sustainability

MDPI

Article

A Quantitative Modeling and Prediction Method for Sustained Rainfall-$PM_{2.5}$ Removal Modes on a Micro-Temporal Scale

Tingchen Wu [1,2,3,4], Xiao Xie [5,*], Bing Xue [5] and Tao Liu [1,2,3,4]

1 Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China; giswtchen@163.com (T.W.); ltaochina@foxmail.com (T.L.)
2 National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, China
3 Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China
4 Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China
5 Key Lab for Environmental Computation and Sustainability of Liaoning Province, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China; xuebing@iae.ac.cn
* Correspondence: xiexiao@iae.ac.cn

Abstract: $PM_{2.5}$ is unanimously considered to be an important indicator of air quality. Sustained rainfall is a kind of typical but complex rainfall process in southern China with an uncertain duration and intervals. During sustained rainfall, the variation of $PM_{2.5}$ concentrations in hour-level time series is diverse and complex. However, existing analytical methods mainly examine overall removals at the annual/monthly time scale, missing a quantitative analysis mode that applies micro-scale time data to describe the removal phenomenon. In order to further achieve air quality prediction and prevention in the short term, it is necessary to analyze its micro-temporal removal effect for atmospheric environment quality forecasting. This paper proposed a quantitative modeling and prediction method for sustained rainfall-$PM_{2.5}$ removal modes on a micro-temporal scale. Firstly, a set of quantitative modes for sustained rainfall-$PM_{2.5}$ removal mode in a micro-temporal scale were constructed. Then, a mode-constrained prediction of the sustained rainfall-$PM_{2.5}$ removal effect using the factorization machines (FM) was proposed to predict the future sustained rainfall removal effect. Moreover, the historical observation data of Nanjing city at an hourly scale from 2016 to January 2020 were used for mode modeling. Meanwhile, the whole 2020 year observation data were used for the sustained rainfall-$PM_{2.5}$ removal phenomenon prediction. The experiment shows the reasonableness and effectiveness of the proposed method.

Keywords: sustained rainfall; $PM_{2.5}$ removal mode; micro-temporal scale; quantitative modeling; mode prediction

Citation: Wu, T.; Xie, X.; Xue, B.; Liu, T. A Quantitative Modeling and Prediction Method for Sustained Rainfall-$PM_{2.5}$ Removal Modes on a Micro-Temporal Scale. *Sustainability* **2021**, *13*, 11022. https://doi.org/10.3390/su131911022

Academic Editor: Constantinos Cartalis

Received: 27 August 2021
Accepted: 1 October 2021
Published: 5 October 2021

1. Introduction

Air pollution is currently a major environmental challenge for both developed and developing countries worldwide, with increasing industrialization, growing urbanization and energy consumption posing a serious threat to public health [1,2]. According to a report from the World Health Organization, $PM_{2.5}$ is unanimously considered to be an important indicator of air quality [3,4]. The rainfall processes are typically regarded as strong drivers to remove $PM_{2.5}$ so as to improve the air quality.

However, in the existing available research on an observation data analysis, there are obvious inconsistency between rainfall and the $PM_{2.5}$ removal effect, especially for sustained rainfall, a kind of typical but complex rainfall process in southern China with an uncertain duration and intervals. For example, Kao et al. [5] found that in a rainforest environment, summers with high precipitation are negatively correlated with $PM_{2.5}$ levels; Preethi et al. [6] proposed that the effect of simulated Indian monsoon rainfall removal

depends to a large extent on climatic wind speeds; Neal et al. [7] pointed out that there has been no systematic improvement in air quality in mid Wales for 17 years in the face of increased rainfall. In addition to the complexity of the sustained rainfall itself, the variations in regional environmental conditions are possibly another important reason. Therefore, studying the relationship between sustained rainfall and the $PM_{2.5}$ removal effect becomes a multi-factor related scientific issue.

