

## Article

# A Machine-Learning Approach Combining Wavelet Packet Denoising with Catboost for Weather Forecasting

Dan Niu<sup>1,2,\*</sup>, Li Diao<sup>3</sup>, Zengliang Zang<sup>4</sup>, Hongshu Che<sup>1</sup>, Tianbao Zhang<sup>5</sup> and Xisong Chen<sup>1,2</sup>

<sup>1</sup> School of Automation, Southeast University, Nanjing 210096, China; 220191565@seu.edu.cn (H.C.); 101010097@seu.edu.cn (X.C.)

<sup>2</sup> Key Laboratory of Measurement and Control of CSE, Ministry of Education Research Laboratory, Nanjing 210096, China

<sup>3</sup> School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; diaoli@sjtu.edu.cn

<sup>4</sup> Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China; zangzengliang@nudt.edu.cn

<sup>5</sup> School of Suzhou United Graduate College, Southeast University, Nanjing 210096, China; 220204780@seu.edu.cn

\* Correspondence: 101011786@seu.edu.cn

**Abstract:** Accurate forecasting of future meteorological elements is critical and has profoundly affected human life in many aspects from rainstorm warning to flight safety. The conventional numerical weather prediction (NWP) sometimes leads to unsatisfactory performance due to inappropriate initial state settings. In this paper, a short-term weather forecasting model based on wavelet packet denoising and Catboost is proposed, which takes advantage of the fusion information combining the historical observation data with the prior knowledge from NWP. The feature selection and spatiotemporal feather addition are also explored to further improve performance. The proposed method is evaluated on the datasets provided by Beijing weather stations. Experimental results demonstrate that compared with many deep-learning or machine-learning methods such as LSTM, Seq2Seq, and random forest, the proposed Catboost model incorporated with wavelet packet denoising can achieve shorter convergence time and higher prediction accuracy.

**Keywords:** weather forecast; Catboost; wavelet packet denoising; machine learning



**Citation:** Niu, D.; Diao, L.; Zang, Z.; Che, H.; Zhang, T.; Chen, X. A Machine-Learning Approach Combining Wavelet Packet Denoising with Catboost for Weather Forecasting. *Atmosphere* **2021**, *12*, 1618. <https://doi.org/10.3390/atmos12121618>

Academic Editors: Avelino F. Arellano and Yunheng Wang

Received: 21 October 2021  
Accepted: 1 December 2021  
Published: 4 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Weather prediction is of great importance and can affect some aspects of daily life, such as air quality, travel plans, energy supply, and so on [1–4]. A conventional prediction method is numerical weather prediction (NWP) method, which solves the numerical solutions of atmospheric hydro-thermo dynamic equations to predict meteorological dynamics [5–7]. However, unsatisfactory prediction results are obtained if inappropriate initial states are set [7,8]. Moreover, conventional NWP-based approaches do not take full advantage of vast amount of existing historical observation data [9]. As the observation techniques develop and the historical meteorological data grow bigger and bigger, some purely data-driven approaches are expected to be introduced into weather forecasting [10]. They do not solve complex differential equations, but quickly model the meteorological dynamics by learning from a large amount of historical datasets.

It is well known that several machine-learning methods including artificial intelligence (AI) methods have been demonstrated to be powerful methods for weather forecasting [9–13]. For one-dimensional timeseries meteorological data, some data-driven machine-learning approaches have been presented for weather forecasting. In [14], autoregressive integrated moving average artificial neural networks (ARIMA-ANN) combined with ARIMA-Kalman is proposed to predict the wind speed and show the effectiveness of the hybrid model. In [15], global radiation is effectively predicted by a hybrid ARIMA/ANN model. In [16],

Gaussian processes method is employed to revise NWP results and achieve a more accurate prediction of the 24 h wind energy. Moreover, a deep hybrid model in [17] is utilized to predict a series of weather related elements. In [18], the relationship between variables is captured by auto-encoder, and rainfall is forecasted by the multi-layer perceptron (MLP). In [19], multi-layer perceptron with spatial–temporal attention is proposed for wind speed and wind direction in Beijing. In addition, the long short-term memory (LSTM) model is also effective for weather forecasting [20,21]. Seq2Seq and its variations, which are well known in the field of natural language processing, can be migrated and applied to solve weather prediction problems [22].

On the other hand, some deep-learning methods employed the two-dimensional radar echo maps to achieve a short-term weather forecast. In [8], a short-range forecast based on radar echo maps is viewed as a spatiotemporal sequence forecasting problems, and the convolutional LSTM (ConvLSTM) method was proposed to solve it. Furthermore, the same author proposed the trajectory GRU (TrajGRU) to further improve weather forecast accuracy [23]. It utilized not only location-invariant structure in ConvLSTM, but the location-variant recurrent structure to efficiently depict the motion patterns such as rotation and scaling in radar echo maps. In addition, the memory in memory (MIM) network [24], the improved predictive recurrent neural network (PredRNN++) [25,26], and generative adversarial gated (GAN) recurrent unit models [27,28] were also proposed to handle the limitations in ConvLSTM model and enhance the prediction performance.

Although the accuracies of above machine-/deep-learning methods are high, these purely data-driven models have the limitations that they ignore the important prior knowledge from NWP and may not fully capture the spatiotemporal dynamics of diverse meteorological variables [19,22,29]. Besides, the training convergence time with some deep-learning methods, especially for abundant weather variable forecasts at large amount of weather stations, is remarkable long. It is well known that decision-tree-based methods, such as random forest methods, are quite effective to deal with big data, while the training convergence time can be decreased greatly [30,31]. The trees in the forest are definitely different and consequently provide an “expert collection” that performs better than any single tree. Among these, Catboost algorithm [32], which is an improved version of gradient-boosting decision tree (GBDT) and uses a special type of depth-first expansion called oblivious trees, has put in great performance at many data-mining applications [33–36]. In this paper, a hybrid model based on Catboost and wavelet packet denoising is proposed to predict multiple meteorological variables at multiple future steps. Some additional features are extracted and fused. Moreover, it will combine the advantages of the NWP and machine-learning method to improve forecasting accuracy, while the training convergence time can be largely decreased compared with some deep-learning methods. The rest of the paper is structured as follows. The problem statement and our method are discussed in Section 2, followed by the experiments and performance analysis. Lastly, we conclude with a brief summary in Section 4.

## 2. Materials and Methods

### 2.1. Problem Statement

In this work, historical meteorological observations from 10 weather stations for more than three years and RMAPS-based preliminary weather forecast data from NWP [22] are supplied. They can be defined as follows.

Historical meteorological observation datasets denoted as

$$O(t) = [o_1(t), o_2(t), \dots, o_{N_1}(t)] \in R^{N_1} \quad (1)$$

where the variable  $o_i(t)$  is one of  $N_1$  meteorological features, for  $t = 1, \dots, T_o$ . Here,  $N_1 = 9$ .

Preliminary weather forecast timeseries obtained by NWP as

$$F(t) = [f_1(t), f_2(t), \dots, f_{N_2}(t)] \in R^{N_2} \quad (2)$$

where the variable  $f_i(t)$  is one of  $N_2$  NWP features, for  $t = T_0 + 1, \dots, T_0 + T_f$ ,  $T_f$  is forecasting step number. In this work,  $N_2 = 29$ . Additionally, a forecast time period is required from 3:00 to 15:00 (UTC) of the next day, then  $T_f = 37$ .

Ground truth of target meteorological variables  $Y(t)$  and their predictions  $\tilde{Y}(t)$

$$Y(t) = [y_1(t), y_2(t), \dots, y_{N_3}(t)] \in R^{N_3} \tag{3}$$

where the variable  $y_i(t)$  is truth value of  $N_3$  meteorological variables, for  $t = T_0 + 1, \dots, T_0 + T_f$ . In this work, relative humidity at 2 m (rh2m), wind at 10 m (w10m), and temperature at 2 m (t2m) are target forecasting variables, and therefore,  $N_3 = 3$ .

