Nankai University

Undergraduate Thesis (Design)

Chinese title: Ultra-short-term solar irradiance study based on predictable meteorological factors

Foreign language title: Ultra-short-term Prediction of Solar Irradiance Based on Predictable Meteorological Factors

学号：1810055姓名：刘汉青年级：2018年级专业：统计学
Department: Department of Probability and Statistics
College: School of Mathematical Sciences Instructor:

Ji-Shou Nguyen
Completion date: 2022.05

!" #$%&'()*+,-. /01 345

I solemnly declare that the dissertation submitted by me is the result of my research work under the guidance of my supervisor. The research results of this dissertation do not contain any published or unpublished works created by others, except for those cited in the text. Other individuals and collectives who have contributed to the research work involved in this dissertation have been clearly indicated in the text. The ｌｅｇａｌ responsibility for the originality of this dissertation rests with me.

Dissertation author's signature.

Year January

Date

I declare that this dissertation is a research result completed by my supervising student, and I have reviewed all the contents of the dissertation and can guarantee the consistency and accuracy of the Chinese and English contents of the title, keywords and abstract part.

Signature of Dissertation Supervisor.

Year January

Date

! #

Photovoltaic power generation is currently an important way to harness solar energy. By accurately predicting the irradiance, the output data of the photovoltaic power generator can be indirectly predicted in the future period. In this paper, the irradiance is predicted for a short period of time in the future using historical data such as wind speed, wind direction, temperature, humidity, and pressure. In this paper, similar samples are reconstructed using pattern recognition methods and the results of power spectrum analysis for each dimension of meteorological factors are used to analyze their predictability in terms of signal science. To meet the requirements for the sampling frequency, the meteorological factors containing high-frequency components are spiked using a linear interpolation method. It was tested that the optimized MMMLP model using the reconstituted samples achieved significantly better prediction results compared to the original model.

!" #irradiance prediction; pattern recognition; power spectrum analysis; multilayer perceptron

Abstract

Photovoltaic power generation is an important way to utilize solar energy and one of the important clean energy sources in China. Using the accurate prediction of irradiance, the output data of photovoltaic power generation elements in the future can be indirectly predicted. Using the accurate prediction of irradiance, the output data of photovoltaic power generation elements in the future can be indirectly predicted. Based on the historical data of wind speed, wind direction, temperature, humidity and pressure, the solar irradiance in a short time in the future is predicted. In this paper, the pattern recognition method is used to reunion the similar samples, and the power spectrum analysis results of each meteorological factor are used to In order to meet the requirement of sampling frequency, the meteorological factors with high After being tested, the optimized MMMLP model using the recombined samples After being tested, the optimized MMMLP model using the recombined samples achieved significantly better prediction results than the original model.

Key words: Irradiance prediction; Pattern recognition; Power spectrum analysis; Multilayer perceptron

Summary 4
Abstract 5
I. Introduction 7
(I) Background and significance of the subject research. 7
1. Status and characteristics of photovoltaic power generation systems 7
2. Requirements and difficulties of PV power forecasting 7
3. Correlation of irradiance prediction and PV power prediction 9
(ii) Review of existing irradiance prediction methods 9
1. Mathematical prediction method based on historical data 10
2. Prediction methods based on image processing 11
3. Forecasting methods based on WRF model weather forecasts 12
II. Pattern recognition-based weather type recognition and diurnal recognition 14
(i) Weather type exploration and similar day identification using clustering methods 14
1. Weather type design and pre-processing 14
2. Cluster Analysis 16
(ii) Circadian recognition based on support vector machine 17
(iii) Sample data restructuring 20
III. Predictability analysis based on power spectrum calculation 21
(i) Power spectrum analysis and predictability analysis based on SFGPS and AFAGPS 21
(ii) Sampling frequency determination and sample interpolation 23
IV. Irradiance prediction using MMMLP 26
(i) Brief description of the algorithm 26
(ii) Predictive performance analysis 26
V. Conclusion 29
References 30

"#$%

$%&'()*+,-. /0

$$%&'()*+,-. /0

As one of the world's richest countries in terms of solar energy resources, China has a promising future for resource development. China's annual theoretical solar energy reserves are equivalent to the energy produced by burning 14.95 billion tons of standard coal.[1] As an important way to utilize solar energy, photovoltaic power generation offers broad application prospects with its flexible application form and simple maintenance.[2] According to China's development plan in the field of renewable energy, the installed capacity of solar power generation will reach 600GW by 2050, by which time photovoltaic The installed capacity of photovoltaic power generation will account for 5% of the country's installed power capacity. At the same time, the compound growth rate of China's installed solar power capacity will reach more than 25% in the next decade or so. How to forecast the efficiency of photovoltaic power generation with relative accuracy, so as to provide reasonable guidance for energy dispatch, in order to improve the grid system's ability to consume and utilize electricity, becomes a key issue.

Photovoltaic power generation systems can be divided into grid-connected and off-grid types according to whether they are connected to the grid or not. Off-grid power generation systems are generally used in relatively remote areas such as islands, where the power generation is small and the installed capacity accounts for a small proportion of the total capacity of photovoltaic power generation, and are mainly used to meet the small-scale local electricity demand at that time. Grid-connected PV power plants are usually divided into

large-scale grid-connected PV power plants and distributed grid-connected PV power generation systems, both of which need to be connected to power dispatch management systems to coordinate with different sources of power. This is currently the main method of photovoltaic power generation. Since most grid-connected PV plants are not equipped with energy storage, the power generated is usually connected directly to the grid and is shut down and operated according to diurnal variations.

2$%&'(34+56.78

Conventional sources of electricity such as thermal, hydroelectric, and nuclear power usually have continuous and controllable characteristics.[3] However, according to the efficiency equation for PV power generation proposed in the literature [4], the output of PV power depends heavily on solar irradiance and other climatic conditions such as temperature, cloudiness, relative humidity, clear sky index, and wind speed.[4] These meteorological factors can have a drastic effect on the power generation performance of PV modules . The literature [2] has investigated the light

It has been pointed out in the study of PV generator system that local shading may lead to serious power loss. If the shaded area accounts for 10% of the total module area, the power of PV output may be reduced to 20% of the original; and when the shaded area reaches 20% or more, the power generated by the module is almost 0. Therefore, factors such as clouds and water vapor can bring serious uncertainties to PV power generation.[2] These factors not only undergo periodic shifts with time, but also are prone to sudden changes. This results in significant nonlinearity and indirect fluctuations in the power of PV generation.

When the proportion of PV power generation is small, the impact of its fluctuation on the balance of power generation - transmission - consumption is not obvious. With the rapid increase of installed capacity of PV power plants, the fluctuation of its output requires the power dispatching system to provide backup power to match its capacity to smooth out the fluctuation to ensure the stability of power and frequency, which not only means that a certain amount of installed power is required to be in the state of readiness for a long time, but also means that frequent grid connection operations are required. If the power of photovoltaic generation is constantly in a state that cannot be stabilized and accurately predicted, it will not only lead to energy waste, but also may pose a threat to the security of the power system. If accurate and timely prediction of PV power can be made, it can provide important data support for grid management and decision making, and help to regulate the power and peak frequency of the grid, as well as reduce the spare capacity and fuel operation cost. From the perspective of PV operators, effective forecasting methods can improve their economic efficiency and return on investment, help them rationalize maintenance and repair of power generation units and inverters, reduce penalties and losses due to uncertainty, and thus further enhance the economic competitiveness of PV plants.

The prediction of PV power has become a key area of concern for

the energy and power industries in recent years, and relevant industry standards have been introduced. According to the requirements in the Technical Requirements for PV Power Forecasting (Draft for Comments) and the Functional Specifications for PV Power Forecasting System (Draft for Comments), the ultra-short-term PV forecasting should be able to forecast the future 15 minutes to 4 hours with a resolution of 15 minutes.