When expressing the effect of rainfall on the removal of $PM_{2.5}$ concentrations in the air, the observed time series is usually the most intuitive expression [8,9]. The observed value will be considered as a sign of the presence of the rainfall process [10]. However, sustained rainfall, which is characterized by two or more sustained rainfall events within a given period of time, is intermittent, slow to change and uncertain in its length of formation, and its complexity needs to be fully considered [11,12]. During sustained rainfall, the variation of $PM_{2.5}$ concentrations in hour-level time series is diverse and complex. However, existing analytical methods mainly examine overall removals at the annual/monthly time scale and mainly use the correlation analysis (CA) between rainfall amount and $PM_{2.5}$ concentration based on macroscopic monitoring data, which refers to the rainfall time series and $PM_{2.5}$ time series sampled on an annual or monthly temporal scale [13,14]. For example, using the data of a macro-temporal scale, Shaibu et al. [15] confirmed that monthly $PM_{2.5}$ concentrations in the Niger Delta region of Nigeria show a significant positive correlation with monthly rainfall. Similarly, through a study of annual datasets from five air quality monitoring stations in Bahrain from 2006 to 2012, Jassim et al. [16] indicated little correlation between rainfall and $PM_{2.5}$, leading to a year-on-year increase in $PM_{2.5}$ concentrations. They are all missing a quantitative analysis mode that applies micro-scale time data to describe the removal phenomenon. In order to further achieve air quality prediction and prevention in the short term, it is necessary to analyze its micro-temporal removal effect for atmospheric environment quality forecasting.

In addition, due to the sustained rainfall and the atmospheric pollution particulate matter itself being complex, the large-scale temporal analysis lacks guidance for specific sustained rainfall-$PM_{2.5}$ removal processes [17]. On this basis, a part of the study proposes to extract the historical single rainfall process using hourly observation data and adopt a predetermined calculation model to quantify the removal effect of the rainfall process: Chhavi et al. [18] analyzed the wet removal effect by calculating the $PM_{2.5}$ concentration difference before and after rainfall, including the positive and negative removal; Kapwata et al. [19], based on the intensity of rainfall, delineate the rainfall classes, counting the percentage of positive removal to summarize the removal effect according to the influence factors, such as rainfall duration and rainfall volume, which have certain guiding significance. The above methods do not take into account the changes in the effects produced by complex sustained rainfall processes at different stages, including effects such as hygroscopic growth and the secondary transformation of gaseous pollutants which cause $PM_{2.5}$ concentrations to rise or rebound [20–22]; at the same time, they lack universal law exploration, having difficulties in serving the scientific prediction and early warning of air quality systems and active prevention [23,24].

In this paper, we propose a quantitative modeling and prediction method for sustained rainfall-$PM_{2.5}$ removal modes on a micro-temporal scale. The detailed contributions are as follows:

- A novel micro-scale analytical framework for quantitatively elucidating the mechanism of $PM_{2.5}$ removal by sustained rainfall was proposed. Compared with the yearly, monthly and daily time scales, the hourly scale is a more suitable form of information for decision making; therefore, the framework would more clearly express the complex characteristics of sustained rainfall than the analysis methods of large-scale data. The innovative hourly scale data analysis in this paper is more useful for practical applications in predicting and assessing air quality.
- A set of quantitative $PM_{2.5}$ removal modes based on a micro-analysis are proposed. The modes would highlight the specific and high-level patterns of the removal effect

of sustained rainfall at the micro-scale than the traditional micro-scale data analysis methods. During sustained rainfall, the variation of $PM_{2.5}$ concentrations in an hourly time series is diverse and complex. The analysis of hourly scales reveals new characteristic modes that are different from the traditional large scale. These "declining, rebounding, or rising" modes not only allow the analysis of historical data from different regions, but also allow the prediction of $PM_{2.5}$ removal at hourly intervals using future hourly rainfall, which can help the relevant systems and departments to make timely decisions on air pollution control.

This paper is organized as follows. The study area and data on analytical framework are viewed in Section 2. Section 3 introduces the quantitative definition of the sustained rainfall-$PM_{2.5}$ removal mode on a micro-temporal scale, then presents the rainfall-$PM_{2.5}$ removal phenomenon predicting algorithm based on the quantitative model and Section 4 discusses the experimental results. Finally, the discussions are presented in Section 5.

2. Study Area and Data

2.1. Study Area

The experimental area of this paper is located in Nanjing, which is in the middle of the lower reaches of the Yangtze River and has a subtropical monsoon climate with cold and dry winters, high temperatures and rainy summers and a relatively large rainfall variability, making it one of the regions in China with more droughts and floods [25,26]. Some studies have shown that (i) due to the influence of strong convective weather and global warming, there is a significant increase in the amount and duration of rainfall in Nanjing, which has a different degree of the purifying effect on air pollutants [27]; (ii) at the same time, according to the results of a pollutant source analysis in Nanjing, $PM_{2.5}$ water-soluble ion concentrations have diurnal and seasonal differences, making the average mass concentration of water-soluble ions higher during the day than at night, and the flushing effect of spring and summer rainfall is generally lower than that of autumn and winter [28,29]; (iii) according to the statistics, the annual average humidity in Nanjing is nearly 80%, and some rainfall in this environment may lead to hygroscopic growth of particulate matter, resulting in higher pollutant concentrations [30].