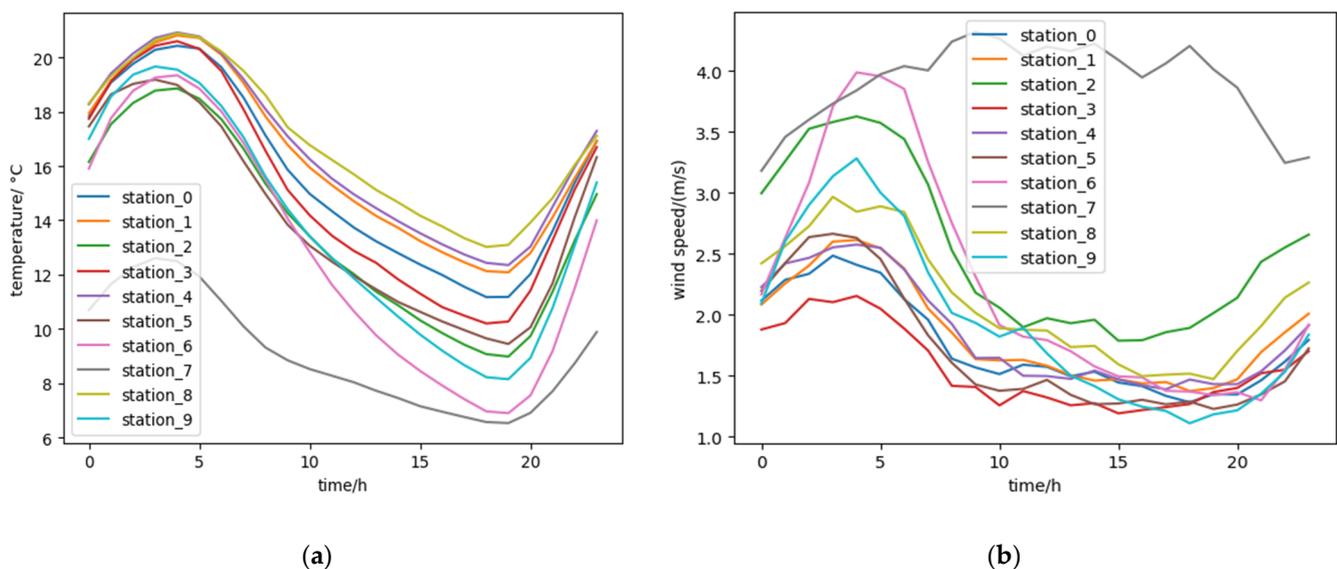
## 2.2. Data Processing

### 2.2.1. Missing Values

In this work, two kinds of missing values, including local missing (local non-continuous) and block missing, are included. For local missing, linear interpolation is employed. As for block missing, the data of those days are deleted directly, and therefore, 40-day data are deleted from a total of 1188 day dataset.

### 2.2.2. Additional Spatiotemporal Features

Figure 1 shows the historical statistic means of meteorological variation t2m and w10m. From this figure, compared with other weather stations, t2m means in station ID 7 has obvious differences and w10m means in station ID 7 and 6 follow different trends. Thus, station ID should be added as a category variable. In this work, the one-hot encoding method is employed to depict the spatial feature.



**Figure 1.** The 24 h means of meteorological variations from 1 March 2015 to 21 May 2018 for 10 stations: (a) t2m, (b) w10m.

As for the temporal features, the historical variation of target variable t2m is presented in Figure 2. From this figure, strong periodicity and seasonality can be observed, and the temporal features should be considered. If the time features such as month and hour are also set by the one-hot encoding method, a large feature space will be occupied. Moreover, the temporal continuity such as between December and January will be destroyed if the month feature (1–12) is directly mapped into the interval of 0–1. In this work, clock projection is utilized to extract the temporal features. Specifically speaking, the month feature and the hour feature are transformed to the order of 1–12 on the clock, which can be seen in Figure 3. The projected horizontal and vertical coordinate values (within 0–1) will

be deemed as new temporal features. Test results demonstrate that the prediction accuracy can be significantly improved by the additional spatiotemporal features.

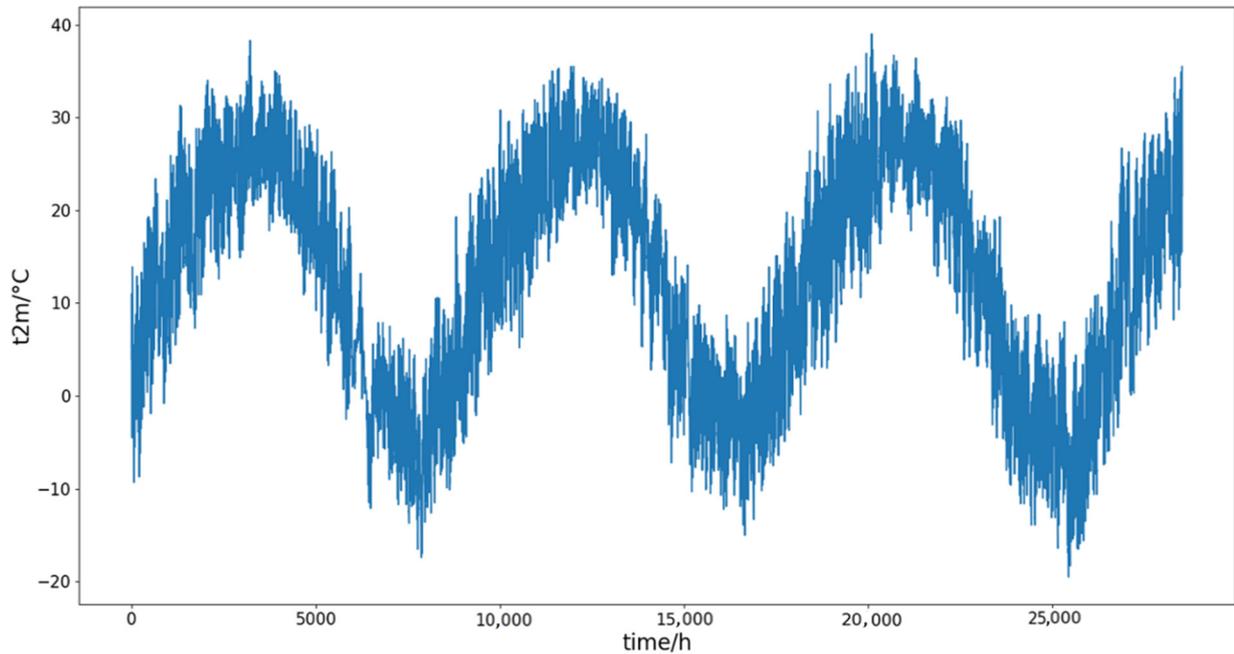


Figure 2. The variation of t2m from 1 March 2015 to 21 May 2018.

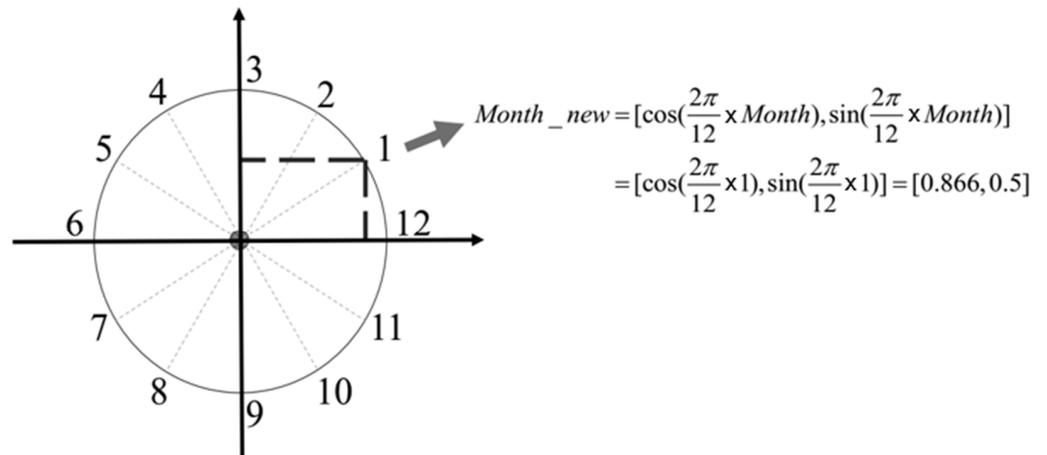


Figure 3. The clock projection method for temporal feature extraction.

### 2.2.3. Feature Selection

In this work, 38 meteorological features are provided for weather forecast in the historical observed dataset and NWP forecasting dataset. Considering the prediction accuracy and the training computational consumption, an ensemble selection method integrating three methods are employed for feature selection, including RFE, correlation matrix, and tree model. In this work, SVM-based recursive feature elimination (RFE) is adopted [37]. The correlation matrix is also called the correlation coefficient matrix, and the map is presented in Figure 4. Moreover, the Catboost model is also used for the ranking of feature importance. By the ensemble selection model, the less important features are screened out.