Output power, and the forecast should be executed once every 15 minutes, and the single calculation time should be less than 5 minutes. Short-term forecasting

The time span should be extended to the next day from zero hour to the next 72 hours, twice a day, and a single implementation

The time should be less than 5 minutes. This requires that the prediction mechanism should have the characteristics of low time complexity, iterability, and self-correction, in addition to a thorough study of the law of change. In this paper, we take into account the adaptability of the model for short-term forecasting and focus on the requirements related to ultra-short-term forecasting.

%$9:;34<%&'(=>34+!?

Among the factors affecting the power of photovoltaic power generation, the sun's radiation is the most important meteorological factor. Solar irradiance refers to the solar radiation energy received by an object per unit time and unit area, and is usually measured directly by meteorological observation equipment. The solar irradiation directly causes the photovoltaic effect, where the sunlight irradiated to the photovoltaic panel induces the directional movement of electrons in the single crystal silicon, which in turn generates the built-in electric field and creates a current congruence when the external load is connected. At 25 degrees Celsius, the stronger the solar irradiance within a certain range, the higher the output current, the higher the output power, and the location of the maximum power point is also correspondingly higher.[1] Therefore, the size of the solar irradiance directly affects the size of the output power of the PV module. At the same time, there is a physically meaningful quadratic curve relationship between solar irradiance and PV module output power.[5] If solar irradiance can be predicted more accurately, and then using the relevant physical models of different PV power generation units, the power generation of PV power plants can be predicted indirectly.

In general, according to astronomical equations, the intensity of radiation in the outer tangent plane of the atmosphere depends uniquely on the intensity and direction of solar irradiation at the upper boundary of the [atmosphere5 and the] relevant value can be found precisely. This value also determines the theoretical maximum value of solar irradiance received by a region. However, solar radiation undergoes continuous transport in the atmosphere
when it is directly disturbed by many natural environmental factors, such as clouds, aerosols and fine particles, water vapor and ozone. How to find potential patterns and make predictions in these complex

and chaotic climate systems is the main difficulty faced by solar irradiance prediction.

$1&23+456789:;<

At present, the main irradiance prediction schemes are focused on ultra-short-term and short-term prediction, and logically there are three main categories of schemes: numerical prediction methods based on historical data, prediction methods based on image processing, and prediction methods based on WRF model weather forecasts. Among them, the mathematical and theoretical forecasting methods based on historical data mainly use the meteorological data and irradiance data in history, and train a forecasting model using statistical, machine learning or empirical methods. Some of these methods also restructure the samples based on season, geographic location, weather type, or use meteorological and astronomical concepts to pre-process and extract features from meteorological data, or fuse the relevant station geographic and astronomical information. The prediction methods based on image processing, on the other hand, mainly consider the impact of moving clouds on irradiance, and mainly use satellite images or ground-based cloud maps to analyze the size, thickness, height and movement trend of clouds above the PV power station sites, and combine the ground

The solar irradiation intensity outside the site is calculated to predict the solar irradiance at the site by calculating the degree of attenuation. The prediction method based on WRF model weather forecasts uses the values provided by established weather forecasting systems and related weather models to predict the irradiance.

$$@ABCDE+DF34GH

Mathematical forecasting methods based on historical data include continuous forecasting, Markov chain forecasting, wavelet Decomposition method, neural network method, stochastic time series method and combinatorial forecasting method etc.[1].

The continuous prediction method assumes a certain degree of consistency between irradiance at future times and irradiance over a relatively short period of time previously, so the sliding average of historical irradiance data over a previous period is used as the predicted value. Although this method is simple to calculate, it is extremely insensitive to sunrise, sunset and sudden weather changes and is not suitable as a short-term, high-precision prediction method. The multiple linear regression algorithm takes multiple meteorological data and time as independent variables, and trains a multiple linear model after variable screening and pre-processing. However, irradiance itself has strong nonlinearity, especially when it is consistently 0 after sunset, and there is an obvious truncation phenomenon, so the prediction effect is not satisfactory. Markov chain prediction methods, on the other hand, use state probability vectors and state probability transfer matrices to describe the likelihood of irradiance being in different states in the future, based on the current irradiance and its changing trend. However, the basic assumptions of Markov prediction also cannot be adapted to the requirements of irradiance prediction, i.e., the trends and probabilities of state changes in irradiance are not the same, and there are frequent random changes in addition to periodic

sunrise, noon, and sunset changes. Therefore, large errors can occur in inter-day and night predictions, as well as when used in scenarios with more drastic weather changes[6].

Wavelet decomposition is usually used as an auxiliary method of neural networks. By performing wavelet decomposition on the historical data of solar irradiance, the high-frequency and low-frequency components are split, and then each component is predicted individually using neural networks, and finally the predicted solar irradiance is obtained by wavelet reconstruction. Several different network models have been used in studies related to neural network algorithms, including the use of BP neural networks to predict irradiance using meteorological data as input, or the use of extreme learning machine (ELM) algorithms to speed up training to accommodate the time complexity required for short-term prediction. To better accommodate long sequences, some studies have used LSTM long and short-term memory networks to increase the sensitivity to time series by using gate structures to control the forgetting of old information and preserve a portion of the information for a longer period of time. By reorganizing the historical samples, better prediction of irradiance can be achieved. However, due to the relatively large depth of the network and the fact that it requires more powerful computational resources to run and perform for a longer period of time if attempting to devote a longer series to prediction. In the ultra-short term prediction

In the measurement, the network structure needs to be retrained after putting in a new iteration of samples every 15 minutes, which may not meet the industry standard of no more than five minutes for a single computation.

For the stochastic event series method, Pandit et al. pointed out that the autoregressive sliding average model (ARMA) is sufficient to forecast nonlinear time series and make predictions, but the predictions are still not satisfactory. Combinatorial forecasting methods, on the other hand, try to combine the above methods with variable selection methods and sample combination methods in different ways so as to complement each other's strengths, including using wavelet decomposition in combination with LSTM to strengthen the regularity of training data; or using hierarchical clustering or support vector machine classification to aggregate training samples and then train them individually; or weighting the periodic variation pattern of irradiance and neural network prediction results or try to use training samples with different time steps to adapt to weather pattern transitions, etc.

2$@AIJKF+34GH

The prediction methods based on image processing mainly include the prediction methods based on satellite data and ground-based cloud map. Among them, satellite data refers to meteorological satellite cloud maps captured by meteorological satellites outside the atmosphere, and ground-based cloud maps refer to the acquisition of cloud maps using ground-based equipment including the TSI all-sky imager.[7] After obtaining the relevant images, the interference information existing between different cloud clusters in the cloud maps due to drawing contour lines is removed by image denoising algorithms, and the cloud clusters among them are extracted using the maximum inter-class variance adaptive threshold segmentation

algorithm. Although the motion of cloud masses and the change of their own morphology are very complicated, for ultra-short-term prediction, cloud masses are not prone to drastic morphological changes in ultra-short time, and it can be basically assumed that cloud masses keep their original shape and pan on the image.[5] Then, we use geography to calculate the depth, shape, range and change trend of the shadow they constitute on the earth's surface, and determine accordingly whether cloud masses constitute a shadow to PV power generation elements The shading factor is calculated by using the geographic means. The values of the shading coefficient and the intensity of extraterrestrial solar irradiation are used to determine the future irradiance of the PV plant.

This strategy can theoretically predict very well the irradiance decay caused by clouds and their motion, but still has some hard drawbacks. For example, in the irradiance prediction using meteorological satellite clouds in the literature [8], the satellite clouds used do not meet the requirements of a single PV plant in terms of resolution. Its temporal resolution is 30 minutes and spatial resolution is 2.5 square kilometers. According to the relevant regulations issued by the Ministry of Land and Resources on the land for PV power projects, most PV power plants in China should cover an area no larger than 0.5 km2 per 10 MW with full consideration of latitude and conversion efficiency, which means that the minimum spatial unit provided by the satellite cloud map is still five times or more than the area covered by the PV power plant, and its temporal resolution does not match

conform to industry standards. On this basis, some compensation can be made if ground-based cloud maps are used for prediction. For example, the commonly used TSI-880 sky imager can capture and photograph sky images with a smaller rotating hemispherical mirror.[8] However, with an effective detection range of 5 km, it can barely monitor the movement of clouds beyond 5 km. And in the longest time scales of both ultra-short-term and short-term predictions, there is a possibility that clouds have completed their movement across the monitoring range of the domain, which can easily lead to the failure of the prediction model.

