t) = (I - K(t)H)\,P'(t) \tag{20}$$

where $P(t)$ and $P'(t)$ are measurement covariance error matrix and prediction covariance error matrix, respectively. $K(t)$ is the gain matrix. $\hat{o}'(t)$ is the measurement residual. I is the identity matrix.

The error can be the entry point to verify the effectiveness of this adjusted process, namely (18) to (20). The whole proof process can be seen in Appendix C.

3.3. Proposed mathematical profit model

The profit model is used to calculate the expected profit curve, and then the recommendation for the cleaning schedule could be obtained. To reduce the complexity of the profit model, this paper here did not consider the energy storage systems in the profit models. Additionally, the time-variant factors into the profit model are future SR and PR, which have already been predicted by our proposed ELNAR model. Subsequently, the detailed mathematical profit model can be defined as follows.

Define theoretical power generation as $\mathbb{X}$, power earnings as $\mathbb{I}$, cost as $\mathbb{C}$, profit as $\mathbb{R}$. SR (MJ/m^2) and PR are defined as $\mathbb{F}_1$ and $\mathbb{F}_2$. Then project capacity is $\mathbb{A}$ (MW) and electricity price $(\cdot/kWh)$ is $\mathbb{P}_e$ and cleaning price per unit capacity $(\cdot/MW)$ is $\mathbb{P}_c$.

$$\mathbb{X} = f_\kappa(\kappa_{dust} * \kappa_{angle} * \kappa_{ff} * \kappa_{glass} * \kappa_{eva} * \kappa_{frame}) * \mathbb{F}_1 * \mathbb{F}_2 * \mathbb{A} * 1000/3.6 \tag{21}$$

where f_κ is the function that maps the scaling coefficients describing the effect of multiple parameters on PV module performance to the range [0, 1], it could be written as

$$f_\kappa(x) = \frac{2}{\sqrt{\pi}} \int_0^{2x} e^{-t^2}\,\mathrm{d}t, \quad x > 0 \tag{22}$$

$$\mathbb{I} = \mathbb{P}_e * \mathbb{X} \tag{23}$$

$$\mathbb{C} = \mathbb{P}_c * \mathbb{A} \tag{24}$$

$$\mathbb{R} = \mathbb{I} - \mathbb{C} \tag{25}$$

Through the calculations above, the profit curve, describing relationships between the expected profit and cleaning cycles, can be generated. The recommendation for flexible cleaning schedules can be ultimately provided according to the analysis of profit curve. This analysis should be done in a flexible way by the operators, taking into account things like the current operation situation, how the staff is set up, etc.

Note that it is only the simplified profit model. For example, the energy storage system, the use of the robotics, and the special cleaning reagent can all influence the profit model. This should be discussed separately and will not be expanded upon in this paper.

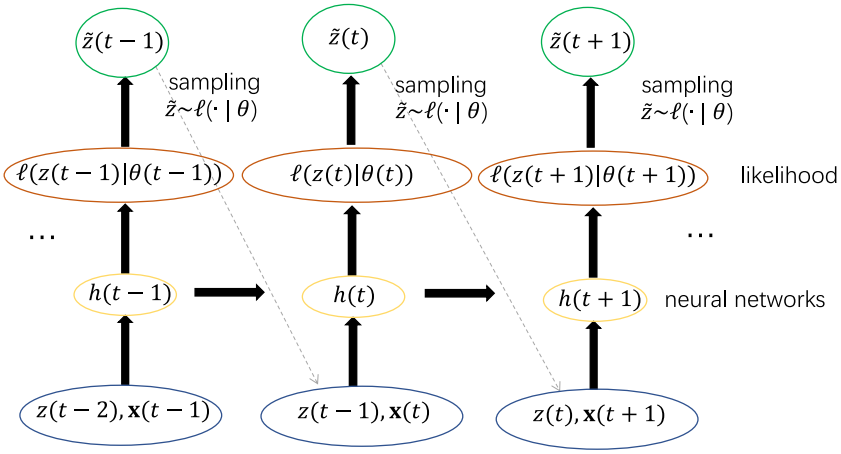


Fig. 9. Summary of the probabilistic inference with autoregressive recurrent networks.

4. Experiments

4.1. Data preparation

All the data presented in this paper was collected from actual solar farms. Within the farms, there exist various types of PV panels due to the project planning being carried out in phases, predominantly composed of polycrystalline silicon panels with a small quantity of monocrystalline silicon panels. Three datasets obtained from three different farms are employed in this paper. The dataset I has the long data length of a year and a half, to test the forecasting ability of our proposed ELNAR model. Considering the data in dataset II and dataset III are both acquired with only the time length of one year, the two datasets are employed as the test for cleaning recommendation. Excessively dense data is redundant and often necessitates downsampling, so the resolution of data within the dataset is measured in daily increments.

Regarding the features of the three datasets, the weather ($\mathbb{W}$), temperature ($\mathbb{T}$), cleaning recordings ($\mathbb{C}_r$), SR ($\mathbb{F}_1$), and PR ($\mathbb{F}_2$) are recorded information. The weather consists of sunshine ($\mathbb{W}_{se}$), cloudiness ($\mathbb{W}_c$), snow ($\mathbb{W}_{sw}$), light rain ($\mathbb{W}_{ln}$), moderate rain ($\mathbb{W}_{mn}$), heavy rain ($\mathbb{W}_{hn}$), sandstorm ($\mathbb{W}_{sm}$), haze ($\mathbb{W}_{he}$), hail ($\mathbb{W}_{hl}$), and fog ($\mathbb{W}_{fg}$). It could be seen that the rain is categorized into fine division, including light rain, moderate rain and heavy rain. This is due to the fact that rainwater, as a force of nature, can have a certain cleansing effect, significantly impacting the accumulation of dust on the surface of PV panels.

Considering the overlapped information of weather, the one-hot encoding [34] is applied to represent the significant weather information. The temperature here selects the average value of one day. The cleaning recordings mean that the days to execute cleaning procedures are labeled as 'true', and otherwise 'false' throughout the whole statistical process. SR and PR are two quantifiable factors that impact the power generation capacity of PV panels, with the former representing irradiation levels and the latter indicating conversion efficiency.