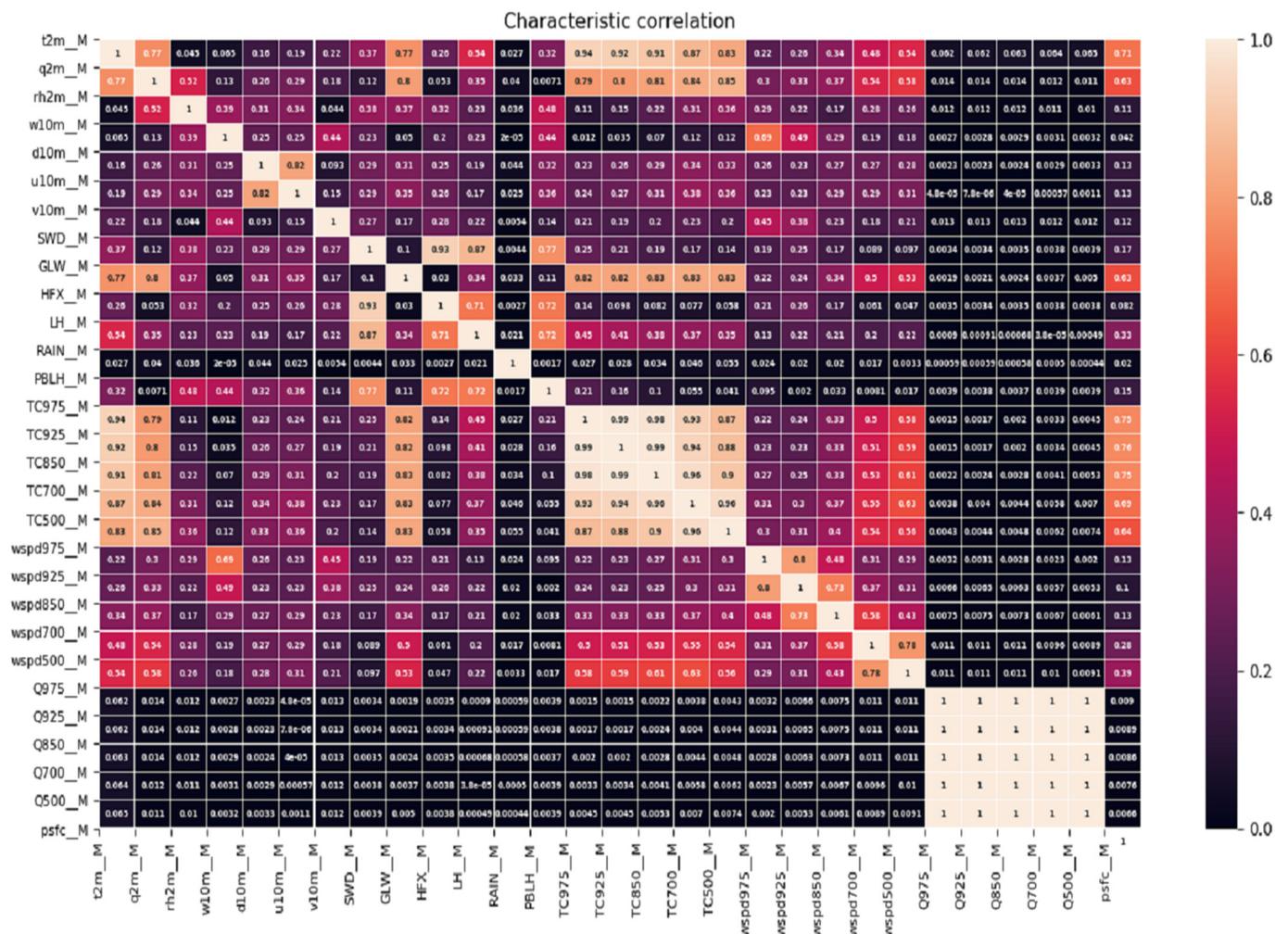


Figure 4. The correlation matrix heat map.

### 2.3. Model Architecture

The proposed short-term weather forecasting model is based on wavelet packet denoising and Catboost. The block diagram of the proposed forecasting model is presented in Figure 5.

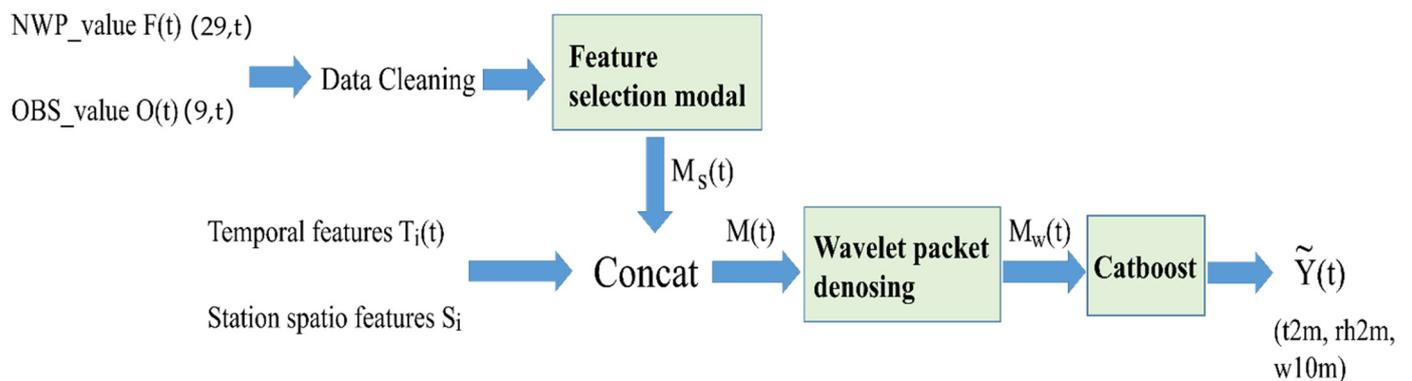


Figure 5. The block diagram of the proposed forecasting model.

In this model, the historical meteorological observation dataset  $O_{T_0} = [O(1), O(2), \dots, O(T_0)] \in R^{T_0 \times N_1}$  and NWP forecasting dataset  $F_{T_f} = [F(1), F(2), \dots, F(T_f)] \in R^{T_f \times N_2}$  will be firstly dealt

with above-mentioned data preprocessing, including data cleaning, spatiotemporal features addition, and feature selection. Then, the wavelet packet denoising method is proposed and achieved as a “pre-learning task”, which can improve the accuracy and shorten the convergence time. At last, the Catboost model is utilized for training and weather forecast.

### 2.3.1. Wavelet Packet Denoising Principle

Signal denoising, especially for sensor noises in historical meteorological observation datasets, is of great importance. This “pre-learning task” will be verified to effectively shorten the subsequent training convergence time and enhance forecast accuracy. For noise reduction in non-stationary signals in meteorological datasets, it is reasonably effective to remove the noises by the wavelet-based denoising methods due to their multi-resolution and multi-scale analysis property [38]. Wavelet transforms have been successfully used in many scientific fields such as image compression, image denoising, and signal processing, to name only a few. The window size in wavelet transform is fixed, while its shape is adjustable. Both the time and frequency window can be adjusted. Hence, it has a strong ability to extract local features of signals and detect singularity in signals [39]. The denoised signals can be obtained by employing the threshold function to filter the wavelet coefficients and conduct the wavelet reconstruction [40].

In wavelet-based denoising methods, the wavelet basis function and the threshold function have great influence on the sparsity of wavelet representation coefficients. In this work, Daubechies (dbN) basis function is adopted due to its fine regularity. It means that it is difficult to perceive the smooth error introduced by this wavelet as a sparse basis, and then, the denoised signal can be much smoother [40]. Moreover, the soft-threshold method is employed here. It has good continuity and makes a more smooth process to the wavelet coefficients.

For a signal  $f(t) \in L^2(\mathbb{R})$ , the wavelet decomposition coefficients can be obtained as

$$W_f(a, \tau) = \langle f(t), \psi_{a,\tau}(t) \rangle = \frac{1}{\sqrt{a}} \int f(t) \psi\left(\frac{t-a}{\tau}\right) dt \quad (4)$$

where  $\psi_{a,\tau}(t)$  is the mother wavelet; the parameter  $a$  indicates the scale index, and  $\tau$  represents the time shifting.

In order to extract local features of signals by more fine-grained details and enhance the time-frequency resolution, the wavelet packet denoising method is utilized in this paper. The recursive formula of wavelet packet transform is as follows.

$$W_{2n}(t) = \sum_k h_k W_n(2t - k) \quad (5)$$

$$W_{2n+1}(t) = \sum_k g_k W_n(2t - k) \quad (6)$$

The signal  $W_1(t)$  will go across the orthogonal filter combining the high-pass filter  $h_k$  and low-pass filter  $g_k$  [39] and wavelet packet decomposition can divide the signal into different frequency spans layer-by-layer. The width of the frequency span  $\Delta f$  can be obtained as

$$\Delta f = \frac{f_s}{2^{i+1}} \quad (7)$$

where  $f_s$  is the sampling frequency;  $i$  is the decomposition layer. The expected frequency span can be obtained when layer  $i$  is set appropriately. The original signal can be separated from noise and interference signals when every frequency is wide enough. The recursive reconstruct formula is

$$W_2(2t) = \sum_k h_{2k+1} W_{2n}(t - k) + \sum_k g_{2k+1} W_{2n+1}(t - k) \quad (8)$$

$$W_2(2t - 1) = \sum_k h_{2k} W_{2n}(t - k) + \sum_k g_{2k} W_{2n+1}(t - k) \tag{9}$$

It is beneficial that we can choose all the frequency bins or some parts of them and set the others (noise or random interruption) as zero. When the signal is decomposed into different frequency bins, it is easy to extract noise by the reconstruction. Figure 6 illustrates the three-layer wavelet packet decomposition.