L$@A MNO PQRS3T+34GH

The WRF (The Weather Research and Forecasting) model is a system for atmospheric research and numerical weather forecasting with advanced numerical computation techniques that can be used to achieve better results in mesoscale weather forecasting operations.[9] By dividing the atmosphere into columns that run vertically upward from the ground to high altitude and studying the temperature, wind direction, wind speed, humidity, and pressure inside the columns , wind direction, wind speed, humidity and pressure and other physical quantities to describe the weather conditions within a meteorological unit. It can complete the description by parameterizing several physical processes that affect irradiance, including parameterization of microphysical processes (water vapor condensation, etc.), parameterization of cumulus clouds, etc.

However, the WRF model is still not a very reliable weather prediction model, especially in terms of its insensitivity to uncommon weather conditions and unexpected weather.[10] For irradiance prediction, its accuracy of irradiance prediction is high in low cloudiness weather, and in high cloudiness weather, there is a significant decrease in prediction accuracy.

decreases and the irradiated rain prediction error increases as the cloudiness increases.[9] The schemes such as post-classification prediction by weather type and variable screening using multiple linear regression are usually used in studies for revisions, but they are of limited practical use. This is because the WRF model has some problems in discretization. Due to its third-order Runge-Kutta time partitioning scheme for time integration.

$$\Phi^{*} = \Phi^{"} + \frac{\Delta t}{3} R\&\Phi^{"}, \tag{1a}$$

$$\Phi^{**} = \Phi^{"} + \frac{\Delta t}{2} R\&\Phi^{*}, \tag{1b}$$

$$\Phi^{"\#\Delta"} = \Phi^{"} + \Delta t R\&\Phi^{**}, \tag{1c}$$

In contrast, the intrinsic WRF model uses a fixed value for the time step Δt, which is specified to be no more than 6 times the outermost grid space step (in kilometers) in seconds. From the signal science point of view, the results of the discretization of the high-frequency and low-frequency components are obviously different due to the fixed sampling interval.

According to the requirements of the sampling theorem, the high frequency component can be satisfied if sufficiently small time steps are used, but the questions of how to take the value, whether the required sampling frequency will be higher than the actual resolution, etc. cannot be solved. If the dynamic adjustment for different meteorological dimensions and different times cannot be realized, then it is bound to have a strong instability. After decomposing the components of different meteorological factors by power spectra, a maximum sampling interval should be determined according to Shannon's sampling theorem to improve the quality of discretization. This is one of the main tasks of this paper.

In addition to the above three types of irradiance prediction schemes mentioned in this paper, researchers have used a combination of the three schemes or cross-corroborated them to further reduce the errors, and some have achieved better results. For example, the literature [5] predicted irradiance by using the irradiance predicted by numerical methods to integrate the shading parameters calculated from ground-based cloud map analysis, and corrected them using meteorological data provided by the WRF model, thus taking advantage of the different methods as much as possible to improve the accuracy of the prediction.

&#'()*+,-. /01+,234+,

Pattern recognition, refers to the analytical processing of representational information, followed by description, interpretation, discrimination and classification. In application, statistical pattern recognition methods can be used for aggregation of similar samples and differentiation between different categories of samples. And objectively, there are significant periodicity, diurnal similarity, diurnal distinction and seasonal distinction in the process of climatic factors occurring changes. In this section, an attempt is made to evaluate the similarity and differentiation among historical samples using statistical pattern recognition methods to discriminate and classify them, and to provide criteria for designing and combining training samples.

$%&=>? @9:+AB@CDE.FGHIJ

$$URVWXY<3KF

In order to be able to differentiate weather data using weather conditions, several weather classification methods have been proposed by relevant researchers according to the needs. The literature [1] advocates classification by day types such as sunny, clear to cloudy, cloudy, cloudy with light rain, light rain to heavy rain, and snow, while the literature [11] argues that too many classifications may lead to problems such as significant data discontinuity and insufficient sample size in some classes, thus advocating classification into four categories according to Table [111].

Class	weather conditions
A	Sunny, sunny to cloudy, cloudy to sunny
B	Cloudy, cloudy, cloudy to cloudy, cloudy to cloudy,

	fog
C	Showers, thundershowers, sleet, drizzle, light snow, snow showers, freezing rain, light to moderate rain, light to Mid-Snow
D	Moderate rain, heavy rain, torrential rain, heavy rain, very heavy rain, moderate snow, heavy snow, blizzard, moderate to heavy Rain, heavy to torrential rain, torrential to heavy rain, moderate to heavy snow, heavy to torrential snow, sand and dust

Table 1 Classification of weather conditions in four categories

However, the two aforementioned studies do not suggest how to address the situation where category shifts occur within a given day. To avoid this problem, this paper ignores the labeling of weather types and chooses to use unsupervised clustering methods to target weather

The data were clustered. To ensure that there is an adequate number of samples in each category of weather conditions, the number of clusters is set to 4, using the recommendation in the literature [5].

Since the sample data are meteorological observations with 15-minute resolution, and the object of meteorological classification should be a day in the meteorological sense, the daily meteorological data are reorganized and used to form a vector describing the daily meteorological conditions. For a natural day, the meteorological conditions are described according to the representation of daily meteorological conditions in the weather forecast as

$$\$X = [T_{\%\$\&}, T_{\%'(}, H, W_{)}, W_{*}, R, F]^{+} \quad (2)$$

The meaning and calculation of each variable are shown in Table 2.

	Calculation of the	physical meaning of variable names
$T_{\%\$\&}$	Minimum temperature	$T_{\%\$\&} = min\ Temmpratur\ e$
$\%'($	T maximum temperature	$\%'(\ T = max\ Tempratur\ e$
H	Average humidity	$H = \frac{1}{96}\sum_{\&-/.}^{,-} Humidity$
$W_{)}$	Average wind speed W_0	$= \frac{1}{}\sum_{\&-/.}^{,-} Wind\ Speed\ 96$
W_1	Main wind direction	$W_1 = \frac{1}{}\sum_{\&-/.}^{,-} Wind\ Direction\ 96$ and assign it to one of eight directions, including east, northeast and north
R	Time corresponding to the first moment when the sunrise time irradiance shifts	from 0 to non-zero
F	The time corresponding to the first moment when the irradiance time	at sunset shifts from non-zero to 0

Table 2 Physical meaning and calculation of meteorological condition vectors

In order to accommodate the need to calculate the inter-sample distances in cluster analysis, it is necessary to transform the data of different dimensions to the same measure. Therefore, for each dimension $x \in X$ above, it is restricted to the $[0,1]$ interval using $x =$ ς

%'((

Inside.

2$ZV[\

Clustering is a typical unsupervised learning method. Its training data does not contain information about its categories, but rather uncovers potential categories based solely on the relationships between its attributes. In clustering tasks, samples with similar properties are usually grouped into similar categories, and those with more different properties are grouped into different categories. Mathematically, correlation coefficient, similarity coefficient and distance are commonly used to measure the relationship between different variables. For unordered sample sets, commonly used clustering methods include systematic clustering, one-at-a-time formation clustering, K-level stepwise formation clustering, and moving center clustering.[12] For the large sample size in the dataset of this paper, it is necessary to choose a less computationally intensive solution to meet the computing time requirements for fast prediction. The moving center clustering method has been empirically proven to be effective for large data sets[12] and is computationally simpler. Therefore, the moving center clustering method is used.

For the previously mentioned feature vector X describing the daily climate conditions$, the Euclidean distance d is used to reflect the differences existing between the different samples and as a basis for clustering. Where the Euclidean distance between the ith sample and the jth sample is defined as[3]

$$d\$3 = \mathrm{LBM}_{(\in 6} x\$ - x^{04} \quad (3)$$

According to the target classification number of 4 classes, the implementation steps of the moving center clustering method are as follows.