Regarding the expansion of the input data dimension, the proposed ELNAR model capitalizes fully on the capacity to capture short-term local details and the proficiency in grasping long-term global aspects inherent in statistical methods and deep learning approaches, respectively. This culminates in the effective extraction of information from various types of time-variant variables within the ELNAR model. This suggests that the ELNAR model's predictive performance improves with an increase in the dimensional input. Fundamentally, the model boasts a high degree of extensibility, indicating that any relevant input information that impacts or reflects the PR and the SR can be incorporated, based on the developer's discretion.

4.2. Definition of indicators

When it comes to this cleaning schedule optimization application, small fluctuations within a range are allowed. However, the traditional indicators may not effectively reflect the forecasting performance under this condition, so it is necessary to define some different indicators in this field. Three indicators (stationarity, precision, and acceptability) are defined. The stationarity can describe the overall bias between the real and predicted values. The precision here shows the prediction accuracy directly, but may exist the negative value. The value of precision nearer to 100% denotes the accuracy is higher. The acceptability is a significant reference indicator that reflects the percentage of the predictive points that are satisfied with the allowed error range.

$$stationarity = \left|\sum_{i=1}^{n}((V^i_{pred} - V^i_{real})/V^i_{real})\right|/n \times 100\% \tag{26}$$

$$precision = \sum_{i=1}^{n}(1 - \left|V^i_{pred} - V^i_{real}\right|/V^i_{real})/n \times 100\% \tag{27}$$

$$acceptability = \sum_{i=1}^{n} flag^i/n \times 100\% \tag{28}$$

where

$$flag^i = \begin{cases} 1, & \left|V^i_{pred} - V^i_{real}\right|/V^i_{real} < 20\% \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

where V_{pred} represents the predicted value, V_{real} is the real value, and i is the index value of the data.

4.3. Results and analysis

4.3.1. Forecasting performance

The forecasting performance of the proposed ELNAR model is compared with the original two methods NARX and Deep AR. They are the basis of ELNAR, corresponding to predictive component 1 and component 2, respectively. The forecasting process primarily includes two aspects, PR and SR, and this paper will compare and analyze them, respectively.

Results from three different methods are given in Table 1 with different performance indicators. The proposed ELNAR has the minimum stationarity of the prediction for both PR and SR, 2.16% and 45.1%, respectively. In terms of the precision comparison, the PR prediction using ELNAR obviously outperforms the Deep AR and the NARX. Finally, the acceptability for PR prediction using ELNAR can reach 98.8%.

Table 2 shows the detailed performance of the ELNAR for PR prediction and SR prediction, respectively. The settings of the prediction

Table 1
Forecasting comparison of different models for PR/SR.

Model type	Stationarity PR/SR (%)	Precision PR/SR(%)	Acceptability PR/SR(%)
NARX (%)	10.0%/55.3%	78.7%/35.7%	83.4%/39.6%
Deep AR (%)	3.32%/45.2%	90.2%/43.2%	91.3%/45.9%
ELNAR (%)	**2.16%/45.1%**	**95.7%/45.7%**	**98.8%/51.0%**

Table 2
Detailed ELNAR performance for PR/SR prediction.

Time span (days)	Stationarity PR/SR(%)	Precision PR/SR(%)	Acceptability PR/SR(%)
First 30 days	0.53%/11.7%	97.7%/67.7%	100%/76.6%
Second 30 days	0.52%/34.3%	96.4%/60.2%	100%/53.3%
Third 30 days	5.44%/89.4%	93.2%/9.24%	96.6%/23.3%
Average	2.16%/45.1%	95.7%/45.7%	98.8%/51.0%

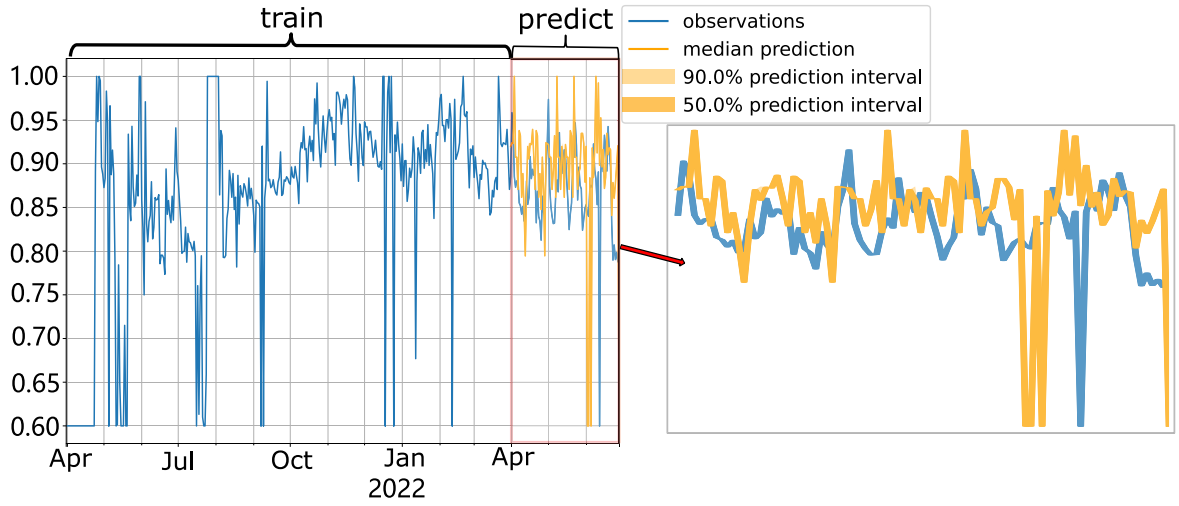


Fig. 10. PR prediction through proposed ELNAR model.

should be explained here. The context length used for reference is 365 days, and the prediction length representing the future forecasting is 90 days. In view of the aforementioned settings, the future 90 days are divided into 3 parts, all with 30 days to analyze specifically. It can be seen that the first 30 days have the highest acceptability for both the PR and the SR prediction. In terms of the other two indicators, stationarity and precision, the third 30 days present dramatically reduced performance compared with the first 30 days and the second 30 days.

Figs. 10 and 11 show the prediction results of PR and SR. The yellow areas describe the prediction intervals, 50% and 90% prediction intervals, respectively. The prediction interval means the interval of an individual sample. For example, A 90% prediction interval indicates that if a random sample of 100 individuals is taken from the population, approximately 90 of the individual values are expected to fall within this interval. It can be seen that the widths of the 50% and 90% prediction interval for the PR prediction are smaller than the SR prediction, which denotes that the fluctuation of the PR is smaller than the SR.