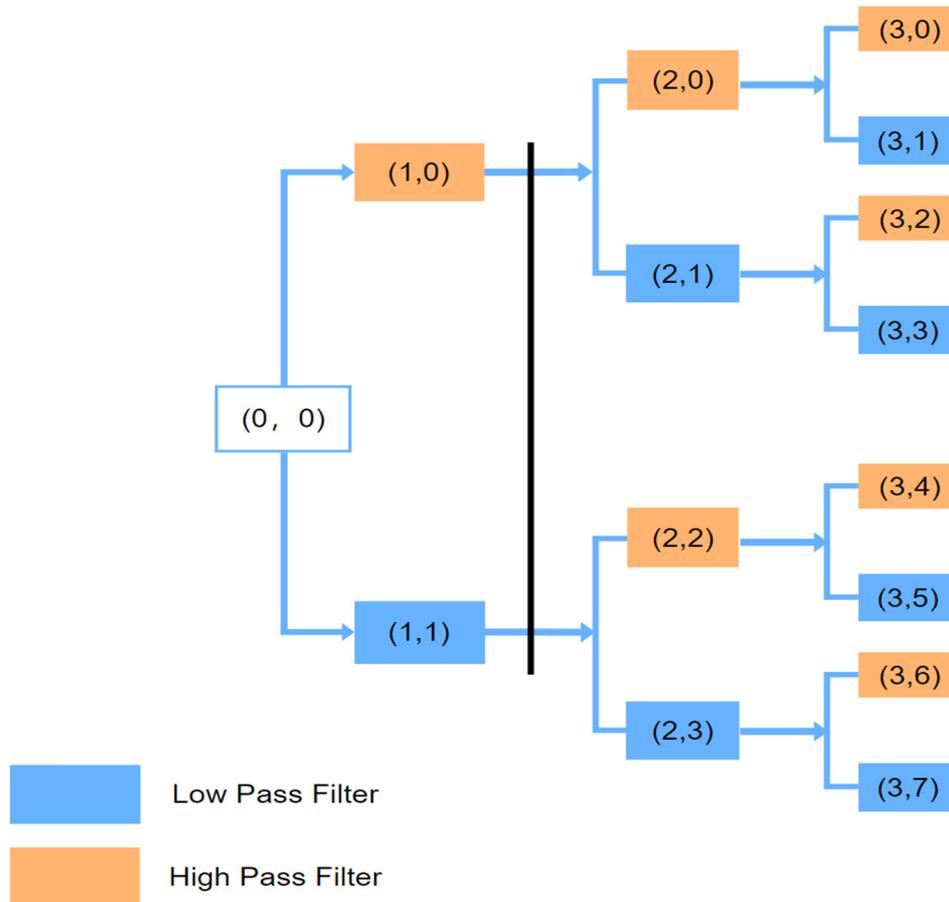


Figure 6. Three-layer wavelet packet decomposition.

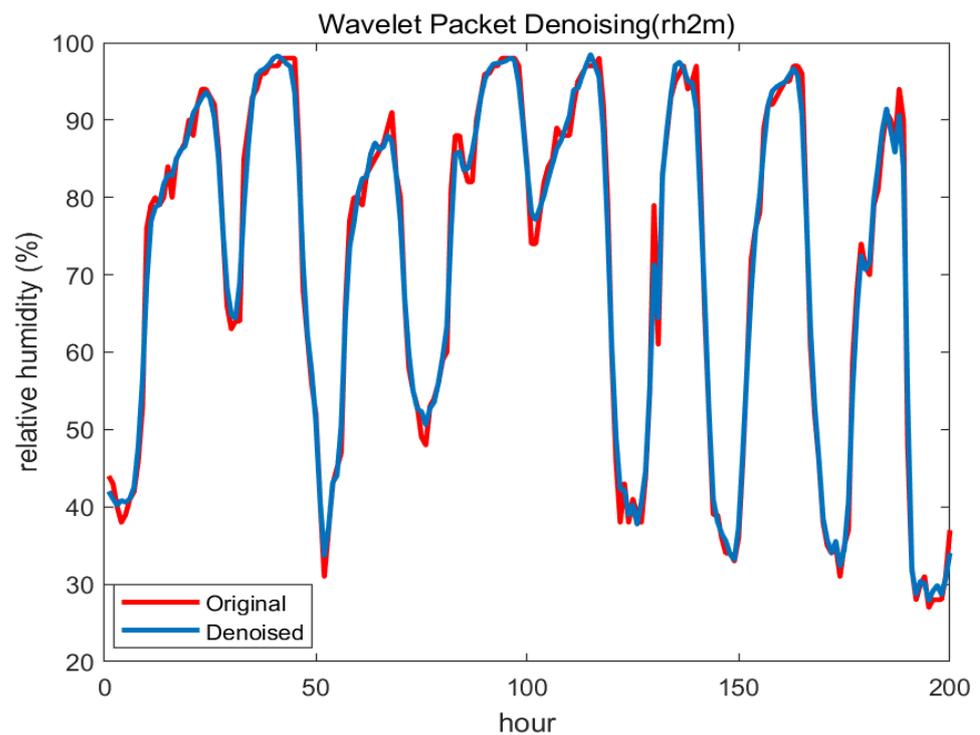
In addition, for the choice of the wavelet packet threshold, the three-layer wavelet packet and the soft-threshold is utilized. The threshold  $Th_{wp}$  is designed to be self-adaptive with the decomposition layers and takes advantage of the mean and standard variance of decomposition coefficients, that is

$$Th_{wp}(p) = \frac{1}{M} \sum_{j=1}^M C_{3,p}(j) + \sqrt{\frac{1}{M} \sum_{j=1}^M \left[ C_{3,p}(j) - \frac{1}{M} \sum_{j=1}^M C_{3,p}(j) \right]^2} \tag{10}$$

$$\tilde{C}_{3,p}(j) = \begin{cases} \text{sign}(C_{3,p}(j)) * (|C_{3,p}(j)| - Th_{wp}(p)), & |C_{3,p}(j)| \geq Th_{wp}(p) \\ 0 & |C_{3,p}(j)| < Th_{wp}(p) \end{cases} \tag{11}$$

where  $Th_{wp}(p)$  is the wavelet packet threshold, combining with the mean and variance.  $C_{3,p}(j)$ ,  $p = 1, 2, \dots, 8$  is the wavelet packet transform coefficients, and  $M$  is the coefficient length. When coefficients  $C_{3,p}(j)$  are greater than the threshold  $Th_{wp}(p)$ , set  $C_{3,p}(j)$  as zero; otherwise, set  $C_{3,p}(j)$  as  $\text{sign}(C_{3,p}(j)) * (|C_{3,p}(j)| - Th_{wp}(p))$ , which makes the  $\tilde{C}_{3,p}(j)$ .

The historical meteorological observation signals are denoised by wavelet packet transform. Figure 7 shows the original observed and denoised curves of 2 m relative humidity.



**Figure 7.** The original observed and denoised curves of 2 m relative humidity.

### 2.3.2. Learning Model: Categorical Boosting

Gradient boosting is an effective and powerful machine-learning technology for solving problems with complex dependencies, noisy data, and heterogeneous characteristics. It has a theoretical explanation on how iteration combines weak models through gradient descent in function space [32] and has demonstrated most advanced performance in a variety of practical tasks, such as ET0 estimation [30], global solar radiation prediction [33], and web searching [35]. Catboost, proposed by Yandex Company, is a novel gradient boosting algorithm [32], which makes many improvements to overcome the model overfitting and deal with parallelism. Thus, the layout can be completed in less time. Generally, high prediction accuracy is the key point; however, good stability and less computational workload should also be laid emphasis when employing machine-/deep-learning models. Some models are inherently unstable and will obtain fewer precision estimates when with new datasets [30].