Step 1:Select 4 samples from n samples as the center of the class by random selection.

Step 2:For all samples in I, calculate their distances to each class d and classify them to the distance most

Nearby categories.

Step 3:Based on the newly classified categories, the arithmetic mean of the samples in the class in each dimension is calculated and used as the

New class centers.

Step 4:Repeat the steps of Step 2 and Step 3 to classify all samples repeatedly.

The cycle is stopped when one of the following two conditions is met.

（1）Achievement of the number of iterations specified.

（2）The same division is obtained for two successive iterations.

After the cycle stops, the last classification is used as the clustering result.

The clustering task is executed according to the above algorithm, and the samples are divided into four natural clusters for

preservation in terms of natural days.

$1&KLMNOPQ+RSIJ

By defining the period when the photoreceptor does not sense sunlight as the night time period and the period when the photoreceptor value is not 0 as the day time period, the meteorological and irradiance data collected by the PV plant can be divided into two parts: daytime data and nighttime data. The irradiance observed during the night time is stable at 0, so there is no practical meaning to predict the irradiance at night. However, if there is a significant difference between the diurnal and nocturnal weather data, it helps to train the potential reflection between weather-irradiance when adding the nocturnal weather data to the training sample; if the difference is not enough to distinguish the diurnal and nocturnal data, it means that the nocturnal period data in the training sample affects the reliability of the model. Therefore, a classification task is attempted to determine whether the disparity existing between their weather data is sufficient to achieve achieving the distinction between day-night.

The essence of the classification task is to learn to obtain an objective function f such that for each attribute set x it can be mapped to a class label y[13]. The classification model predicts the class label of an unknown sample by fitting a potential mapping relationship between the class label and the attribute set in the input data. In this task, this model can be used to describe the gap between classes and for distinguishing objects contained in different classes.

> For a set of linearly divisible samples, i.e., there exists a hyperplane such that the distribution of positive examples in the separation

the positive class side of the hyperplane. where the distance between the sample point and the hyperplane can be used to express the degree of confidence in its classification as a positive class. As in Figure 1, for the classification rule represented by the line f, the farther

the point A is from the hyperplane, the greater the degree of certainty in predicting it to be a positive class.

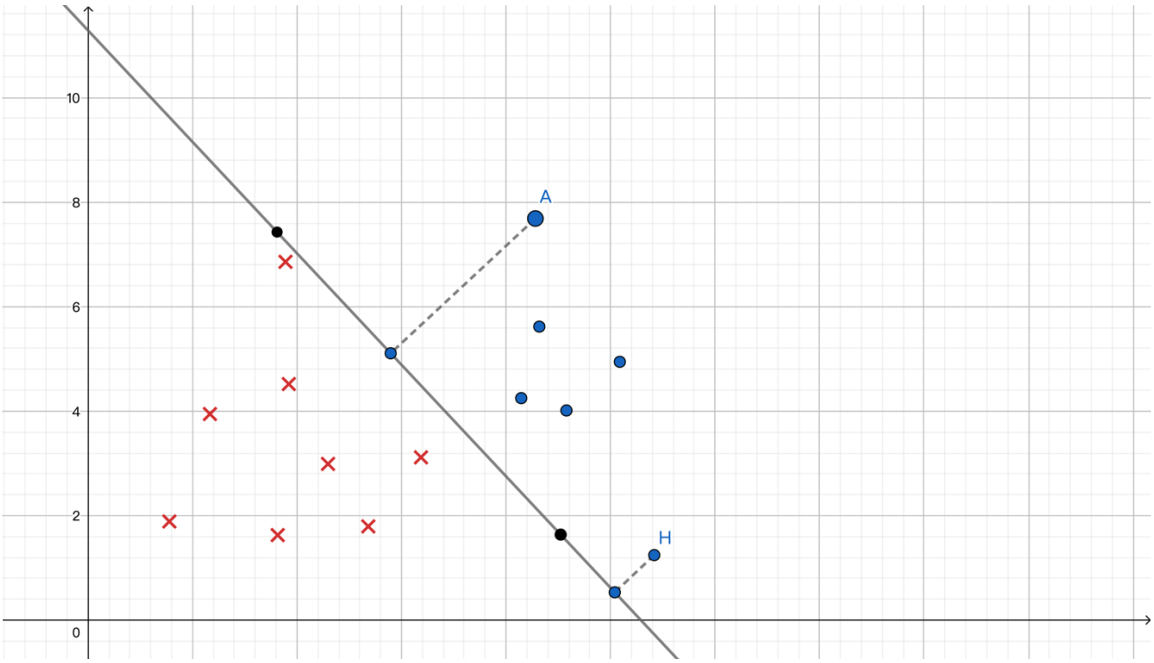


Figure 1 Examples of classification laws and degrees of certainty

Under the setting that the hyperplane $w \times x + b = 0$ remains determined, $|w \times x + b|$ can relatively represent its distance from the hyperplane, and whether its sign is consistent with that of the class marker y indicates its classification correctness. Considering that the above results change when (w, b) undergoes a proportional change, it is useful to use

$$r_{\$} = \frac{w}{\mathbb{V}|w|\mathbb{V}} \cdot x_{\$} + \frac{b}{w} \tag{4}$$

denotes the distance between the sample point x and the hyperplane (w, b). At this point if (w, b) is changed proportionally, the distance from the sample point to the hyperplane remains unchanged. If the geometric interval of the hyperplane about the training set is defined as the minimum geometric distance from all sample points in the sample set to the hyperplane, i.e.

$$r = min\, r_{\$} \tag{5}$$

For the support vector machine approach, then, the aim is to find a method of separating hyperplanes that can correctly partition the training set and obtain the largest possible geometric interval, where the sample points closest to the hyperplane are called support vectors. In the process of finding the hyperplane, only a small number of support vectors act as constraints on it.

Considering that the distinction between day and night is essentially a distinction between meteorological states, and that the five-dimensional data {wind speed, wind direction, temperature, humidity, and pressure} serve as a low-dimensional description of meteorological states, and that there is noise and idiosyncratic points in the data, the data are linearly indistinguishable in the original dimension, making the support vector machine (SVM) learning method in the linearly separable problem no longer applicable, and an attempt should be made to use a nonlinear model that correctly classifies the sample data. The SVM learning method in this problem uses the kernel function to design a nonlinear transformation ϕ to map

the sample $x \in X$ under the original space to the sample $z \in Z$ in the new space satisfying $z = \phi(x)$. After the transformation, the samples are linearly divisible in the new higher dimensional space, thus the linearly indivisible problem in the original space can be converted into a linearly divisible problem in the new space. Considering the existence of idiosyncratic points and noise, the relaxation variable ϵ and the penalty coefficient C are introduced so that for some idiosyncratic points, the correct classification can be obtained by adding the function interval to the relaxation variable. And the larger the penalty coefficient $(C > 0)$, the larger the penalty for the case of misclassification and the stricter the trained hyperplane.

Based on this classification method, the nighttime observations are assigned label 1 and the daytime observations are assigned label 0. 70% of the total sample data are selected as the training set, on which a support vector method is used to try to train the best possible classification hyperplane, and the classification performance is evaluated in the test set. To exclude the effect of different dimensions of the selected meteorological data, the same random number of seeds is set and 22 combinations of five dimensions of temperature, wind speed, wind, humidity and pressure are subsequently taken out for evaluation. To find the stronger classification criteria

The accuracy is set to 5.0, the penalty factor is set to 5.0, and the linear kernel function is used for binary classification training. The classification accuracy is as

Figure 2.