4.3.2. Cleaning schedule recommendation for Farm 1 and Farm 2

To ascertain the multiple parameters on the PV module, certain information of solar side is required. The module type composition in farm 1 is relatively complex, incorporating YL235P-29b, CS3K-295, and SEGP6-60-265, with the quantity proportions being roughly equal. In contrast, farm 2 solely utilizes the CETC-535M(H)/144 solar panel module type. Farm 1 has a cover glass with an anti-reflective coating, while farm 2 cover glass is without an anti-reflective coating. All PV panels are constructed using EVA materials, and given their relatively recent deployment, the performance of the EVA should manifest in an optimal state. In these two farms, very few PV panels are frameless, and the majority of the panels have an anodized aluminum alloy (ANA) frame. Electric parameters to calculate FF for the two farms could be obtained by search the module type in Appendix A. For both farm 1 and farm 2, their tilt angles are set at 30 degrees. However, the theoretical optimal tilt angles are 33 degrees and 40 degrees, respectively. Subsequently, the dust particle radius for farm 1 and farm 2 is approximately 35–50 μm and 50–75 μm, respectively. The values for dust deposited density are 9.6 g/m^2 and 2.3 g/m^2, respectively. For ease of calculation, we assume that the number of spherical and cubic particles is the same, and we use the maximum value from the size range.

The Farm 1 and Farm 2 are carefully picked with significantly different range of PR. The Farm 1 with PR around 50% to 70% and the Farm 2 with PR around 70% to 90% can construct a control-experiment related to the variable range of PR. In fact, the PR is the primary factor that influences the decision for cleaning, so the control-experiment is necessary.

In order to obtain the predictive performance of the current ELNAR model, in the actual operation process, an initial 90 days will be set aside as the baseline 'real value' for comparison with the 'predictive value'. The rolling-horizon forecasting process will only be initiated when the accuracy of the ELNAR model reaches the desired state, at which point the corresponding profit will be calculated.

Table 3 denotes the forecasting abilities of the proposed ELNAR model in the application of the Farm 1. The PR prediction performs with good average stationarity (2.87%), precision (91.4%), and acceptability (98.8%). The SR prediction performs terrible during the third 30 days, with stationarity value of 177% and precision value of −120%.

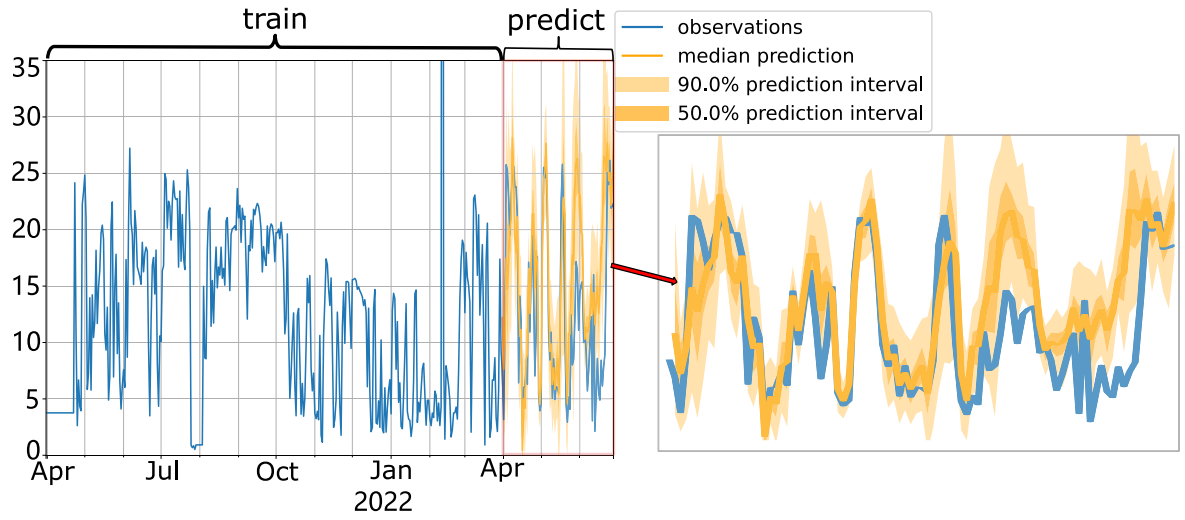


Fig. 11. SR prediction through proposed ELNAR model.

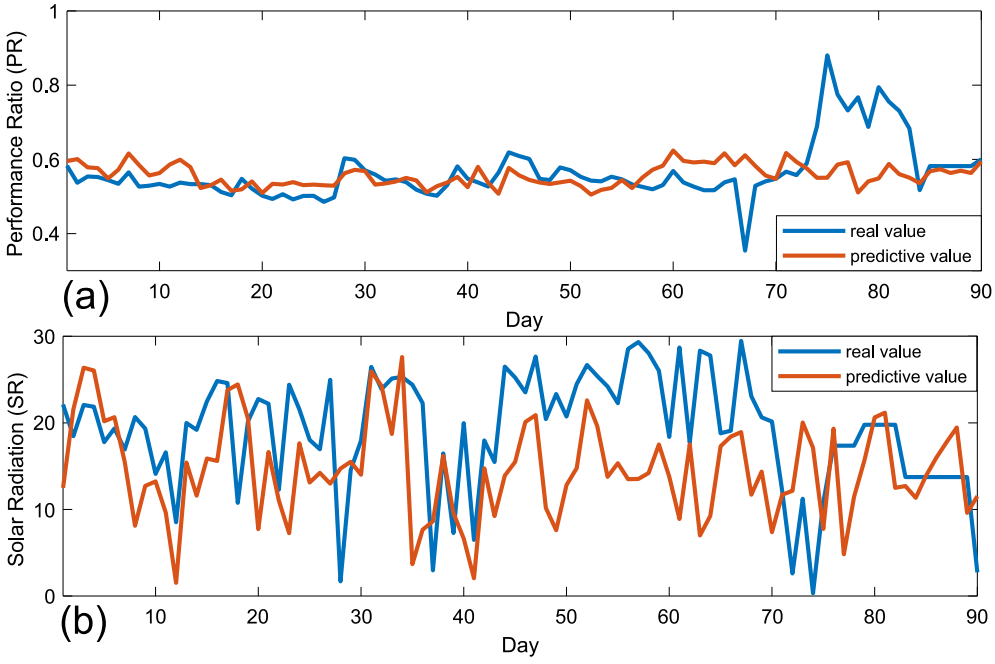


Fig. 12. Prediction of Farm 1 within 90-day span, (a) PR (b) SR.