2.2. Dataset

The experimental data in this paper were obtained from hourly $PM_{2.5}$ concentrations and meteorological observations released by the China General Environmental Monitoring Station and the National Meteorological Information Centre for the period from January 2016 to December 2020. Based on the geographical location of the monitoring stations in Nanjing and the completeness of the data, the $PM_{2.5}$ data observed at the CCM (Caochangmen) site (32.0572 °N, 118.7486 °E) and the rainfall data observed at the PK (Pukou) site (32.177 °N, 118.706 °E) were used for the experiment.

Figure 1 demonstrates the phenomenon of $PM_{2.5}$ concentration removal by rainfall observed on a macro-temporal scale (yearly and monthly) to a micro-temporal scale (hourly): from the annual scale, the average annual $PM_{2.5}$ concentration in Nanjing decreases year by year from 2016 to 2020, while the pollutant concentration is relatively lower in years with a higher average annual rainfall; from the monthly scale, the monthly rainfall shows an overall stable trend of first increasing and then decreasing, and the average monthly $PM_{2.5}$ concentration in the period of concentrated rainfall are significantly lower than those during periods of low rainfall, indicating that rainfall has a significant effect on the removal of air pollutants; on an hourly scale, the hour-by-hour $PM_{2.5}$ concentration changes during a single rainfall process, thus rainfall has an obvious effect on pollutant removal while there are situations that cause concentrations to increase, and it is not possible to obtain the mode of the effect of regional rainfall on $PM_{2.5}$ from a particular rainfall process. In summary, the mechanism of $PM_{2.5}$ removal by sustained rainfall in Nanjing is complex, and it is difficult to satisfy the study of the mechanism of $PM_{2.5}$ removal by rainfall with the

macro time statistics and the time series of a single rainfall process, so the micro time-series effect model proposed in this paper is applied for a deeper analysis.