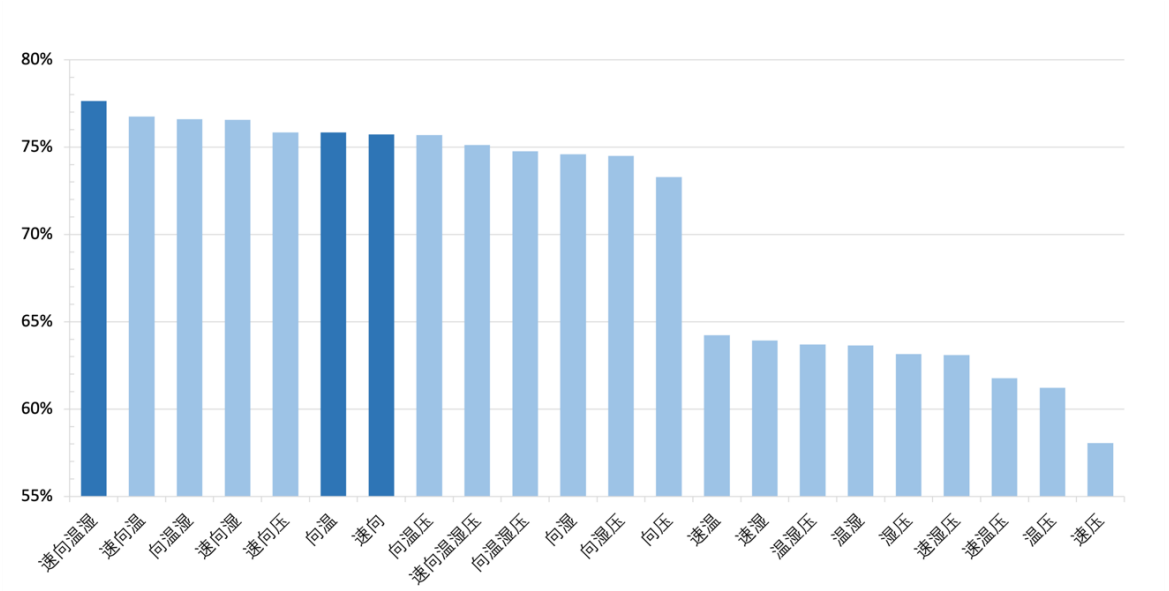


Figure 2 Classification accuracy with different data combinations

The training results show that, among these 22 combinations, the best classification result is achieved with 77.64% accuracy when four dimensions of {wind speed, wind direction, temperature, and humidity} are used, and the worst classification result is achieved with 58.05% accuracy when only two dimensions of {wind speed and pressure} are used. Also considering that the number of both daytime and nighttime samples in the training and test samples are equal, the classification effect of this model is not significantly better than that of random guesses.

Meanwhile, the training accuracy can be maintained above 75% with eight combinations, among which using {wind speed, wind direction} and {wind direction, temperature} can achieve better classification effectiveness with minimum sample data, with accuracies of 75.72% and 75.84%, respectively. This indicates that wind and temperature are the most significant differences in climate between day and night, while pressure variations can interfere with diurnal distinctions. Humidity brings no significant effect on the task of diurnal differentiation.

This suggests that when attempting the task of diurnal differentiation, physical quantities related to describing wind and temperature should be selected whenever possible, and physical quantities describing air pressure should be ignored.

However, it is worth noting that the accuracy of the four dimensions {wind speed, wind direction, temperature, and humidity} in the classification was only improved to a limited extent, from 75.12% to 77.64%, after excluding the disturbing factor of barometric pressure. No matter what combination is used, the classification accuracy cannot exceed 78%, which indicates that trying to use the existing meteorological data, the support vector machine method makes it difficult to classify no less than 22% of the samples with multiple values and misclassification, and to carry out accurate segmentation of diurnal and nocturnal time only by

The method of manually adding labels is used to differentiate. When training the prediction model, then the nighttime weather data should be excluded.

$T&UVWXYZ

The raw data are six-dimensional meteorological data of {wind speed, wind direction, temperature, humidity, pressure, and irradiance} at 15-minute resolution for a 24-hour period. Based on the above clustering analysis and support vector machine results, the daily diurnal data were extracted from the total data and the samples were grouped into different categories according to the daily category labels. According to their relative temporal order, they are formed into a continuous signal sequence to form a sample of meteorological data under different weather types.

5#'(6789:-;<=>? @

Considering the meteorological data as a column of random signals, the distribution of the sequence $(y_{",}, y_{"#/}, y_{"#4}, \ldots, y_{"#0})$ is independent of t for any s according to the requirement of smoothness on the signal. For meteorological data, its seasonal variation may affect its smoothness.[14] However, considering the decomposition of the clustering results for seasonal variation in the previous section, the reconstructed meteorological data series still has some smoothness. In order to further improve its smoothness for smooth signal processing, the signal sequence is split into several shorter smooth sequences and cut into a total of 9984 records with a length of no more than 104 natural days. At this point, the signal does not show significant trends due to seasonal changes. And since an arbitrary signal can always be represented as a superposition of harmonic signals and white noise, the cut signal sequence is considered suitable for studying the theoretical possibility of its being fitted by a finite number of simple harmonics.

$%&KL [\]^[- __]^[`abcdef78gcd

Consider the cut smooth sequence as the superposition of a series of harmonic sequences with white noise, i.e., the weather signal $X_b(t)$ can be

$$X(t) = \sum_{7-/}^{8} A_7 \cos(\omega_7 t + \phi_7) + e_{"} \tag{6}$$

fit, where $\phi_7 \in [-\pi, \pi]$ satisfies a uniform distribution and $e_{"}$ is a white noise process. For the sequence $X(t)$, the power spectrum is

$$P(\omega) = \frac{\pi}{2} \sum_{7-/}^{8} A_7 \delta(\omega - \omega_7) + \sigma^4 \tag{7}$$

where σ^4 is the variance of $e_{"}$.

Therefore, by performing a power spectrum analysis on the smooth sequence X b, it can be determined from the analysis that the composition $X(t)$

Power and frequency parameters of harmonic processes.

The SFGPS (Simple and Fast algorithm to Genetrate Power Spectrum) algorithm states that in contrast to the Welch method of calculating the power spectrum by segmenting the overall signal using a sliding window, if the periodogram is filtered directly in the following way.

$$\begin{matrix} {}^{A4}(m) = \\ +\sigma^4 \end{matrix} \begin{matrix} \mathrm{i}A^4(m), \quad if A^4(m) > \mathrm{j}A\mathrm{j}4\mathrm{j} \\ {}^4\sigma, \quad else \end{matrix} \tag{8}$$

The time complexity of computing the power spectrum can be reduced from $O(N^4\ logN)$ to $O(N\ logN)$. and the superposition for discrete harmonic processes is consistent with the effectiveness of the Welch method. It should be noted that the filtering method has certain maladaptive situations for sequences with very significant fluctuations, such as when the variance σ^4 is too large, the filtering will instead appear counterproductive. To solve this problem, the original signal can be deflated to a certain extent to achieve the desired [effect15].

In addition, the AFAGPS (Accurate and fast algorithm to Generate Power Spectrum) can also capture the power spectrum very accurately. The AFAGPS algorithm makes three important improvements.

1. The time complexity is reduced from $O(N^9)$ for solving the Yule-Walker equation to $O(NNO)$ by bypassing the direct solution for the Yule-Walker equation and instead rewriting it by using the fast Fourier transform and the inverse transform to obtain the required parameters with respect to the amplitude ($NlogN$).

2. The method of estimating the correlation function is changed from the traditional asymptotic unbiased estimation

$$Rq(m) = \frac{1}{N - |m|} \sum_{7=} (x(k) - \mu)(x(m+k) - \mu), \quad m = 0,1,2,\ldots \tag{9}$$

Change to circumferential convolution

$$Rq(k) = \frac{1}{N-k} z(k), \quad , k = 0,1,2,\ldots, K < v\frac{N}{2} \tag{10}$$

where $z(n)$ is the circular convolution of $x(n)$ and its inverse sequence $x(-n)$. At this point the time complexity is reduced from $O(N^4)$ to $O(NlogN)$ and can accommodate longer data lengths.