Table 3
Prediction performance of PR/SR for Farm 1.

Time span (days)	Stationarity PR/SR (%)	Precision PR/SR(%)	Acceptability PR/SR(%)
First 30 days	4.61%/4.32%	94.1%/35.5%	100% /93.3%
Second 30 days	1.06%/3.01%	95.0%/54.9%	100%/93.3%
Third 30 days	2.96%/177%	85.1%/-120%	96.6% /80.0%
Average	2.87%/61.4%	91.44%/−10.0%	98.8%/88.8%

However, the acceptability is relatively rational, denoting that a few extreme points drag down total performance. As shown in Fig. 12, discrepancies between the real and predictive values for PR and SR are evident from around day 60. Despite these anomalies, the overall predictive results are satisfactory, barring a few outliers.

Table 4 shows the forecasting abilities of the proposed ELNAR model in the application of the Farm 2. The acceptability values of the PR prediction all reach 100% for the three types of time span, and this average value is only 78.8% for the SR prediction, but both of them are acceptable. As depicted in Fig. 13, extending the prediction scale results in substantial divergence between real and predictive PR and SR values starting from the 60th day. Despite this, the PR and SR predictions maintain a considerable accuracy.

As illustrated in Figs. 14 and 15, the profit curves for Farm 1 and Farm 2 are delineated. These figures serve to provide suitable cleaning recommendations conducive for informed decision-making. When the function f_κ is not taken into account, the profit model can be regarded as naive. Conversely, when the function f_κ is considered,

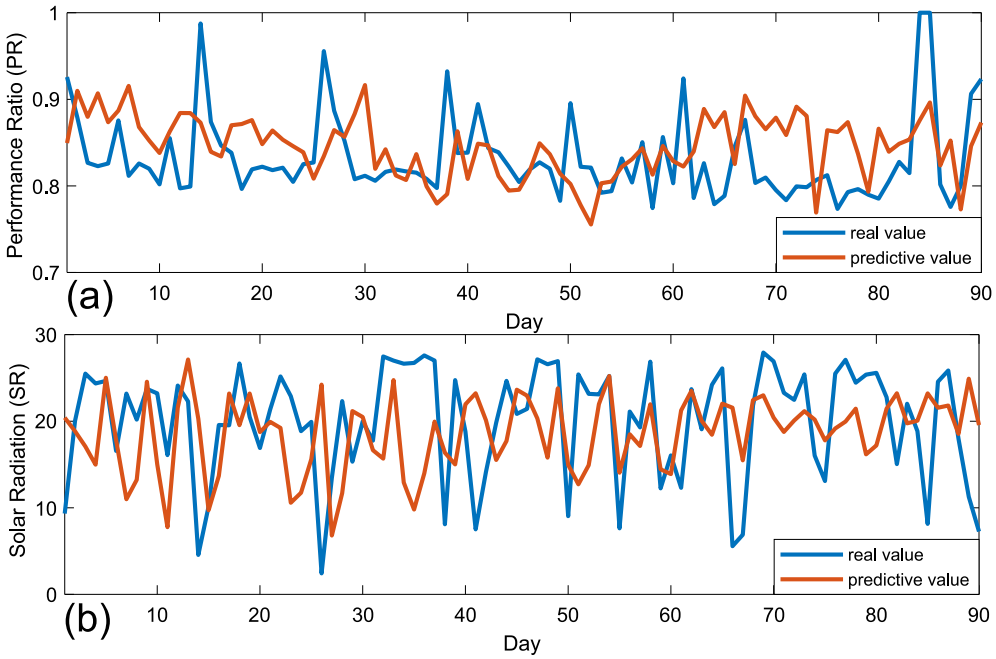


Fig. 13. Prediction of Farm 2 within 90-day span, (a) PR (b) SR.

Table 4
Prediction performance of PR/SR for Farm 2.

Time span (days)	Stationarity PR/SR (%)	Precision PR/SR(%)	Acceptability PR/SR(%)
First 30 days	3.18%/3.04%	93.9%/31.9%	100%/83.3%
Second 30 days	1.01%/0.05%	96.7%/62.7%	100%/83.3%
Third 30 days	3.66%/26.2%	92.4%/53.3%	100%/70.0%
Average	2.61%/9.78%	94.3%/49.3%	100%/78.8%

it is an adjusted profit model. The adjusted profit model implies that multiple parameters on PV module performance are considered. This means that the derived profit model is more reasonable, closer to the actual situation, and incorporates more empirical information from the PV side. Intuitively, the adjusted profit curve tends to be lower and may even exhibit negative values, especially when the cleaning cycle is too short, as evidenced by the first point on the blue line in Fig. 14. From an analysis of profitability pertaining to Farm 1, as depicted in Fig. 14, it can be discerned that the apex of profit is attained between day 7 and day 13. The improvement of profitability can reach a magnitude of 30%, thereby underscoring the indispensability of a flexible cleaning operation. In relation to the anticipated profitability of Farm 2, Fig. 15 shows that the interval spanning from day 13 to day 19 is characterized by high profit margins. Under optimal conditions, the profit improvement is restricted to a mere 6%.

Combined with the Table 5, the reference suggestions of cleaning schedule for the two farms can be given. Before using the recommendation system, both farms adhered to a standardized cleaning cycle of 30 days. The resultant profit improvement for Farm 1 and Farm 2 are calculated as (2.63-2.12) / 2.12 * 100% = 24% and (1.06-1.04) / 1.04 * 100% = 2%, respectively. Referring to the aforementioned analysis based on the traditional setting, the specific cleaning recommendation can be suggested further. For Farm 1, the optimal cleaning interval is pinpointed to a single day within the span of days 7 to 13, yielding a profit enhancement of 30%. Considering the uncontrolled factors that complicate the execution of the anticipated plan, the 90% maximum profit threshold offers another wide reference range spanning from day 4 to day 18. Regarding the suggestions for Farm 2, the ideal cleaning should be executed on one day from day 13 to day 19. Nevertheless, based on the maximal improvement of 6% and the 90% maximum interval (spanning from day 4 to day 90), the urgency level for cleaning the PV panels is relatively low. Consequently, it is not advocated for this farm to engage in cleaning procedures for the subsequent 90 days.