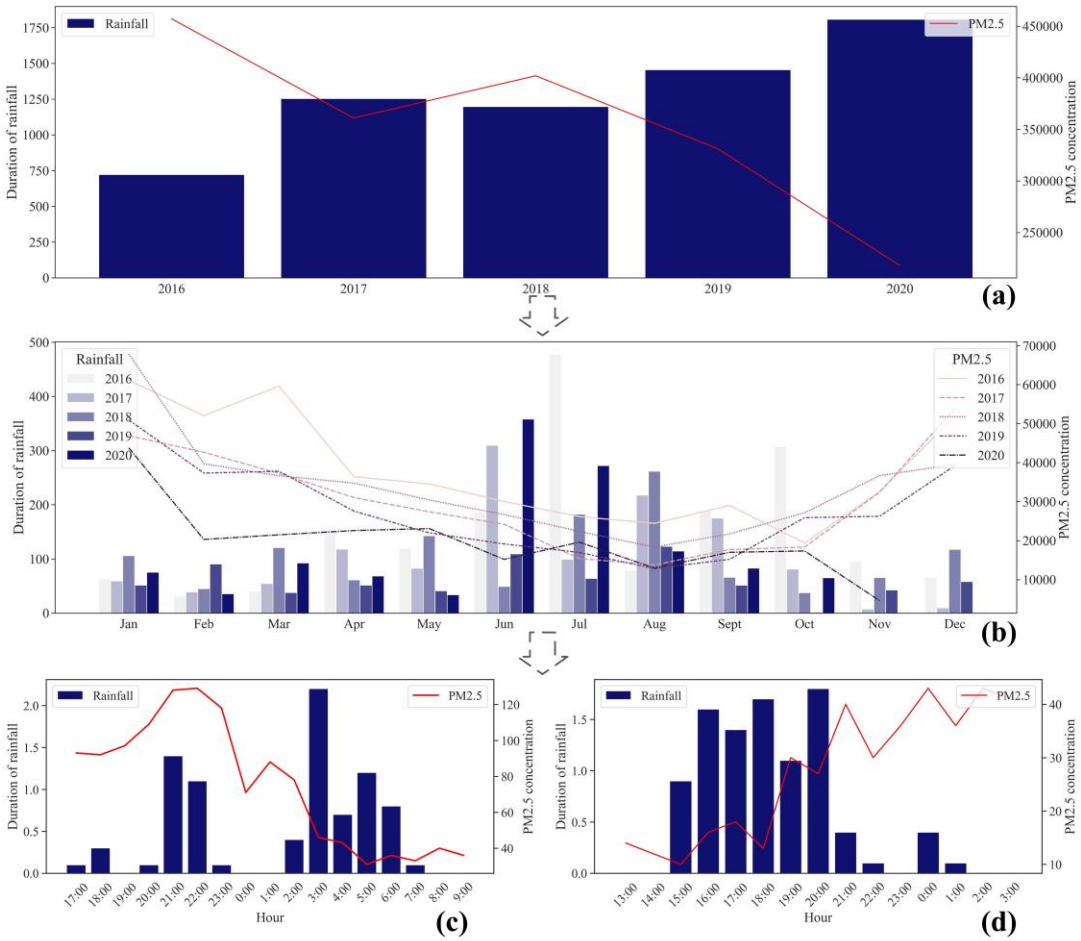

Figure 1. Macro/micro-temporal scale of sustained rainfall-$PM_{2.5}$ observation data series and removal phenomenon in Nanjing from 2016 to 2020: (**a**) Yearly temporal-level; (**b**) Monthly temporal-level; (**c**,**d**) Hourly temporal-level.

3. Methods

3.1. A Quantitative Modeling for Sustained Rainfall-$PM_{2.5}$ Removal Mode in Micro-Temporal Scale

3.1.1. Overview

In order to accurately describe the study object and the study boundary from the observation dataset, this section first defines a microscopic time-series fragment model of the sustained rainfall process, on the basis of which the time-series process of the removal effect is structured, further modeling the concomitant factors affecting the role of rainfall in $PM_{2.5}$ removal, and finally establishes an evaluation mode to quantitatively describe the removal effect.

3.1.2. Micro-Temporal Modeling of Sustained Rainfall Process

The model needs to accurately identify a complete sustained rainfall process in micro-scale observations due to the large variation in time duration and the existence of uncertain intervals. The sustained rainfall process is characterized by intermittent multi-fragmentation on hourly observation data series.

Time-series segments (TS) are mathematical frameworks for describing and modeling event sequences in the time domain, and are commonly used to construct interpretable time-series analysis models, such as trend prediction and anomaly detection [31,32]. In this

paper, the sustained rainfall time-series segment (SRS) is defined at the micro-temporal scale rainfall: $\{R_t : t = hour\}$. SRS visualizes the variation of rainfall in the time dimension of a sustained rainfall process, as shown in Figure 2, and is defined as follows:

$$SRS = \left\{ [R_{t'}, \ldots, R_{t''}] \middle| t', \Delta t, t'' \right\} \quad (1)$$

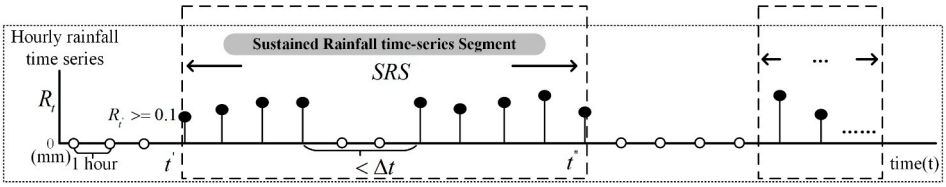

Figure 2. Schematic diagram of a micro-temporal model of the sustained rainfall process.

The quantitative measurement steps are:

Step 1: Using the hourly rainfall values as a benchmark, the moment of the first occurrence of 0.1 mm and above rainfall is taken as the starting point t' of the time-series fragment.
Step 2: Since rainfall and its resulting effects will remain in space for a certain period of time, a threshold value Δt is set to indicate the intermittent duration of the sustained rainfall process, i.e., the sequence before and after when rainfall is zero does not exceed Δt is regarded as the same time-series fragment.
Step 3: Satisfy the above conditions and the last occurrence of rainfall greater than zero is the end point t'', and obtain a complete SRS.

The micro-temporal model of the sustained rainfall process provides a quantitative basis for determining the baseline interval for the clearance effect analysis.