3. [4] The approximate power estimated with $\&R(0), R(1), \ldots, R(K)'$ at the setting $K \&\leq y$ is sufficiently large

The spectrum can be used to estimate the true power spectrum at[15].

The power spectrum analysis of the signal sequences in each of the five meteorological dimensions was performed with the above two algorithms, and the intercepted signal length was 9984, and the

obtained results are shown in Fig. 3. According to the power spectrum analysis results, the five-dimensional meteorological data can indeed be fitted with discrete or continuous π since their maximum cutoff frequencies are all far from the center π, and the power spectrum results under the two algorithms have strong consistency. harmonic processes for fitting. Under the AFAGPS algorithm, wind speed and wind direction exhibit that they may consist of continuous harmonics, and temperature, humidity, and pressure may consist of discrete harmonics; the results under the SFGPS algorithm verify that the three dimensions of temperature, humidity, and pressure may consist of discrete harmonics. For the wind speed and wind direction dimensions, if more discrete harmonics are used, it may be possible to achieve a fit while tolerating some errors.

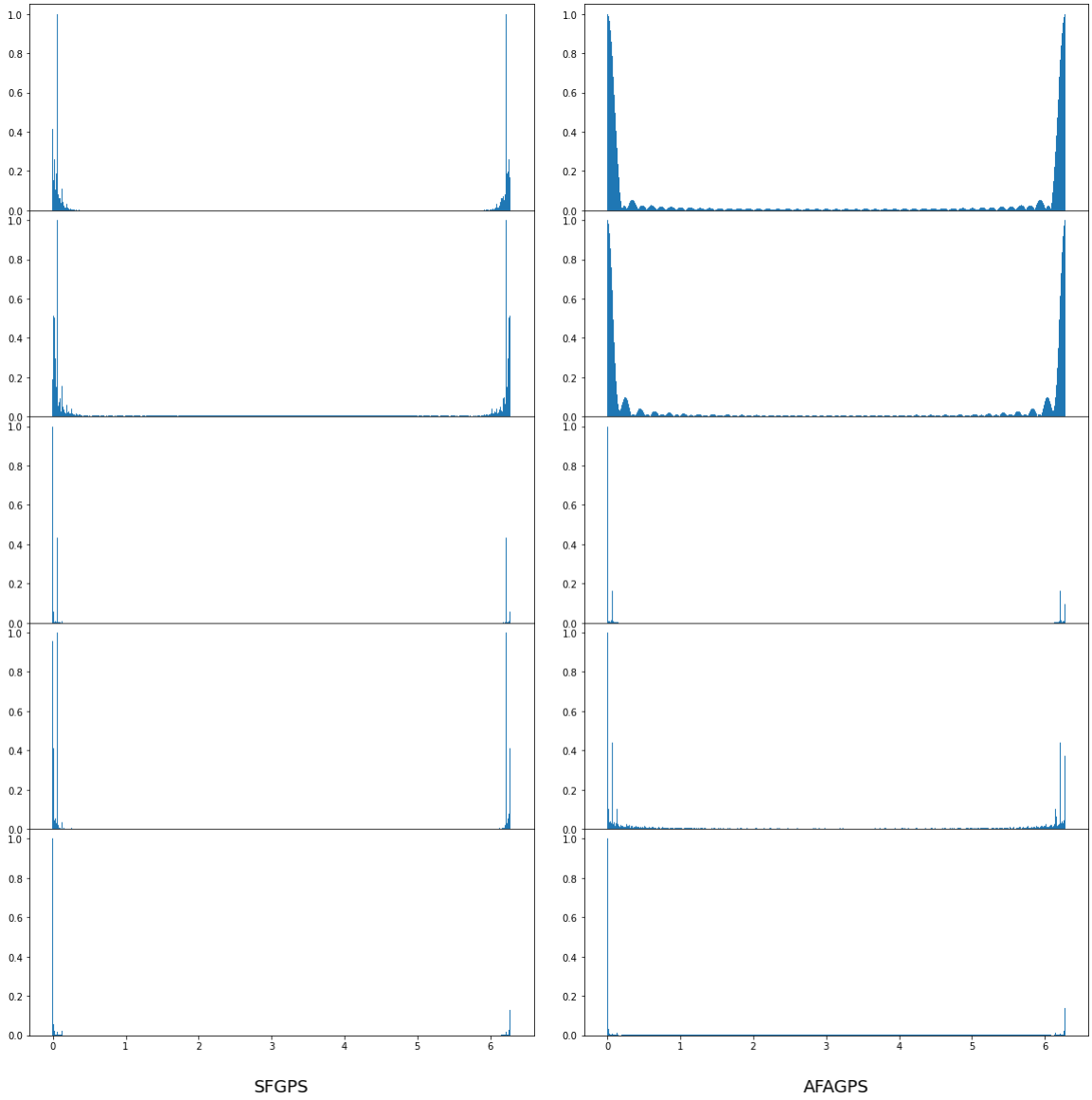


Figure 3 Power spectrum images of five-dimensional meteorological data under SFGPS and AFAGPS algorithms

$1&hUiajkeUVIm

For a signal of finite length, when the support of its spectrum is bounded, it is determined that its frequency band is finite, that is, there exists a maximum cutoff frequency. For a signal in the time domain, the classical Shannon's sampling theorem states that if its maximum cutoff frequency $f_>$ exists, then to recover the original sequence from the sampled sequence of the signal, it is necessary to

To sample the column signal $xMn\Delta O$ the sampling interval $\Delta \leq$. At this point the signal can be sampled using the

4?!

The following reconfiguration

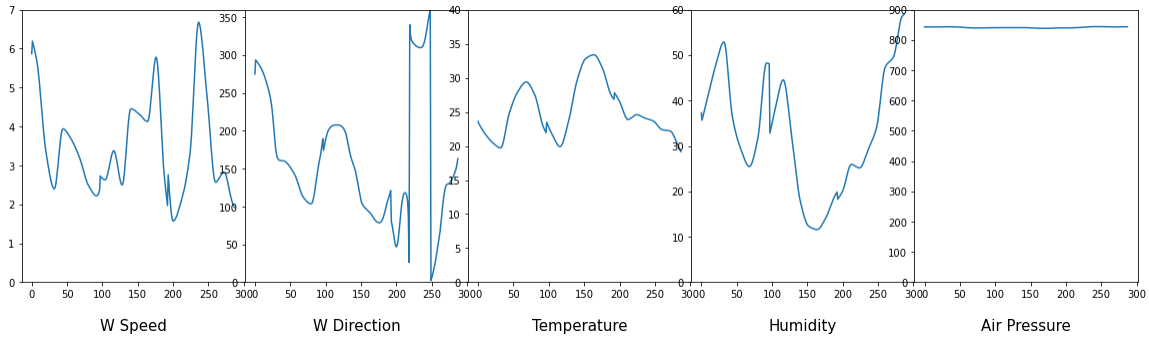
$$x(t) = \sum_{n=-\infty}^{+\infty} x(n\Delta) \frac{\sin \frac{\pi}{\Delta}(t - n\Delta)}{\frac{\pi}{\Delta}(t - n\Delta)} \tag{11}$$

In other words, from the signal science point of view, for a smooth random sequence, it can be approximated by a harmonic process. And for an observation of a harmonic process, its sampling frequency satisfies the requirements of Shannon's sampling theorem before this observation can be recovered as a complete harmonic process, and thus this harmonic process can be used to make a reasonable prediction of future data.

If the sampling frequency is lower than the minimum sampling frequency required for these high-frequency components, there is a high probability that these high-frequency components will be lost in the forecast. These high-frequency harmonic components often correspond to sudden changes in weather and the occurrence of unconventional phenomena in meteorology. Existing prediction models are generally less effective for such weather conditions, largely due to the fact that the frequencies of the observed data used do not meet the requirements for recovering the high-frequency components, and only the low-frequency components can be better restored, thus allowing only reasonable predictions of routine weather conditions. If we want to make timely and accurate predictions of extreme weather, we need to recover the high-frequency components, which requires researchers to use higher-frequency meteorological observation methods or interpolate and refine the existing data to artificially increase their sampling frequency.