Finally, the real operating environment of the two farms can be analyzed to explain the differences of cleaning recommendations between Farm 1 and Farm 2. The Farm 1 is constructed around a heavily polluting enterprise without adequate light, resulting in the low PR throughout the year. With respect to the Farm 2, it is located at the basin with excellent light resources, generating the condition of high PR. This also reveals the general rule that the PV panels lying in the dirty environment without proper light should be cleaned regularly.

5. Analysis and discussion from the PV side

In order to obtain a better understanding of this cleaning recommendation system, this section will provide further elaboration from the perspective of the PV side.

For PV panels, ensuring surface cleanliness is pivotal for excellent energy output. However, under differing conditions such as varying locations and levels of irradiance, the energy output of PV panels can exhibit substantial variations. For instance, in regions with higher irradiance, the effect of cleanliness on power production is relatively negligible unless the panel surfaces are heavily contaminated and fully blanketed in dust. This becomes particularly pronounced in areas with lower irradiance, where the cleanliness of the panels can result in marked disparities in power generation. For instance, results from Farm 1 show that the low PR value fluctuates between 50% and 70%, a substantial improvement of 30%, potentially attributable to low irradiance. Therefore, keeping the excellent energy output for a farm is difficult due to large amounts of uncertainties.

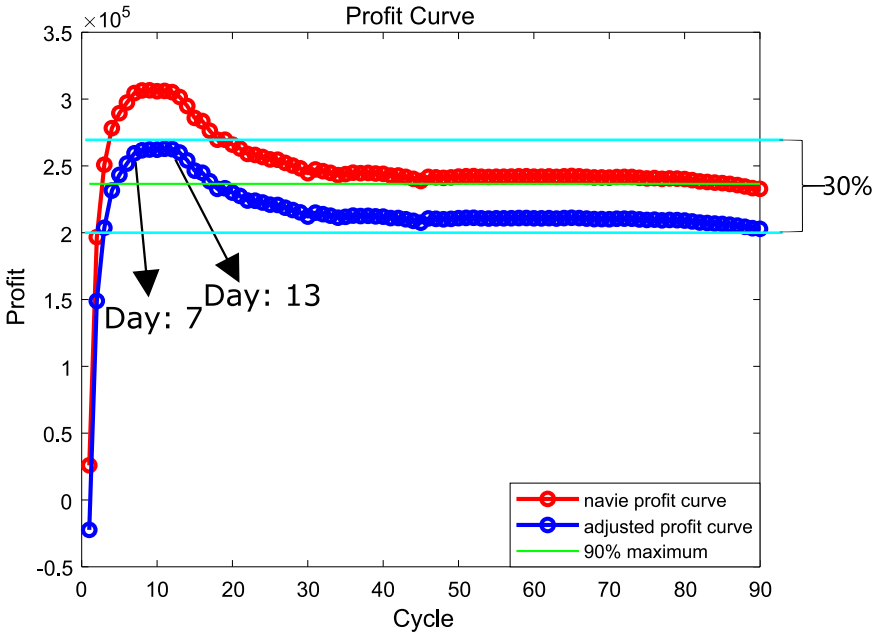


Fig. 14. Profit curve about the cleaning cycle of Farm 1.

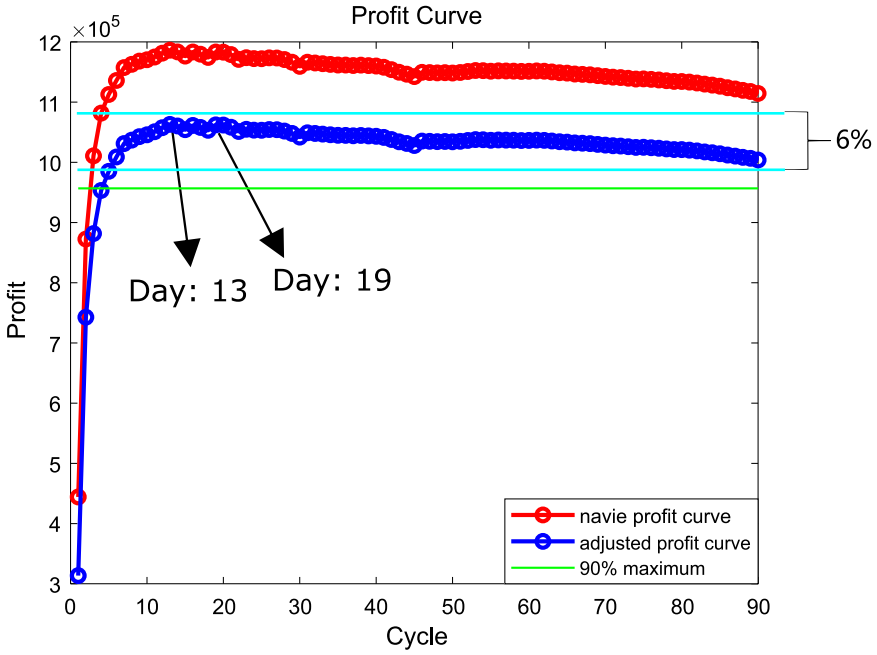


Fig. 15. Profit curve about the cleaning cycle of Farm 2.

Table 5
Profit quantitative results for Farm 1 and Farm 2.