3.1.3. Sustained Rainfall Removal Concomitant Factor Modeling

Sustained rainfall processes are often concomitant by changes in the accompanying meteorological factors such as temperature, humidity, wind speed and direction, which will affect the removal effect to varying degrees [33]. Therefore, in this paper, we consider the effect and intensity of the existing influence factors [34–37], and balance the availability of their own observational data and the predictability of future trends to establish a set of accompanying influence factors F, including the (i) direct factor (F_D) to describe rainfall characteristics, and the (ii) indirect factor (F_I) to describe environmental characteristics. The factors and impact effects are shown in Table 1.

On this basis, the removal effects time-series segment (RES) is defined in conjunction with the $PM_{2.5}$ time series. The RES refers to the process and subsequent effects of rainfall on $PM_{2.5}$ at the scale of the mechanism of removal, while avoiding the chance of $PM_{2.5}$ concentration values before rainfall anomalies by expanding the series values for t and t' forward and backward for n hours, respectively (Figure 3):

Table 1. Table of qualitative descriptions of concomitant factor.

	Class	Factor	Label	Impact Effects
F	F_D	Rainfall Total	P	High correlation with air pollutant concentrations, which can directly influence the removal effect
		Rainfall Duration	D	
	F_I	Temperature	$\overline{T}$	When the temperature near the ground is high, atmospheric convection is intensified, which tends to reduce $PM_{2.5}$ concentrations, and conversely $PM_{2.5}$ is not easily dispersed
		Humidity	$\overline{H}$	Changes in $PM_{2.5}$ are closely related to the moisture content of the air, with "hygroscopic increase" occurring due to the adsorption of particulate matter concentrations
		Wind Power	W	Stronger winds also facilitate the dilution and uplift of pollutants
		Initial $PM_{2.5}$	C_1	The effect of removal is influenced by the magnitude of $PM_{2.5}$ concentrations before rainfall, and has little effect on particulate concentrations when air quality is good
		Seasonal	S	The removal effect is mostly higher at night than during the day; the total positive removal of sustained rainfall will be slightly higher in autumn than in other seasons
		Day and night	K	

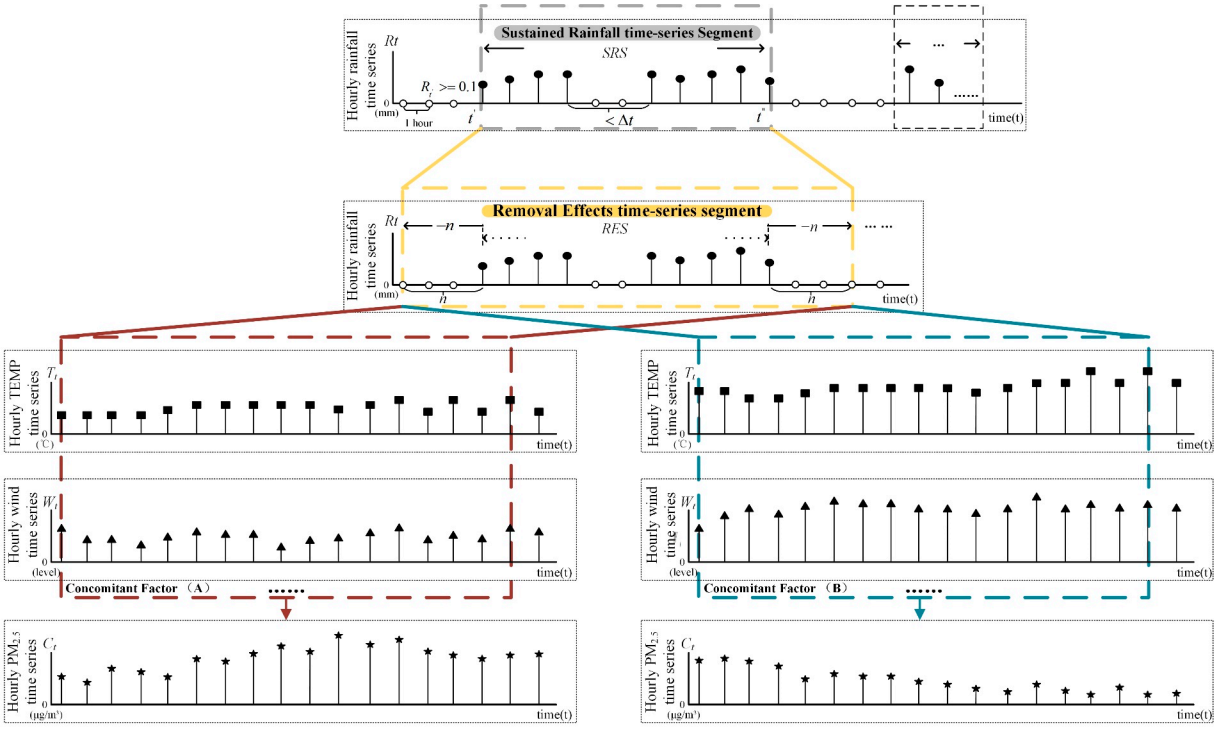

Figure 3. Schematic diagram of a micro-temporal model of the removal effects process and concomitant factor.