Catboost can deal with categorical features well and is employed here as a learning model for short-term weather forecast. It has the following advantages [32]:

**Category features:** In order to reduce overfitting and utilize the whole dataset for training, an efficient strategy called target statistics (TS) with minimum information loss is employed in the Catboost. Specially, for the input example sets  $D = \{(x_k, y_k)\}_{k=1, \dots, n}$ , a plurality of random permutation is performed. Then, the average label values will be calculated for the sequence with the same category value. Finally, all classification features will be substituted with the following formula:

$$\tilde{x}_k^j = \frac{\sum_{j=1}^n [x_j^i = x_k^i] \cdot y_j + \beta \cdot P}{\sum_{j=1}^n [x_j^i = x_k^i] + \beta} \quad (12)$$

where the parameter  $\beta > 0$ , namely, the weight of the prior, can dampen the low frequency category noise.  $P$  is a prior value.  $y_k \in \mathbb{R}$  is the target, and  $x_k = (x_k^1, \dots, x_k^M)$  is a random vector of  $M$  features.

**Feature combinations:** A greedy way is utilized for Catboost when the tree constructs a new split. No combination is considered for the first split. However, for the subsequent splits, Catboost contains all the combination and classification features in the current tree of the dataset with all categorical features. Moreover, all splits selected in the tree are treated as categories with two values and are similarly utilized in the combination.

**Unbiased boosting:** In Catboost, the ordered boosting is developed by theoretical analysis to solve the gradient bias, which is inevitable when the traditional GBDT employs the TS method to convert categorical features into numerical values. Moreover, multiple permutations of the training data are employed to enhance the robustness. Different permutations will be utilized for training distinct modals, which can deal with the overfitting problem.

**Fast scorer:** Oblivious trees, which are balanced and less inclined to overfitting, are used as base predictors in Catboost. Moreover, in order to calculate the model predictions, each leaf index in the Catboost model evaluators is encoded as a binary vector, whose length is equal to the depth of the tree. Figure 8 shows the structure of the Catboost algorithm.

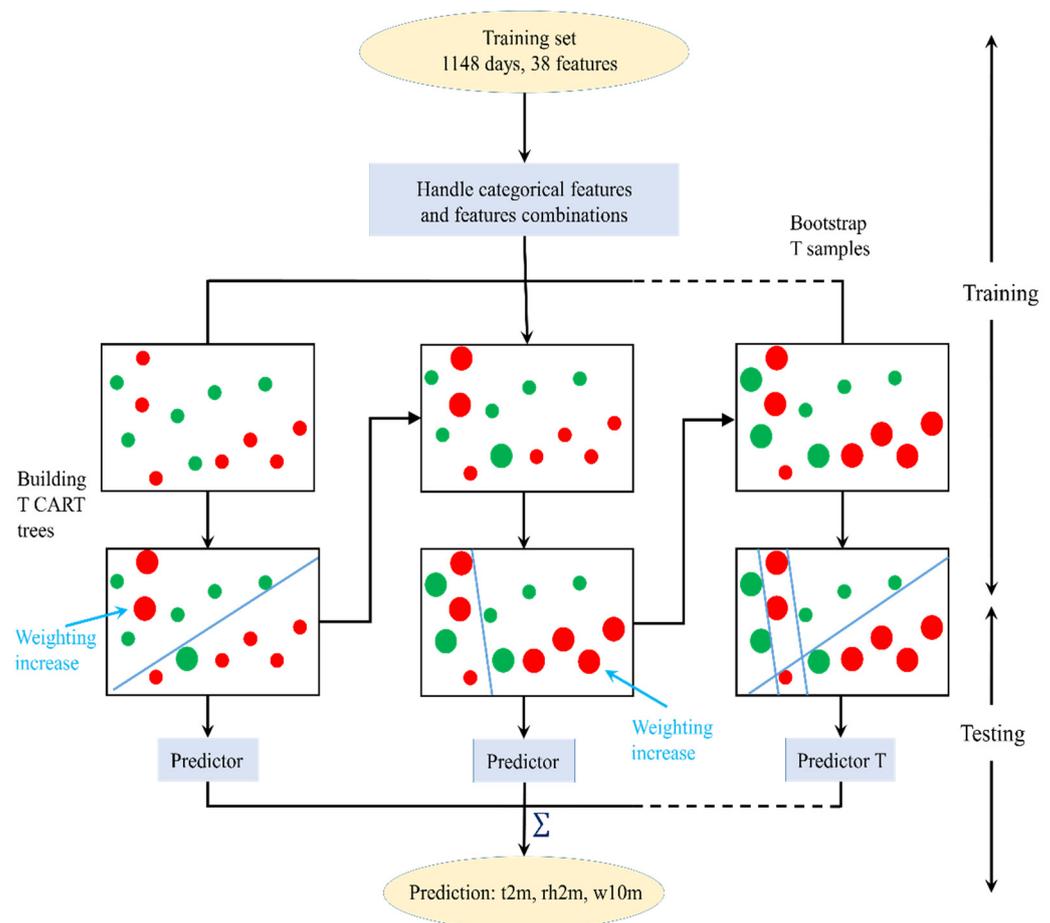


Figure 8. The Catboost algorithm structure.

### 3. Performance Analysis and Comparisons

In this section, the performance analysis and comparisons of the proposed weather forecasting model with other machine-learning and deep-learning methods are illustrated. Firstly, the evaluation metrics are shown.

### 3.1. Statistical Evaluation

The root mean squared error (*RMSE*) for three objective variables from 10 stations is calculated as daily evaluation.

$$RMSE_{i,ml} = \sqrt{\frac{\sum_{s=1}^{10} \sum_{t=T_o+1}^{T_o+T_f} (y_{i,s}(t) - y_{i,s}^{ml}(t))^2}{10 \cdot T_f}} \quad (13)$$

where  $y_{i,s}(t)$  and  $y_{i,s}^{ml}(t)$  are the ground truth and the forecasting value (by the proposed machine-learning method) of the objective variable  $i$  (here,  $i$  denotes t2m, rh2m, and w10m) of station  $s$  at time  $t$ , respectively. Similarly,  $RMSE_{i,NWP}$  can be obtained, which uses the predicted value  $F(t)$  by NWP method.

$$RMSE_{i,NWP} = \sqrt{\frac{\sum_{s=1}^{10} \sum_{t=T_o+1}^{T_o+T_f} (y_{i,s}(t) - F_{i,s}(t))^2}{10 \cdot T_f}} \quad (14)$$

Then, the associated skill score  $S_i$  is employed to compare the forecasting improvement with the classic NWP method.

$$S_i = \frac{RMSE_{i,NWP} - RMSE_{i,ml}}{RMSE_{i,NWP}} \quad (15)$$

$S_i > 0$  means that the proposed machine-learning method can obtain lower *RMSE* and better prediction accuracy than the NWP method. The higher  $S_i$  is, the better the forecasting performance by the proposed method is.

$$S_{day} = \frac{S_{t2m} + S_{rh2m} + S_{w10m}}{3} \quad (16)$$

$S_{day}$  is the average skill score of the three objective variables and is the ultimate prediction criterion for each day.

### 3.2. Baselines and Experimental Settings

In this work, the machine-learning method random forest and two deep-learning methods LSTM and Seq2Seq are implemented as baselines for comparisons. Seq2Seq method is well known in the field of natural language processing and is effective to solve timeseries prediction problems.

The tests were performed on a GPU server with GTX 1080Ti GPU, 11GB of video memory, and a Pytorch programming environment. As mentioned above, 38 meteorological features are provided from observed and NWP datasets. The three least correlated characteristics are removed, and the spatiotemporal features are added in the phase of data preprocessing. The set hyperparameter  $T_o = 28$  means that the previous 28 h observation data are selected to modal the recent meteorological dynamics. The data from 1148 days (during 1 March 2015 to 1 June 2018) are used for the training set, and data from 87 days (during 3 June 2018 to 29 August 2018) are for the validation set. Since the evaluation is based on online daily forecasting, the test day index is 1. The main parameters in the random forest method and Catboost method are the maximum depth and the number of trees, which are optimized by the grid search method.

The pseudocode are added to describe the whole methodology, as shown in Algorithm 1.

**Algorithm 1.** Combining WPD and Catboost Method for Weather Forecasting

**Input:** Historical Observation Datasets  $O(t)$ , NWP Datasets  $F(t)$ , Ground Truth  $Y(t)$

**Output:** Prediction  $\tilde{Y}(t)$  ( $t2m, rh2m, w10m$ )

- 1: Data Cleaning
- 2:  $M_s(t) \leftarrow$  Feature Selection:  $O(t), F(t)$
- 3:  $T_i(t) \leftarrow$  Temporal Features: *Month, Hour*
- 4:  $S_i(t) \leftarrow$  Spatio Features: *Station ID*
- 5:  $M(t) \leftarrow$  Concat  $M_s(t), T_i(t), S_i(t)$
- 6:  $M_w(t) \leftarrow$  Wavelet Packet Denosing:  $M(t)$
- 7:  $\tilde{Y}(t) \leftarrow$  Catboost:  $M_w(t)$
- 8: End

3.3. Performance Analysis

This work is based on an attended online competition dataset for daily weather forecasting. The forecasting daily evaluation score  $S_{day}$  and average score  $S_{avg}$  for continuous five competition days are presented in Table 1, which are based on incremental data released daily according to real-world forecasting processes. In this table, LSTM, Seq2Seq, and random forest (“RF” for brevity) are implemented for comparison. “ST” means that spatiotemporal features are added. “FS” denotes that feature selection is employed. “WPD” indicates that the wavelet packet denoising method is combined.