Farm (ID)	Maximum profit (Value)	Profit on Day 30 (Value)	Profit on Day 90 (Value)	Profit above 90% maximum (Range)
Farm 1	2.63×10^5	2.12×10^5	2.03×10^5	Day 4 – Day 18
Farm 2	1.06×10^6	1.04×10^6	1.00×10^6	Day 4 – Day 90

To address this challenge, this paper identified the critical factors. In the current version of the cleaning recommendation system, key factors influencing the final energy output of solar farms include SR and PR. The future SR and PR are influenced by a variety of things such as weather, temperature, historical SR and PR data, and cleaning records. By subtracting the incurred expenses of cleaning costs from the revenue derived through ultimate energy production, the overall net profit of a solar farm can be determined. This situation thus mandates that the farm strike a judicious tradeoff between the earnings from energy output and the associated cleaning costs. Furthermore, considering that cleaning operations require labor engagement, and the cost of labor

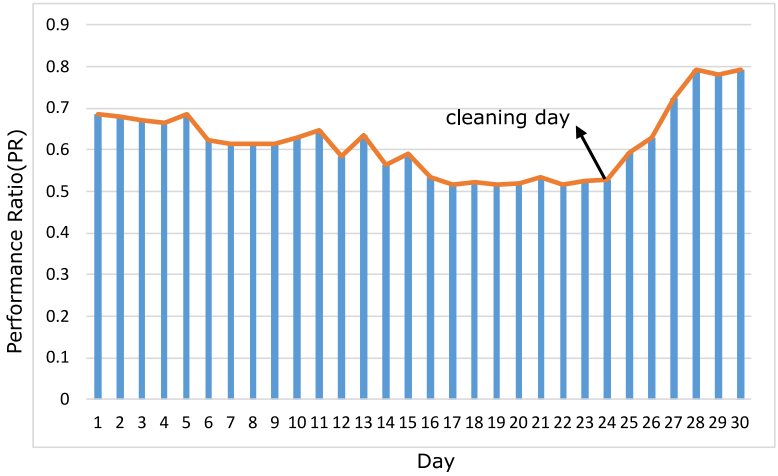


Fig. 16. PR fluctuation within one month.

differs across regions, this can result in minor variances in these costs, thereby indirectly influencing the ultimate revenue of the power plant.

Here, we used PR data of one month that executes the cleaning operation. Fig. 16 can be used to analyze the necessity of cleaning from the efficiency value of PV energy conversion, i.e., the PR. This figure depicts the change in PR values for farm 1 over a month, where we can observe that the PR significantly continues to decrease until the 24th day. However, after cleaning was conducted locally on the 24th day, the PR value began to rise from around 0.5 to around 0.8.

Additionally, certain significant but time-invariant factors from the PV side, such as solar modules and tilt angle, can greatly influence energy production. In traditional methods, these substantial time-invariant factors should be incorporated into the system. However, the advantage of our recommendation system lies in the fact that while the addition of these time-invariant factors could theoretically enhance prediction accuracy and recommendation performance, their exclusion would not significantly impede the recommendation performance. This is attributable to the fact that these time-invariant factors can be considered a constant bias in the recommendation system and would be adjusted by the system's self-learning capabilities.

6. Conclusion

A rolling-horizon cleaning recommendation system for PV panels is presented in this paper, which could consider time-variant influencing factors based on a prediction model and profit model. (1.) The time-variant factors SR and PR are predicted by the proposed ensemble long-term and nonlinear autoregressive (ELNAR) model, which handles time-variant and fickle influencing factors by optimizing predictive output with minimum mean square error estimation and accounting for nonlinear and long-term relationships in time series. (2.) The profit curve is generated by a mathematical profit model considering the rolling-horizon SR and PR. This model is suitable for industrial deployment because it only requires consideration of two crucial factors, making it highly convenient. (3.) A flexible cleaning schedule recommendation date could be inferred based on the generated profit curve, which involves the relationship between the farm profit and cleaning cycle. And this cleaning recommendation system is tested in two real farms. Compared to fixed cleaning period, the profit of solar farms could be improved by 6% and 30%.

CRediT authorship contribution statement

Chao Zhang: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yunfeng Ma:** Writing – review & editing, Formal analysis. **Zengqiang Mi:** Writing – review & editing, Resources. **Fan Yang:** Resources, Investigation. **Long Zhang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This paper would like to express special thanks to Amazon.com, Inc., and Goldwind Science & Technology Co., Ltd. for their contribution and support. The authors would further like to gratefully acknowledge the support of the China Scholarship Council (No. 202106730115).

Appendix A. Solar panel information

This section presents detailed information about the solar cells installed in the PV farm.

Appendix B. Solutions for predictive component 1

According to the sparse Bayesian learning, the determination of Θ can be viewed as l_1-norm problem:

$$\boldsymbol{\Theta} = \arg\min \left\{ \frac{1}{2}\|\boldsymbol{P\Theta} - \mathbf{y}\|_2^2 + \lambda\|\boldsymbol{\Theta}\|_1 \right\} \tag{B.1}$$

where λ is the adjustable penalty parameter. According to augmented Lagrangian, the quadratic penalty (B.1) can be transformed into a unconstrained optimization problem with intermediate variable υ:

$$\min_{\boldsymbol{\Theta},\upsilon\in\mathbb{R}^M} f_1(\boldsymbol{\Theta}) + f_2(\upsilon) + \frac{\mu}{2}\|\boldsymbol{\Theta} - \upsilon\|_2^2 \text{ s.t. } \upsilon - \boldsymbol{\Theta} = 0 \tag{B.2}$$

where $f_1(\boldsymbol{\Theta}) = \frac{1}{2}\|\boldsymbol{P\Theta} - \mathbf{y}\|_2^2$ and $f_2(\upsilon) = \lambda\|\boldsymbol{\Theta}\|_1$. The μ indicates the Lagrange multiplier. The solution of (B.2) is therefore more equivalent to the weighted l_1-norm minimization problem (B.1) as the value of μ rises.

In order to estimate the coefficient matrix $\boldsymbol{\Theta}$, this paper refers to [35–37] to obtain the final iteration formulas, which has been repeatedly proved efficient in many researches.

$$\widehat{\boldsymbol{\Theta}}_{k+1} = \left(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} + \mu\boldsymbol{I}\right)^{-1}\left(\boldsymbol{P}^{\mathrm{T}}\mathbf{y} + \mu\left(\upsilon_k + d_k\right)\right) \tag{B.3}$$

Table A.6
Material parameters information.

Module type	Frame	Front cover	Encapsulant	Type of solar cell
CETC-535M(H)/144	ANA	3.2 mm ultra-white, AR-coated tempered glass	EVA	Mono-crystalline
YL235P-29b	ANA	3.2 mm tempered Glass	EVA	Poly-crystalline
CS3K-295	ANA	3.2 mm tempered glass	EVA	Poly-crystalline
SEGP6-60-265	None	2.5 mm*2 tempered glass	EVA	Poly-crystalline

1 ANA: anodized aluminum alloy.

Table A.7
Electric parameters information.