The concomitant factor model $\{F, RES\}$ for sustained rainfall removal identifies the information dimensions and effect intervals for the quantitative modeling.

3.1.4. Quantitative Evaluation Modeling of Removal Effects

The effect evaluation refers to the evaluation of the change in the effect produced by a geographical process considering the influence of relevant factors, and usually uses quantitative indicators to measure and analyze its dynamic change characteristics in spatial

and temporal modes [38,39]. The quantitative evaluation model of removal effects is proposed based on the above-mentioned models (SRS, RES):

$$\mathrm{M} = \left\{ M_{i\in(\mathrm{T,P,D,U,R,L})} \middle| \langle RP, RF, RS\rangle, (TSR, TSE) \right\} \quad (2)$$

(1) Quantitative Measurement of Removal Effect Indicators

Based on the model (SRS, RES), the following measurable evaluation indicators (RP, RF, RS) that quantitatively characterize the temporal variability in the removal processes are defined below:

(i) Rate of process RP, the ratio of the difference between the very small value of $PM_{2.5}$ concentration change when it exists and the pre-start $PM_{2.5}$ concentration.

(ii) Rate of final RF, the ratio of the difference between the initial $PM_{2.5}$ concentration before the start and the concentration after the end, the magnitude of which allows a quantitative evaluation of the intensity of removal.

(iii) Rate of rebound RR, the ratio of the difference between the minimum value and the ending concentration for the entire rainfall process.

(2) Modeling of Removal Mode based on Effect Indicators

The (SRS, RES) and (RP, RF, RR) were used to modeling the removal process of sustained rainfall and the indicators were combined with zero boundaries, corresponding to the rising and falling changes in $PM_{2.5}$ concentrations in Figure 4, and combined with the ensemble theory to classify the effect patterns in Table 2.

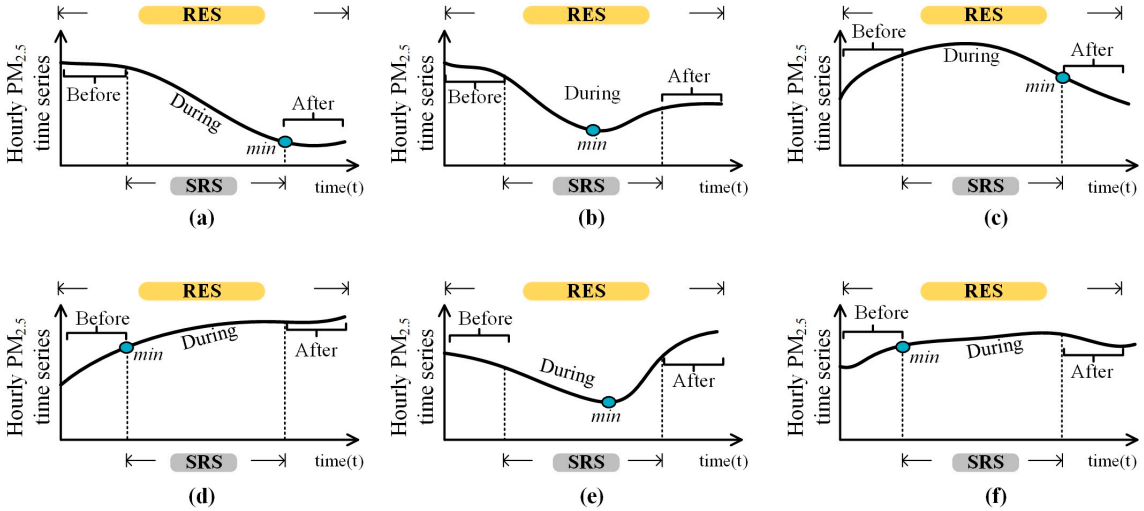

Figure 4. Schematic diagram of the sustained rainfall-$PM_{2.5}$ removal modes. (**a**) Totally Removal Mode; (**b**) Partly Removal Mode; (**c**) Delayed Removal Mode; (**d**) Totally Ascent Mode; (**e**) Rebounding Ascend Mode; (**f**) Lasting Ascent Mode.