**Table 1.** The S performance of different methods on 5 days.

Method	$S_{day1}$	$S_{day2}$	$S_{day3}$	$S_{day4}$	$S_{day5}$	$S_{avg}$
LSTM + FS	0.0203	0.1112	0.0503	0.1303	0.3143	0.1253
LSTM + FS + ST	0.1000	0.2978	0.1821	0.2596	0.3488	0.2377
RF + FS + ST	0.1146	0.1962	0.3582	0.4270	0.3859	0.2964
RF + FS + ST + WPD	0.1821	0.2358	0.3990	0.4513	0.4553	0.3447
Seq2Seq + FS + ST + WPD	0.2587	0.3793	0.4606	0.5320	0.5102	0.4282
Catboost + ST	0.2933	0.3692	0.4610	0.5136	0.5406	0.4355
Catboost + ST + FS	0.2992	0.3921	0.4654	0.5175	0.5434	0.4435
Catboost + ST + FS + WPD	<b>0.3273</b>	<b>0.4088</b>	<b>0.4908</b>	<b>0.5447</b>	<b>0.5530</b>	<b>0.4649</b>
Catboost + ST + FS + WPD + noNWP	0.183	0.139	0.195	0.197	0.207	0.1842
Catboost + ST + FS + WPD + noOBS	0.016	0.314	0.343	0.396	0.401	0.294

First, comparing “LSTM + FS” with “LSTM + FS + ST”, it can be clearly observed that the prediction accuracy by adding spatiotemporal features has been remarkably improved. Similarly, it can be obtained that the feature selection is also effective by comparing “Catboost + ST + FS” with “Catboost + ST”. The less important features are screened out by the ensemble selection model, which can reduce redundant information. In order to verify the importance of wavelet packet denoising, the results of “RF + FS + ST” and “RF + FS + ST + WPD” as well as “Catboost + ST + FS” and “Catboost + ST + FS + WPD” could be taken into consideration and comparisons. Obviously, wavelet packet denoising can greatly promote the prediction scores. Wavelet packet denoising also takes effect as the “pre-learning” process and can effectively shorten the entire learning convergence time. For example, after wavelet packet denoising the data, Catboost only needs an approximately 40 epoch convergence time, which is rather short in contrast with the original one, about 100 epoch.

Second, considering the effect of information fusion, the effectiveness of fusing NWP forecasting can be validated by comparing “Catboost + ST + FS + WPD” with “Catboost + ST + FS + WPD + noNWP”, where the NWP forecasting is masked by zero values. Similarly, comparing “Catboost + ST + FS + WPD” with “Catboost + ST + FS + WPD + noOBS” demonstrates the advantage of modeling recent meteorological dynamics. Moreover, the  $S_{day} > 0$  in the “Catboost + ST + FS + WPD + noOBS” indicates that the machine-learning method can enhance the NWP alone performance. Meanwhile, the  $S_{day} > 0$  in

the “Catboost + ST + FS + WPD + noNWP” without NWP information illustrates that the proposed method has superiority in modeling meteorological data. Catboost successfully handles categorical features and uses a new schema for calculating leaf values when selecting the tree structure, which helps to reduce overfitting. These comparisons clearly exhibit that conducting information fusion is better, and modeling alone with OBS or NWP is not good enough. NWP forecasting contains important prior knowledge.

Moreover, the raw RMSE values for temperature, relative humidity, and wind speed with all the methods are presented in Tables 2–4, respectively. It is also clear that the proposed “Catboost + ST + FS +WPD” method can also achieve the best RMSE values.

**Table 2.** The RMSE performances of t2m with different methods.

Method	RMSE <sub>day1</sub>	RMSE <sub>day2</sub>	RMSE <sub>day3</sub>	RMSE <sub>day4</sub>	RMSE <sub>day5</sub>	RMSE <sub>avg</sub>
LSTM + FS	4.3167	4.5871	4.8221	5.1009	3.9671	4.5588
LSTM + FS + ST	2.4113	2.5097	2.6214	2.7719	2.6444	2.5917
RF + FS + ST	2.3121	2.3584	1.7002	1.4721	1.5651	1.8816
RF + FS + ST + WPD	2.1167	2.1871	1.6672	1.4323	1.5333	1.7873
Seq2Seq + FS + ST + WPD	1.7189	1.7228	1.2688	1.2554	1.4117	1.4755
Catboost + ST	1.1699	1.7337	1.2561	1.3019	1.1887	1.3300
Catboost + ST + FS	1.1642	1.6993	1.2343	1.2982	1.1832	1.3158
Catboost + ST + FS + WPD	<b>1.0596</b>	<b>1.6738</b>	<b>1.0977</b>	<b>1.1276</b>	<b>1.1253</b>	<b>1.2168</b>
Catboost + ST + FS + WPD + noNWP	2.1002	3.2278	2.5180	3.0988	4.4333	3.0756
Catboost + ST + FS + WPD + noOBS	4.9663	2.8534	2.1721	2.2564	1.5477	2.7592

**Table 3.** The RMSE performances of rh2m with different methods.

Method	RMSE <sub>day1</sub>	RMSE <sub>day2</sub>	RMSE <sub>day3</sub>	RMSE <sub>day4</sub>	RMSE <sub>day5</sub>	RMSE <sub>avg</sub>
LSTM + FS	13.2332	13.8361	13.3123	14.5241	12.1123	13.4036
LSTM + FS + ST	11.7341	8.4723	8.2121	11.0852	9.4516	9.7910
RF + FS + ST	12.1946	12.1678	6.5207	9.0303	8.6171	9.7061
RF + FS + ST + WPD	11.1348	11.4681	6.2913	8.9234	8.4588	9.2553
Seq2Seq + FS + ST + WPD	9.9271	8.5661	5.7026	6.8698	7.3950	7.6921
Catboost + ST	9.4694	8.6883	6.6771	7.3775	7.0220	7.8469
Catboost + ST + FS	9.3367	8.3433	6.6651	7.3379	7.0112	7.7388
Catboost + ST + FS + WPD	<b>9.2295</b>	<b>8.3106</b>	<b>6.4138</b>	<b>7.1386</b>	<b>6.9638</b>	<b>7.6113</b>
Catboost + ST + FS + WPD + noNWP	10.4456	12.3314	8.0278	11.4123	12.9973	11.0429
Catboost + ST + FS + WPD + noOBS	13.4584	9.8877	7.7854	10.5543	8.5771	10.0526

**Table 4.** The RMSE performances of w10m with different methods.

Method	RMSE <sub>day1</sub>	RMSE <sub>day2</sub>	RMSE <sub>day3</sub>	RMSE <sub>day4</sub>	RMSE <sub>day5</sub>	RMSE <sub>avg</sub>
LSTM + FS	2.1244	1.3122	1.4667	1.1883	1.2671	1.47174
LSTM + FS + ST	1.6000	0.9260	0.9670	0.8750	0.8140	1.0364
RF + FS + ST	1.5527	1.0406	0.7638	0.8242	0.8862	1.0135
RF + FS + ST + WPD	1.5488	1.0091	0.7505	0.8009	0.8801	0.99788
Seq2Seq + FS + ST + WPD	1.5357	0.9507	0.7392	0.7340	0.8982	0.97156
Catboost + ST	1.4889	0.9355	0.7311	0.7502	0.8663	0.9544
Catboost + ST + FS	1.4019	0.9031	0.7229	0.7442	0.8547	0.92536
Catboost + ST + FS + WPD	<b>1.3631</b>	<b>0.8883</b>	<b>0.7178</b>	<b>0.7365</b>	<b>0.8114</b>	<b>0.90342</b>
Catboost + ST + FS + WPD + noNWP	1.5439	1.0218	0.9512	0.9338	1.3304	1.15622
Catboost + ST + FS + WPD + noOBS	2.2883	0.9802	0.8856	0.9117	0.8834	1.18984

For cross-validation, the new training set (during 1 March 2015 to 15 August 2017), validation set (during 16 August 2017 to 30 April 2018), and test set (during 1 May 2018 to 29 August 2018, 120 days) are assigned again (ratio is about 7:2:1). The test results are shown in Table 5.  $\bar{S}_{day01-20}$  is the average  $S$  value of the first day to the twentieth day in the test set.  $\bar{S}_{avg}$  is the average  $S$  value of the first day to the one hundred and twentieth

day in the test set. From this table, the proposed “Catboost + ST + FS +WPD” method can also achieve the best average  $S$  values for the new cross-validation dataset.