Module type	P_{MPP}	Power tolerance	V_{MPP}	I_{MPP}	V_{oc}	I_{sc}	FF
CETC-535M(H)/144	535 W	±3%	41.5 V	12.9 A	49.35 V	13.78 A	0.79
YL235P-29b/1650 × 990SERIES	235 W	±5 W	29.5 V	7.97 A	37 V	8.54 A	0.74
CS3K-295	305 W	±10 W	32.8 V	9.28 A	39.5 V	9.73 A	0.79
SEGP6-60-265	265 W	±3%	31.4 V	8.44 A	38.1 V	8.98 A	0.77

$$v_{k+1} = \max\left(0, \left(\hat{\boldsymbol{\theta}}_{k+1} - d_k\right) - \mu/\lambda\right) - \max\left(0, -\left(\hat{\boldsymbol{\theta}}_{k+1} - d_k\right) - \mu/\lambda\right) \tag{B.4}$$

$$d_{k+1} = d_k - \left(\hat{\boldsymbol{\theta}}_{k+1} - v_{k+1}\right) \tag{B.5}$$

where d and v, as the temporary parameters during the iteration, are both set as 0 at the initial state, namely $d_0 = 0$ and $v_0 = 0$.

Appendix C. Prove the effectiveness of adjusted process

Firstly, the error between the true value and ideal optimized value is defined as follows:

$$e(t) = o(t) - y(t) \tag{C.1}$$

Then measurement covariance error matrix can be described below:

$$P(t) = E\left[e(t)e(t)^T\right] \tag{C.2}$$

Put (C.1) into (C.2):

$$\begin{aligned} P(t) = E\Big[&\left[(I - K(t)H)\left(o(t) - \hat{o}'(t)\right) - K(t)v(t)\right] \\ &\times \left[(I - K(t)H)\left(o(t) - \hat{o}'(t)\right) - K(t)v(t)\right]^T\Big] \\ = (I - K(t)H)\, E&\left[\left(o(t) - \hat{o}'(t)\right)\left(o(t) - \hat{o}'(t)\right)^T\right] \\ &\times (I - K(t)H)^T + K(t)E\left[v(t)v(t)^T\right]K(t)^T \end{aligned} \tag{C.3}$$

where $E\left[v(t)v(t)^T\right] = R$. $P(t)$ can be rewritten as follows:

$$P(t) = (I - K(t)H)\, P'(t)\, (I - K(t)H)^T + K(t)RK(t)^T \tag{C.4}$$

To get minimum mean square error estimation, the trace of matrix $P(t)$ needs to be calculated, because the mean square error is exactly the trace of matrix $P(t)$, namely $\mathrm{tr}\,(P(t))$.

$$\mathrm{tr}\,(P(t)) = \mathrm{tr}\left(P'(t)\right) - 2\,\mathrm{tr}\left(K(t)HP'(t)\right) + \mathrm{tr}\left(K(t)\left(HP'(t)H^T + R\right)K(t)^T\right) \tag{C.5}$$

To minimize the mean square error, namely trace of matrix $P(t)$, we calculate the derivative of $\mathrm{tr}\,(P(t))$ related to the gain matrix $K(t)$:

$$\frac{\partial tr\,(P(t))}{\partial K(t)} = -\frac{\partial tr\left(2K(t)HP'(t)\right)}{\partial K(t)} + \frac{\partial tr\left(K(t)\left(HP'(t)H^T + R\right)K(t)^T\right)}{\partial K(t)} \tag{C.6}$$

Suppose the symbol M_1 and M_2 represent two matrices, we have:
theorem 1:

$$\frac{\partial tr(M_1M_2)}{\partial M_1} = M_2^T \tag{C.7}$$

theorem 2:

$$\frac{\partial\,\mathrm{tr}\left(M_1M_2M_1^T\right)}{\partial M_1} = 2M_1M_2 \tag{C.8}$$

then we can simplify (C.6) into:

$$\frac{\partial tr\,(P(t))}{\partial K(t)} = -2\left(HP'(t)\right)^T + 2K(t)\left(HP'(t)H^T + R\right) \tag{C.9}$$

make it equal to zero:

$$K(t) = P'(t)H^T\left(HP'(t)H^T + R\right)^{-1} \tag{C.10}$$

Similarly, the prediction error covariance matrix $P'(t)$ can be deduced easily, which is symmetric and $P'(t) = (P'(t))^T$. Firstly, the error between the true value and the predicted value is shown as follows:

$$e'(t) = o(t) - y(t) \tag{C.11}$$

Then prediction covariance error matrix is:

$$P'_k = E\left[e'_k e'^T_k\right] = E\left[(o(t) - y(t))\,(o(t) - y(t))^T\right] \tag{C.12}$$

If $E\,[w(t)] = 0$, we can obtain:

$$\begin{aligned} P'(t) &= E\left[(o(t) - y(t))\,(o(t) - y(t))^T\right] \\ &= E\left[(Ao(t-1) + Bs(t) + w(t) - A\hat{o}(t-1) - Bs(t))\,(Ao(t-1)\right. \\ &\quad \left. + Bs(t) + w(t) - A\hat{o}(t-1) - Bs(t))^T\right] \\ &= E\left[(A\,(o(t-1) - \hat{o}(t-1)) + w(t))\,(A\,(o(t-1) - \hat{o}(t-1)) + w(t))^T\right] \\ &= E\left[(Ae(t-1))\,(Ae(t-1))^T\right] + E\left[w(t)w(t)^T\right] \\ &= AP(t-1)A^T + Q \end{aligned} \tag{C.13}$$