Table 2. Table of sustained rainfall-$PM_{2.5}$ removal modes, quantitative classification and qualitative description.

Removal Mode M_i	Effect Indicators			Trends in $PM_{2.5}$ Concentrations		
	RP	*RS*	*RR*	During	After	Min
M_T	>0	<0	>0	Continued decline	Decline	Non-existent
M_P	>0	>0	>0	Decline, rebound	Decline	Existent
M_D	<0	<0	>0	Continued rise	Decline	Existent
M_A	<0	>0	<0	Continued rise	Rise	Non-existent
M_R	>0	>0	<0	Decline, rebound	Rise	Existent
M_L	<0	<0	<0	Continued rise	Rise	Existent

The removal mode, classified according to the above rules, covers all of the combinations of indicators and corresponds to realistic removal phenomena, providing a category a priori information for predicting future sustained rainfall.

3.2. The Mode Predicting of Sustained Rainfall-$PM_{2.5}$ Removal Effect Using the Quantitative Model

3.2.1. Overview

In this section, based on the quantitative model of sustained rainfall-$PM_{2.5}$ removal mode, the future rainfall is further fitted using a multiple regression to fit the removal modes, the core steps are as follows (Figure 5): Step 1, using the time-series window according to the constitute time-series sample of removal modes with high-dimensional characteristics; Step 2, extracting the principal components and component features of the pattern classification features through the principal component analysis method for the concomitant factor intensity; Step 3, establishing factorization classification regression machines for removal modes according to the factor intensity, which can predict the removal phenomenon based on the future meteorological and current air quality information.