The distribution of the five-dimensional meteorological data over three consecutive days is shown in Figure 4, with the starting coordinates of the y-axis set to 0. Wind is more random and discontinuous than temperature and pressure in climate, and is more prone to sudden changes.

changes, so the sampling frequency required for the two dimensions of wind speed and wind direction is high. The power station where the data used in this paper is located is in the northern region of China, and the intra-day variation of its humidity is very obvious, so a higher frequency of sampling data should be required to finely



describe the sudden changes that occur. In contrast, the temperature changes throughout the day are relatively smooth, with small and slow changes, while the pressure always remains at similar values, both of which should correspond to relatively low sampling frequencies.

Figure 4 Line graph of five-dimensional weather data over three days

Using the power spectrum data calculated by AFAGPS and SFGPS, the sampling frequencies required for the signal sequences of different dimensions are quantified. After filtering out the harmonics with power less than 1% of the maximum power, the frequency corresponding to the last harmonic with power not 0 is the cutoff frequency. The maximum cutoff frequency, minimum sampling frequency, minimum sampling resolution (sampling interval), and post-interpolation resolution of the signal sequence for the five-dimensional meteorological data of a station under both algorithms are shown in Table 3.

Dimensionality	Maximum cut-off frequency (AFAGPS)	Maximum cut-off frequency (SFGPS)	Minimum sampling frequency	Minimum sampling interval partition	Interpolation posterior interval partition
Wind speed	0.454π	0.065π	0.908π	2.203min	1.5min
Wind direction	0.525π	0.115π	1.050π	1.904min	1.5min
Temperature	0.023π	0.021π	0.046π	43.478min	15min
Humidity	0.242π	0.042π	0.484π	4.132min	3min
Pressure	0.042π	0.042π	0.084π	23.810min	15min

Table 3 Maximum screenshot and minimum sampling interval for five-dimensional meteorological data

There is a large difference between the maximum cutoff frequencies calculated by SFGPS and AFAGPS for {wind speed, wind direction, humidity}, while for {temperature, pressure}, there is some error in the calculated maximum cutoff frequencies, which is to some extent caused by the difference in the filtering criteria selected by the two

algorithms. For the signal sequence, the sampling frequency can be used to restore the original signal as long as it is higher than the minimum frequency limited by the sampling theorem. Therefore, the larger cutoff frequency under both algorithms is chosen as the cutoff frequency of this signal, and the minimum sampling frequency and the minimum sampling resolution are calculated from it. The results show that the sampling resolution required for each of the three dimensions {wind speed, wind direction, and humidity} is higher than the 15-minute resolution of the original data, necessitating interpolation and fine-tuning. In order to fit the resolution of the original data, an interpolation resolution no greater than the minimum resolution and divisible by the original resolution was chosen. In contrast, the temperature and pressure were sampled at a resolution greater than the 15-minute requirement and did not require special interpolation. For the locations where interpolation plus refinement is required, the adjacent observations are considered $(x)_{=,y=}$ with $(x_{/,y/})$, the data were spiked according to the resolution requirements using linear interpolation

$$y = y + \frac{y/ - y =}{x/ - x=}(x - x \quad) \qquad (12)$$

Reconstitute the signal sequence that satisfies the requirements of Shannon's sampling theorem.

A#BC DDDEF -GHI<=

$%&n:o<

Multilayer Perception is a complex computational network consisting of a large number of neurons

network with strong nonlinearity. One of the neural networks trained using error backpropagation is compared with the learning sample in the output layer by processing the sample data through the weights in the hidden layer.[16] Based on the error, it enters the backpropagation process, adjusting the weights of each unit along the fastest direction of error decline, and continues to cycle until the training round requirement is reached or the error is lower than the desired training error.

The conventional multilayer perceptron can only pass a single vector in one input and cannot take into account the time-series relationship between the training data. In order to make the multilayer perceptron adaptable to the need of inputting multiple vectors and preserving their time order in irradiance prediction, its input vector approach is improved to obtain the multivariate multistep multilayer perceptron model (MMMLP). Where multivariate refers to the selection of multiple dimensional autovariants, and multistep refers to the prediction using multiple time steps of training data to obtain multiple time steps of prediction data. For the standard MLP model. The input layer requires a single vector input, so multivariate multi-step MLP flattens the input data so that the data of multiple time steps are combined in one output. Data for the first 8 hours of the prediction time are selected to predict irradiance for the next 4 hours. The Adam optimizer and the ReLU activation function are selected to perform a 500-cycle learning process, and the training error is estimated using the mean square error (MSE).

In order to fairly compare the prediction performance, the unprocessed training data and the restructured training data in this paper are used as the learning samples of MMMLP separately, keeping other training parameters the same, and their prediction performance is evaluated in the same test data.

$1&78gpcd

The optimized MMMLP training results and the original MMMLP training results are shown in Fig. 5, intercepting the irradiance predictions within the first 11 days. It is noted from Fig. 5 that both methods achieve better predictions in general, both models are more conservative in peak predictions, and both have some ability to detect anomalous weather.

From the comparison of the two, it is clear that the optimized MMMLP has an overall advantage in terms of prediction performance. Within the first day, the prediction results of both algorithms differed significantly from the measured results. This is largely due to the small number of meteorological samples used on the first day. In the second day, the prediction error of the optimized MMMLP is significantly smaller than that of the original MMMLP, indicating that the optimized method is more well adapted to the prediction conditions with small data volume. On the third day, the original MMMLP showed a serious lack of prediction accuracy in the face of the special weather when the peak irradiance decreased, and fluctuations against the natural pattern were observed; in contrast, the optimized MMMLP fitted well with the measured data and predicted the peak of the day accurately.

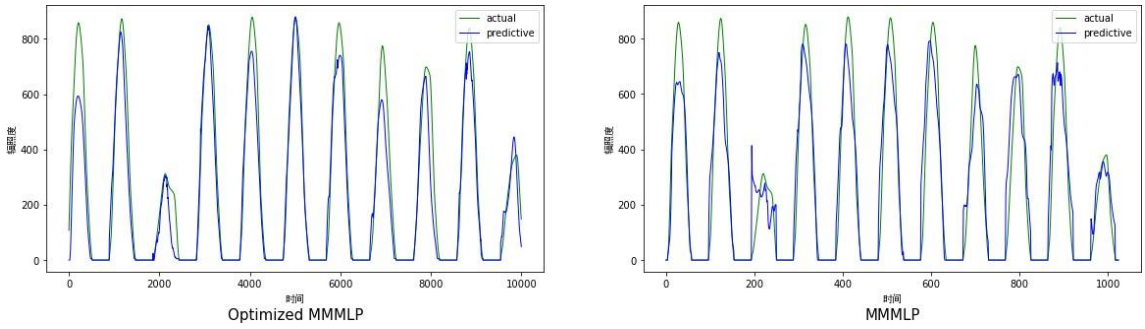


Figure 6 Predicted and actual values of optimized MMMLP and original MMMLP

In addition, the prediction accuracy of the optimized MMMLP method is very high under meteorological conditions such as the fourth and sixth day, and is significantly improved compared with the original method, especially in terms of peak size and time prediction. On the whole, all the peaks predicted by the original method appear to be significantly early, while the optimized MMMLP method is generally accurate in predicting the peak times.

From the numerical results of the data analysis, the performance of the optimized MMMLP is also significantly improved compared to the MMMLP. Using the mean absolute error to evaluate the prediction

accuracy of the two algorithms, the mean absolute error of the optimized MMMLP method is 44.732 and the mean absolute error of the original MMMLP method is 56.002. The optimized method reduces the mean absolute prediction error by 20.12% based on the original method. Defining the maximum value of daily irradiance as the peak irradiance of the day, the average absolute error of the peak of the optimized MMMLP method is 94.863, which is much lower than the average absolute error of the peak of the original MMMLP method of 451.94. Figure 6 shows the histogram of the frequency distribution of the peak prediction error under the two algorithms.