**Table 5.** The statistical average  $S$  performances with different methods.

Method	$S_{day01-20}$	$S_{day21-40}$	$S_{day41-60}$	$S_{day61-80}$	$S_{day81-100}$	$S_{day101-120}$	$S_{avg}$
LSTM + FS	0.0992	0.2273	0.1372	0.1391	0.1335	0.1763	0.1521
LSTM + FS + ST	0.3109	0.2755	0.2859	0.2679	0.1677	0.2056	0.2522
RF + FS + ST	0.3771	0.3902	0.3561	0.2997	0.1984	0.2559	0.3129
RF + FS + ST + WPD	0.4358	0.4376	0.3997	0.3208	0.2110	0.3278	0.3555
Seq2Seq + FS + ST + WPD	0.5119	0.5005	0.4267	0.3329	0.2655	0.3792	0.4028
Catboost + ST	0.5134	0.5137	0.4215	0.3517	0.3287	0.4116	0.4234
Catboost + ST + FS	0.5264	0.5528	0.4293	0.3623	0.3319	0.4479	0.4418
Catboost + ST + FS + WPD	<b>0.5672</b>	<b>0.5837</b>	<b>0.4370</b>	<b>0.3878</b>	<b>0.3536</b>	<b>0.4732</b>	<b>0.4671</b>
Catboost + ST + FS + WPD + noNWP	0.449	0.341	0.269	0.188	0.109	0.165	0.2535
Catboost + ST + FS + WPD + noOBS	0.075	0.465	0.373	0.295	0.237	0.317	0.2937

For demonstrating the effect of varying important hyper-parameters, the proposed method with different “iterations” (number of trees) is tested, and the results are shown in Table 6. “Iterations” is an important hyper-parameter in the Catboost method.  $Catboost_{500}$  means the proposed method with “iterations” = 500 is employed. It is verified that hyper-parameters are also important for prediction performance.

**Table 6.** The  $S$  performances of the proposed method with different “iterations”.

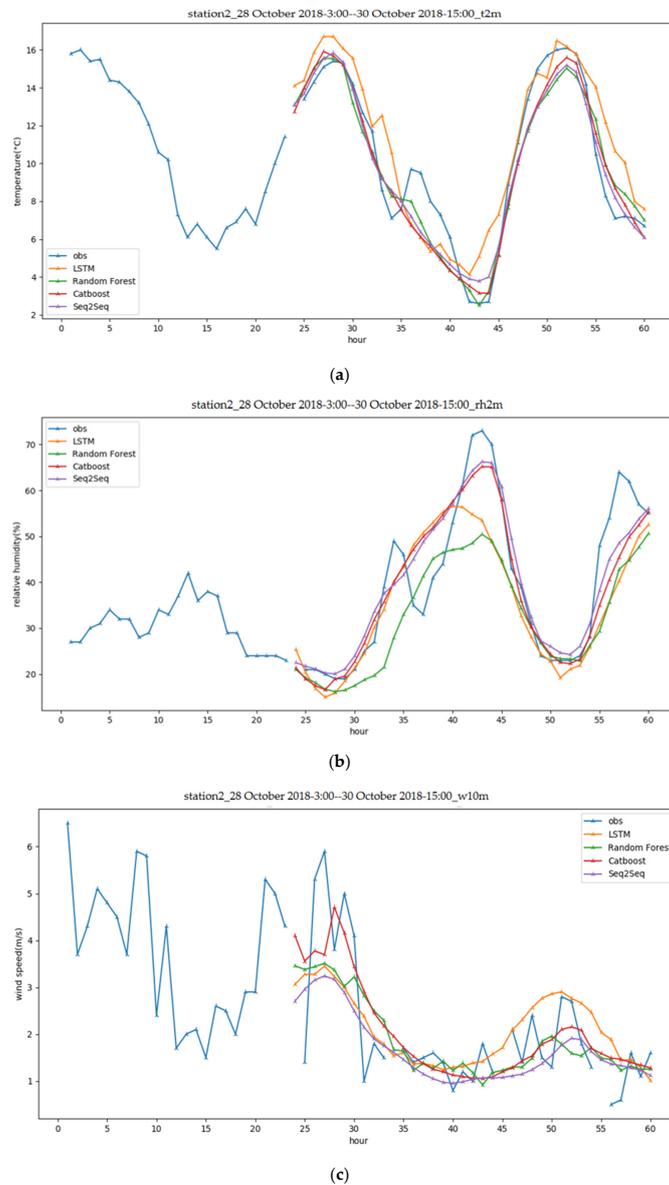
Method	$S_{day1}$	$S_{day2}$	$S_{day3}$	$S_{day4}$	$S_{day5}$	$S_{avg}$
$Catboost_{500}$	0.3109	0.3913	0.4755	0.5226	0.5337	0.4396
$Catboost_{1000}$	0.3188	0.3991	0.4837	0.5316	0.5411	0.4549
$Catboost_{2000}$	0.3229	0.4025	0.4869	0.5365	0.5469	0.4591
$Catboost_{3000}$	<b>0.3273</b>	<b>0.4088</b>	<b>0.4908</b>	<b>0.5447</b>	<b>0.553</b>	<b>0.4649</b>
$Catboost_{4000}$	0.3246	0.4053	0.4889	0.5414	0.5492	0.4619
$Catboost_{5000}$	0.3238	0.4031	0.488	0.5377	0.5471	0.4599

In addition, apart from the accuracy, it is also vital for model construction to minimize the complexity of models [30]. To illustrate the superiority of the proposed algorithm in convergence computing efficiency, a convergence computational time cost experiment under three levels of datasets was implemented. Table 7 presents the computational time costs of the proposed method and LSTM, RF, and Seq2seq models for different amounts of stations. We test three datasets with different levels of data size, including Level 1 (data from one station), Level 2 (data from five stations), and Level 3 (data from 10 stations). It can be seen that the average computing time was algorithm specific, and the average time consumed by the proposed algorithm was much less than that of the LSTM, RF, and the Seq2seq algorithm, especially for the two deep leaning methods. A more noteworthy observation is that the computing time costs of LSTM, RF, and the Seq2seq will increase rapidly as the training data size increases, but it does not change too much for the proposed method. For the Level 2 and Level 3 datasets (just five and 10 stations), the computational costs of LSTM and Seq2Seq were 128.7–13,200.8 and 320.1–77,421.1 times the cost of the proposed method, which possesses a tremendous time cost advantage when predicting the meteorological elements of hundreds of weather stations in practice. The tree-based algorithms are generally competent to build decision trees in parallel, which would more or less decrease the computing time [30]. Therefore, the proposed method conducts apparent optimization in time complexity, especially when the size of the input meteorological dataset is large.

**Table 7.** Computational costs of the four algorithms at different levels of input data size.

Method	Level 1 Dataset (1 Station)	Level 2 Dataset (5 Station)	Level 3 Dataset (10 Station)
RF	39 s	222 s	724 s
LSTM	1791 s	8754 s	37,774 s
Seq2Seq	87,291 s	897,654 s	9,135,680 s
The proposed method	37 s	68 s	118 s

Furthermore, Figure 9a–c illustrates a forecasting instance 2 m temperature, 2 m relative humidity, and 10 m wind speed curves at one station on a competition day. The horizontal coordinate values indicate 37 h, and the vertical coordinate values represent the corresponding unit and the value range of each variable. In each sub-figure, the left blue line is the observed meteorological value during the previous 28 h, the right blue line is the ground truth, the brown line is the Seq2Seq prediction, and the red line is the proposed prediction. It is also clear that the proposed method can achieve higher prediction accuracy.



**Figure 9.** A test sample at one station is chosen to visualize the forecasting of 3 target variables in the future 37 h. (a) 2 m temperature. (b) 2 m relative humidity. (c) 10 m wind speed.

#### 4. Conclusions

Based on the historical meteorological observation datasets and numerical weather prediction (NWP) provided by Beijing weather stations, a short-term weather forecasting model based on wavelet packet denoising and Catboost is put forwarded. Correlation heat map and tree method are combined for feature selection. Moreover, specific spatiotemporal features are extracted for the periodicity and spatial differences of weather features. Then wavelet packet denoising is utilized to provide more effective denoised features, which processes a part of the “learning” task in advance. Test results have demonstrated that, compared with the conventional LSTM, random forest, and Seq2Seq methods, the proposed method incorporating wavelet packet denoising with Catboost can significantly shorten the convergence time of the learning model and decrease the computational cost, as well as notably improve the prediction accuracy.