References

[1] Gebara C, Laurent A. National sdg-7 performance assessment to support achieving sustainable energy for all within planetary limits. Renew Sustain Energy Rev 2023;173:112934.
[2] Chow TT. A review on photovoltaic/thermal hybrid solar technology. Appl Energy 2010;87(2):365–79.
[3] Lu H, Lu L, Wang Y. Numerical investigation of dust pollution on a solar photovoltaic (pv) system mounted on an isolated building. Appl Energy 2016;180:27–36.
[4] Ghazi S, Sayigh A, Ip K. Dust effect on flat surfaces–a review paper. Renew Sustain Energy Rev 2014;33:742–51.
[5] Patil P, Bagi J, Wagh M. A review on cleaning mechanism of solar photovoltaic panel. In: 2017 international conference on energy, communication, data analytics and soft computing (ICECDS). IEEE; 2017, p. 250–6.
[6] Cleaning solar panels: Why, when, and how you should do it. 2017, https://www.energymatters.com.au/renewable-news/solar-panels-cleaning/, Last accessed on 2022-8-15.
[7] El-Shobokshy MS, Hussein FM. Degradation of photovoltaic cell performance due to dust deposition on to its surface. Renew Energy 1993;3(6–7):585–90.
[8] Mani M, Pillai R. Impact of dust on solar photovoltaic (pv) performance: Research status, challenges and recommendations. Renew Sustain Energy Rev 2010;14(9):3124–31.

[9] Jiang Y, Lu L, Lu H. A novel model to estimate the cleaning frequency for dirty solar photovoltaic (pv) modules in desert environment. Sol Energy 2016;140:236–40.
[10] Prasad RD, Bansal R, Raturi A. Multi-faceted energy planning: A review. Renew Sustain Energy Rev 2014;38:686–99.
[11] Debnath KB, Mourshed M. Forecasting methods in energy planning models. Renew Sustain Energy Rev 2018;88:297–325.
[12] Qazi A, Fayaz H, Wadi A, Raj RG, Rahim N, Khan WA. The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. J Cleaner Prod 2015;104:1–12.
[13] Amato U, Andretta A, Bartoli B, Coluzzi B, Cuomo V, Fontana F, et al. Markov processes and fourier analysis as a tool to describe and simulate daily solar irradiance. Solar energy 1986;37(3):179–94.
[14] Reikard G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy 2009;83(3):342–9.
[15] Prema V, Rao KU. Development of statistical time series models for solar power prediction. Renew Energy 2015;83:100–9.
[16] Li L-L, Wen S-Y, Tseng M-L, Wang C-S. Renewable energy prediction: A novel short-term prediction model of photovoltaic output power. J Clean Prod 2019;228:359–75.
[17] Zheng J, Zhang H, Dai Y, Wang B, Zheng T, Liao Q, et al. Time series prediction for output of multi-region solar power plants. Appl Energy 2020;257:114001.
[18] Marzouq M, El Fadili H, Zenkouar K, Lakhliai Z, Amouzg M. Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data. Renew Energy 2020;157:214–31.
[19] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. Energy 2018;148:461–8.
[20] Goliatt L, Yaseen ZM. Development of a hybrid computational intelligent model for daily global solar radiation prediction. Expert Syst Appl 2023;212:118295.
[21] Talavera DL, Muñoz-Cerón E, Ferrer-Rodríguez JP, Nofuentes G. Evolution of the cost and economic profitability of grid-connected pv investments in spain: Long-term review according to the different regulatory frameworks approved. Renew Sustain Energy Rev 2016;66:233–47.
[22] Tao JY, Finenko A. Moving beyond lcoe: impact of various financing methods on pv profitability for sids. Energy Policy 2016;98:749–58.
[23] Zakeri B, Cross S, Dodds PE, Gissey GC. Policy options for enhancing economic profitability of residential solar photovoltaic with battery energy storage. Appl Energy 2021;290:116697.
[24] Said SA, Hassan G, Walwil HM, Al-Aqeeli N. The effect of environmental factors and dust accumulation on photovoltaic modules and dust-accumulation mitigation strategies. Renew Sustain Energy Rev 2018;82:743–60.
[25] Said SA, Al-Aqeeli N, Walwil HM. The potential of using textured and anti-reflective coated glasses in minimizing dust fouling. Sol Energy 2015;113:295–302.
[26] Kempe MD, Moricone T, Kilkenny M. Effects of cerium removal from glass on photovoltaic module performance and stability. In: Reliability of photovoltaic cells, modules, components, and systems II. Vol. 7412, SPIE; 2009, p. 203–14.
[27] Tummalieh A, Beinert AJ, Reichel C, Mittag M, Neuhaus H. Holistic design improvement of the pv module frame: Mechanical, optoelectrical, cost, and life cycle analysis. Prog Photovolt, Res Appl 2022;30(8):1012–22.
[28] Wu Z, Yan S, Ming T, Zhao X, Zhang N. Analysis and modeling of dust accumulation-composed spherical and cubic particles on pv module relative transmittance. Sustain Energy Technol Assess 2021;44:101015.
[29] Chen S, Billings SA, Luo W. Orthogonal least squares methods and their application to non-linear system identification. Int J Control 1989;50(5):1873–96.
[30] Pan W, Yuan Y, Gonçalves J, Stan G-B. A sparse bayesian approach to the identification of nonlinear state-space systems. IEEE Trans Automat Control 2015;61(1):182–7.
[31] Salinas D, Flunkert V, Gasthaus J, Januschowski T. Deepar: Probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 2020;36(3):1181–91.
[32] Welch G, Bishop G, et al. An introduction to the kalman filter. 1995.
[33] Auger F, Hilairet M, Guerrero JM, Monmasson E, Orlowska-Kowalska T, Katsura S. Industrial applications of the kalman filter: A review. IEEE Trans Ind Electron 2013;60(12):5458–71.
[34] Qu Y, Cai H, Ren K, Zhang W, Yu Y, Wen Y, et al. Product-based neural networks for user response prediction. In: 2016 IEEE 16th international conference on data mining (ICDM). IEEE; 2016, p. 1149–54.
[35] Tang X, Zhang L, Li X. Bayesian augmented lagrangian algorithm for system identification. Systems Control Lett 2018;120:9–16.
[36] Liu Z, Tang X, Wang X, Mugica JE, Zhang L. Wind turbine blade bearing fault diagnosis under fluctuating speed operations via bayesian augmented lagrangian analysis. IEEE Trans Ind Inf 2020;17(7):4613–23.
[37] Liu Z, Yang B, Wang X, Zhang L. Acoustic emission analysis for wind turbine blade bearing fault detection under time-varying low-speed and heavy blade load conditions. IEEE Trans Ind Appl 2021;57(3):2791–800.