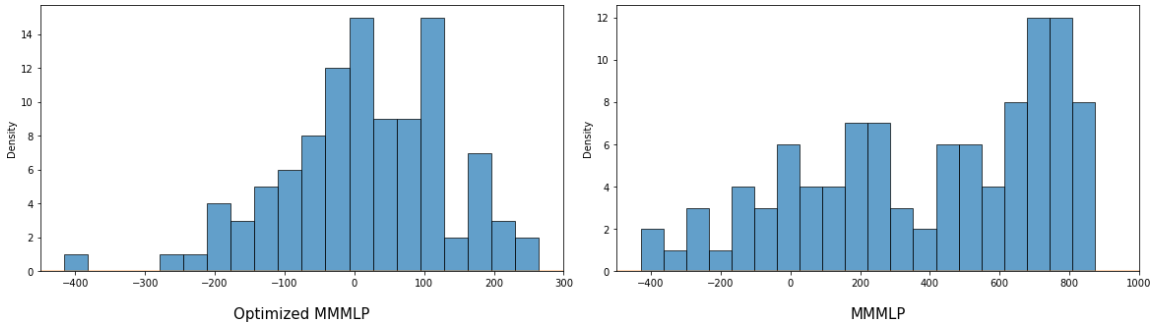


Fig. 6 Histogram of peak prediction error of optimized MMMLP and original MMMLP

Obviously, the original MMMLP shows a clear tendency of conservative prediction, and there are a large number of samples with peak prediction errors higher than 600, i.e., there are more peaks present without successful prediction. And most of the samples have positive prediction errors, which indicates that the model has an obvious conservative tendency in predicting peaks. In the model-optimized MMMLP, the errors are closer to the 0-centered normal distribution, and the prediction errors are generally smaller. Therefore, the optimized MMMLP is also much more effective than the original MMMLP method in terms of peak prediction.

J#KL

Based on the background of irradiance prediction and for the summary of three common classical algorithms, a better irradiance prediction scheme based on historical data is implemented in this paper using pattern recognition, signal processing and neural network methods. Using clustering methods, daytime samples with similarities are reorganized into continuous sequences, and the differences between daytime and nighttime meteorological data are analyzed using support vector machine methods, and it is concluded that daytime and nighttime data need to be processed separately. Subsequently, the conditions for meteorological data of different dimensions to have predictability when used as a smooth signal are explored from a signaling perspective, and the sample data are complemented and refined according to their requirements for sampling frequency as revealed in the power spectrum analysis. Considering the strong nonlinearity of irradiance prediction, a multivariate multistep multilayer perceptron method with reorganized data optimization is applied to predict irradiance, and a prediction model with significantly better prediction results than the original one is obtained. It is worth noting that the prediction scheme in this paper is derived entirely from the mathematical model trained based on historical data, without reference to satellite cloud maps and meteorological forecast data, which are highly relevant to irradiance prediction, and more abundant and diversified data can be used in future work to create possibilities for further improvement of model performance.

MNOP

[1] Wu Shuo. A review of the research on power prediction methods for photovoltaic power generation systems[J]. Thermoelectric Power Engineering,2021,36(08):1-7. DOI:10.16146/j.cnki.rndlgc.2021.08.001.

[2] Wang Fei. Grid-connected photovoltaic power generation prediction method and system[D]. North China University of Electric Power, 2013.

[3] Wu Tianyu. Research on PV output prediction considering meteorological influences[D]. Yanshan University, 2021. DOI:10.27440/d.cnki.gysdu.2021.001196.

[4] Skoplaki E , Boudouvis A G , Palyvos J A . A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting[J]. Solar Energy Materials & Solar Cells, 2008, 92(11):1393-1402.

[5] ZHU X, JU RONGRONG, PROCEDURE, DING YY, ZHOU H. Combined numerical weather prediction and ground-based cloud map for PV ultra-short-term power prediction model[J]. Power System Automation,2015,39(06):4-10+74.

[6] Yang Liufeng. A review of research on the application of artificial intelligence algorithms in photovoltaic power generation forecasting[J]. Solar Energy,2020(08):30-35.

[7] Zhou H, Zhu X, Jinshan H, Zhu T, Zhang Xuesong, Wei HK. Multi-model prediction of ultra-short-term solar irradiance[J]. Chinese scientific and technical papers,2017,12(23):2695-2700.

[8] Kassianov E , Long C N , Ovtchinnikov M . Cloud Sky Cover versus Cloud Fraction: Whole-Sky Simulations and Observations[J]. Journal of Applied Meteorology, 2005, 44(1):págs. 86-98.

[9] Ye Lin. Research on photovoltaic power generation prediction based on WRF model output[D]. Ningxia University,2018.

[10] Li Jing,Xu Honghua,Zhao Haixiang,Peng Yanchang. Dynamic modeling and simulation analysis of grid-connected photovoltaic power plants[J]. Power System Automation,2008,32(24):83-87.

[11] ZHONG Zhifeng,ZHANG Yi,ZHANG Tiantian,YANG Chenxi,SU Yong. A simple study on short-time irradiance prediction[J]. Computer Measurement and Control,2017,25(07):181-185.DOI:10.16526/j.cnki.11-4762/tp.2017.07.045.

[12] Zhang Runchu. Multivariate statistical analysis [M]. Science Press, 2006:266.

[13] Tweedale J W , Jain L C . Advances in Modern Artificial Intelligence [M]. Springer International Publishing, 2014.

[14] Hyndman R J , Athanasopoulos G . Forecasting: Principles and Practice[J]. London: Bowker-Saur. Pharo, 2014.

[15] Ruan Jishou. Lecture Notes in Signal Science, 2022.

[16] Heaton J . Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks. 2015.

q r

Tomorrow's world, for whom mankind is curious and fascinated enough to write a secret history related to science. With the eyes lent to me by mathematics and statistics, I see how those proven tools can help us catch a reed of certainty in a sea of randomness. But even the most powerful tools cannot face the reality that we cannot escape in the fourth dimension: we are stepping into the unknown, and time does not change its pace depending on whether you see the future or not.

I can recall one spring afternoon when I was lost in thought and saw the grass outside the second main building shining brightly; I can also remember how the sparks in my chest beat in the dark night on the deserted streets; I can also recall how sincerity and warmth flowed in the quiet room, allowing all dead things to germinate. They all remind me that I have plowed hard at the reality under my feet, devoting my whole passion, my whole body to lift the big brush of life and write hard on the pages of my book. The stinging loss, the joy of discovery, the anxiety of waiting, the jumping happiness ..... that had occurred in my field, I did not let it flow carelessly. This is the best answer I can give to time.

The journey was more lonely than I had ever imagined, and more beautiful than I had ever imagined. My strength and faith would have been compromised without the time I spent in the office of Professor Gishou Ruan, without the frank conversations with my beloved teachers, and without the scientific and empirical armament that Nankai Mathematics and Nankai Statistics gave me. I will always be grateful for the protection and support they brought to my independent mind.

How to predict the future is my scientific problem, while how to face the gap between the known and the unknown is my enduring life proposition. She polishes my soon-to-be-dusty light back into

precious display when I am about to inadvertently fall into the currents of disillusionment; my parents clutch my hand and guide me to bag my failures and start over forever; my friends, who always pull my body with a trickle of trust. These connections have allowed me to regain meaning that was close to decay and anchored me in moments of vulnerability. Together, they have drawn the underpinnings of my life.

When I look back at the complete story of the self-promise I left behind in high school, its plot has long exceeded all my imagination. And I can finally say, without too much shame, that I kept the promise of sincerity and courage that I left myself all those years ago. My life at 94 and 92 Weijin Road will be gently folded and put in my shirt pocket. To my country, my time and my life, in sowing seeds for you, I have buried my deepest heart's desire in the ground as well, and I will not hesitate to get my hands dirty for you. There is my torch in the mists of the future, and you will travel far away with me.