In this paper, some hyper-parameters (such as the soft-threshold  $Th_{wp}$  in the wavelet packet denoising and the number of trees “iterations” in the Catboost) are important for prediction performance and need to be carefully selected. Future works include making the effort to automatically tune hyper-parameters (e.g., self-adaptive soft threshold) or try some ensemble methods. Moreover, the effect of historical observation sequence length and the fusion strategy of NWP forecast data (e.g., attention mechanism), as well as integrating some deep-learning model (e.g., transformer structure) in the prediction model can be further explored to enhance prediction performance.

**Author Contributions:** Conceptualization, D.N.; data curation, Z.Z., H.C. and T.Z.; funding acquisition, Z.Z. and X.C.; methodology, D.N. and L.D.; project administration, D.N. and Z.Z.; software, D.N., T.Z. and L.D.; validation, H.C. and T.Z.; visualization, H.C. and T.Z.; writing—original draft, L.D.; writing—review and editing, D.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (No. 2018YFC1506905), Natural Science Foundation of Jiangsu Province of China (No. BK20202006), Zhishan Youth Scholar Program of Southeast University, the Key R&D Program of Jiangsu Province (No. BE2019052, BE2017076).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank Yichao Cao and Junhao Huang for helpful discussions and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Gneiting, T.; Raftery, A.E. Weather forecasting with ensemble methods. *Science* **2005**, *310*, 248–249. [[CrossRef](#)] [[PubMed](#)]
2. Jones, N. Machine learning tapped to improve climate forecasts. *Nature* **2017**, *548*, 379–380. [[CrossRef](#)] [[PubMed](#)]
3. Ham, Y.G.; Kim, J.H.; Luo, J.J. Deep learning for multi-year ENSO forecasts. *Nature* **2019**, *573*, 568–572. [[CrossRef](#)] [[PubMed](#)]
4. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, M. Skillful Precipitation Nowcasting using Deep Generative Models of Radar. *Nature* **2021**, *597*, 672–677. [[CrossRef](#)]
5. Marchuk, G. *Numerical Methods in Weather Prediction*; Elsevier: Amsterdam, The Netherlands, 2012.
6. Tolstykh, M.A.; Frolov, A.V. Some current problems in numerical weather prediction. *Izv. Atmos. Ocean. Phys.* **2005**, *41*, 285–295.
7. Juanzhen, S.; Ming, X.; James, W.W.; Zawadzki, I.; Ballard, S.P.; Onvlee-Hooimeyer, J.; Pinto, J. Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 409–426.
8. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the 28th International Conference on Neural Information Processing Systems (NeurIPS), Montreal, QC, Canada, 7–10 December 2015; pp. 802–810.
9. McGovern, A.; Elmore, K.L.; Gagne, D.J.; Haupt, S.E.; Karstens, C.D.; Lagerquist, R.; Williams, J.K. Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 2073–2090. [[CrossRef](#)]

10. Basha, C.Z.; Bhavana, N.; Bhavya, P.; Sowmya, V. Rainfall prediction using machine learning & deep learning techniques. In Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2–4 July 2020; pp. 92–97.
11. Imam Cholissodin, S. Prediction of rainfall using improved deep learning with particle swarm optimization. *Telkommika* **2020**, *18*, 2498–2504. [[CrossRef](#)]
12. Khan, M.I.; Maity, R. Hybrid deep learning approach for multi-step-ahead daily rainfall prediction using GCM simulations. *IEEE Access* **2020**, *8*, 52774–52784. [[CrossRef](#)]
13. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag.* **2009**, *4*, 24–38. [[CrossRef](#)]
14. Ling, C.; Xu, L. Comparison between ARIMA and ANN models used in short-term wind speed forecasting. In Proceedings of the Power and Energy Engineering Conference (APPEEC), Wuhan, China, 25–28 March 2011; pp. 1–4.
15. Cyril, V.; Marc, M.; Christophe, P.; Marie-Laure, N. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy* **2012**, *39*, 341–355.
16. Chen, N.; Qian, Z.; Nabney, I.T.; Meng, X. Short-term wind power forecasting using gaussian processes. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China, 3–9 August 2013.
17. Grover, A.; Kapoor, A.; Horvitz, E. A deep hybrid model for weather forecasting. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 379–386.
18. Hernández, E.; Sanchez-Anguix, V.; Julian, V.; Palanca, J.; Duque, N. Rainfall prediction: A deep learning approach. In Proceedings of the International Conference on Hybrid Artificial Intelligence Systems, Seville, Spain, 18–20 April 2016; pp. 151–162.
19. Li, Y.; Lang, J.; Ji, L.; Zhong, J.; Wang, Z.; Guo, Y.; He, S. Weather forecasting using ensemble of spatial-temporal attention network and multi-layer perceptron. *Asia-Pac. J. Atmos. Sci.* **2021**, *57*, 533–546. [[CrossRef](#)]
20. Salman, A. Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Comput. Sci.* **2018**, *135*, 89–98.
21. Fu, Q.; Niu, D.; Zang, Z.; Hao, H.; Li, D. Multi-stations' weather prediction based on hybrid model using 1D CNN and Bi-LSTM. In Proceedings of the 2019 Chinese Control Conference, Guangzhou, China, 27–30 July 2019; pp. 3771–3775.
22. Wang, B.; Lu, J.; Yan, Z.; Luo, H.; Li, T.; Zheng, Y.; Zhang, G. Deep uncertainty quantification: A machine learning approach Bashafor weather forecasting. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2087–2095.
23. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. In Proceedings of the 30th International Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–7 December 2017; pp. 5617–5627.
24. Wang, Y.; Zhang, J.; Zhu, H. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 9154–9162.
25. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–7 December 2017; pp. 879–888.
26. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Philip, S.Y. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the International Conference on Machine Learning (PMLR), Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132.
27. Wang, C.; Wang, P.; Wang, P.; Xue, B.; Wang, D. Using Conditional Generative Adversarial 3D Convolutional Neural Network for Precise Radar Extrapolation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5735–5749. [[CrossRef](#)]
28. Liu, H.B.; Lee, I. MPL-GAN: Toward Realistic Meteorological Predictive Learning Using Conditional GAN. *IEEE Access* **2020**, *8*, 93179–93186. [[CrossRef](#)]
29. Niu, D.; Huang, J.; Zang, Z.; Xu, L.; Che, H.; Tang, Y. Two-Stage Spatiotemporal Context Refinement Network for Precipitation Nowcasting. *Remote Sens.* **2021**, *13*, 4285. [[CrossRef](#)]
30. Huang, G.; Wu, L.; Ma, X. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* **2019**, *574*, 1029–1041. [[CrossRef](#)]
31. Karimi, S.; Shiri, J.; Marti, P. Supplanting missing climatic inputs in classical and random forest models for estimating reference evapotranspiration in humid coastal areas of Iran. *Comput. Electron. Agric.* **2020**, *176*, 105633. [[CrossRef](#)]
32. Prokhorenkova, L.; Gusev, G.; Vorobev, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–8 December 2018; pp. 6638–6648.
33. Wu, L.; Huang, G.; Fan, J. Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Convers. Manag.* **2019**, *183*, 280–295. [[CrossRef](#)]
34. Chia, M.Y.; Huang, Y.F.; Koo, C.H. Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—A review. *Agronomy* **2020**, *10*, 101. [[CrossRef](#)]
35. Kang, P.; Lin, Z.; Teng, S. Catboost-based framework with additional user information for social media popularity prediction. In Proceedings of the the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2677–2681.

36. März, A. CatBoostLSS-An extension of CatBoost to probabilistic forecasting. *arXiv* **2020**, arXiv:2001.02121.
37. Huang, X.; Zhang, L.; Wang, B. Feature clustering based support vector machine recursive feature elimination for gene selection. *Appl. Intell.* **2018**, *48*, 594–607. [[CrossRef](#)]
38. Qian, H.; Ma, J.C. Research on fiber optic gyro signal de-noising based on wavelet packet soft-threshold. *J. Syst. Eng. Electron.* **2009**, *20*, 607–612.
39. Wu, Y.J.; Gao, G.J.; Cui, C. Improved wavelet denoising by non-convex sparse regularization under double wavelet domains. *IEEE Access* **2019**, *7*, 30659–30671. [[CrossRef](#)]
40. Pan, Y.; Zhang, L.M.; Wu, X.G.; Zhang, K.N.; Skibniewski, M.J. Structural health monitoring and assessment using wavelet packet energy spectrum. *Saf. Sci.* **2019**, *120*, 652–665. [[CrossRef](#)]