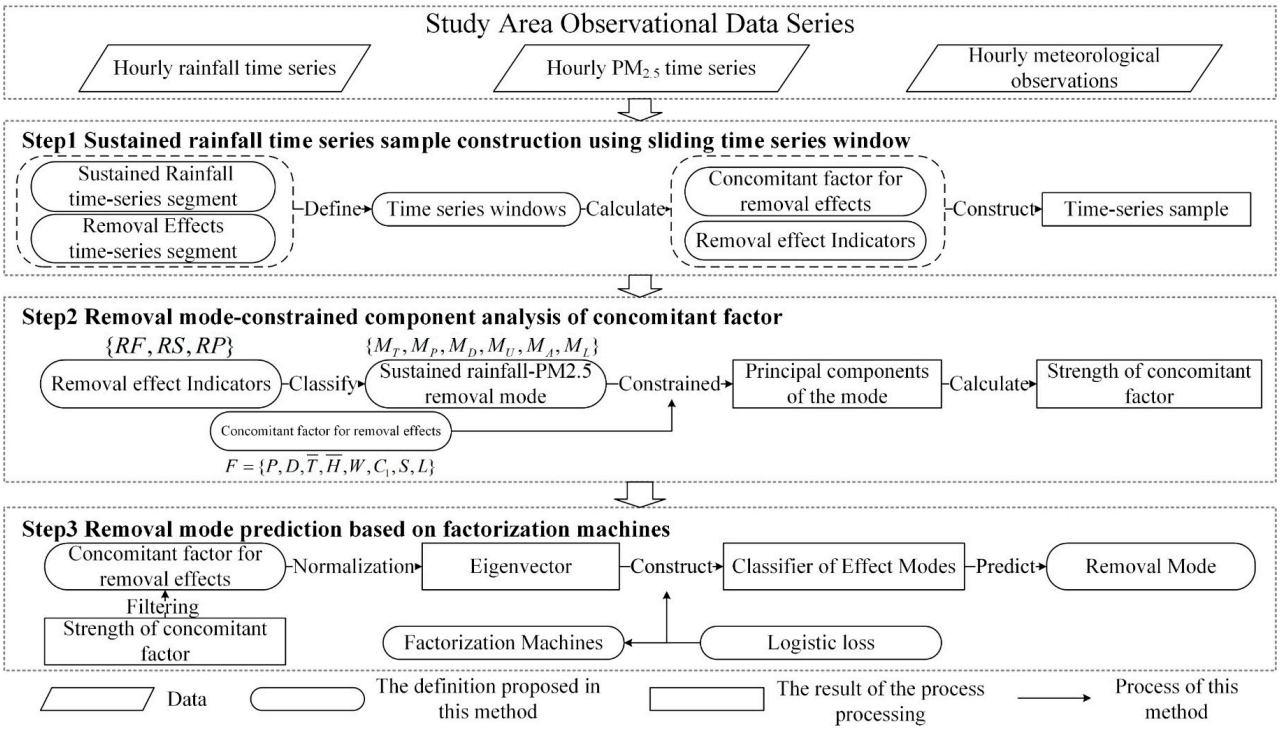

Figure 5. Flowchart for the prediction of sustained rainfall-$PM_{2.5}$ removal effect using the quantitative model.

3.2.2. Sustained Rainfall Time-Series Sample Construction Using Sliding Time-Series Window

Firstly, according to the definitions of the SRS and RES proposed in Section 3.1, a small-scale sliding window is applied to process the hourly rainfall time series R_t, hourly $PM_{2.5}$ time series C_t and hourly meteorological observations, setting up to capture all of the sustained rainfall processes in the historical data [40].

Secondly, the statistics of the corresponding direct and indirect factors are calculated within the extracted time-series range (SRS, RES) according to the concomitant factor model F proposed in this paper, as Table 3.

Table 3. Table of quantitative calculation of concomitant factor.

	Class	Label	Time-Series Range	Quantitative Calculation	
F	F_D	P	SRS	$P = \sum_{t=t'}^{t''} R_t$	(3)
		D		$D = \sum_{t=t'}^{t''} t$	(4)
	F_I	$\overline{T}$		$\overline{T} = \frac{\sum_{t=t'}^{t''} T_t}{t''-t'}$	(5)
		$\overline{H}$		$\overline{H} = \frac{\sum_{t=t'}^{t''} H_t}{t''-t'}$	(6)
		W		$W \in \{0, 1, \ldots, 17\}$	(7)
		C_1	RES	$C_1 = \frac{\sum_{t=(t'-n)}^{t'} C_t}{n}$	(8)
		S	SRS	$S \in \{Spr, Sum, Fal, Win\}$	(9)
		K		$K \in \{Am, Pm\}$	(10)

The quantitative evaluation indicators (RP, RF, RR) of the extracted sustained rainfall processes were calculated based on the $PM_{2.5}$ concentration series C_t according to the evaluation model of removal effects described in this paper.

$$\begin{cases} RP = \frac{(C_1 - C_{min})}{C_1} \\ RR = \frac{(C_2 - C_{min})}{C_2} \\ RF = \frac{(C_1 - C_2)}{C_1} \end{cases} \tag{11}$$

$$C_2 = \frac{\sum_{t=t'}^{t''+n} C_t}{n} \tag{12}$$

$$C_{min} = min(C_{t'}, \ldots, C_{t''}) \tag{13}$$

where C_1 is the initial $PM_{2.5}$ concentration during the sustained rainfall process, C_2 is the $PM_{2.5}$ concentration at the end of the sustained rainfall process and C_{min} is the minimum value for the whole sustained rainfall process.

Finally, sustained rainfall time-series samples $\vec{V}$ were constructed in the form of multi-dimensional feature vectors, as follows:

$$\vec{V} = \left(P, D, \overline{T,H}, W, C_1, S, K, [RF, RS, RP]\right) \tag{14}$$

It provides a computable basis for the accurate extraction of analytical reference intervals from hourly observations.

3.2.3. Removal Mode-Constrained Component Analysis of Concomitant Factor

According to the removal mode division rules in Table 2, the $\vec{V}$ were classified to obtain six types of labeled samples corresponding to M_i, as shown in the following equation. Due to the different degrees of influence of the removal factors on effect indicators in different effect modes, such as the change of effective removal rate magnitude with increasing rainfall duration in complete removal mode, a multi-factor analysis is needed to explore its change pattern.

$$\vec{V} = \left(P, D, \overline{T,H}, W, C_1, S, K \middle| M_i\right) \tag{15}$$

The principal component analysis is a statistical method that reduces the information of multi-dimensional variables to a few characteristic components by a linear transformation and reflects as much information of the original variables as possible [41]. In this section, the main components $\{F_1, F_2, \ldots, F_n\}$ of the sample features under each mode will be extracted separately, and the weights of each effect factor will be calculated by the

eigenvalues λ of each principal component and the variance contribution ratio $E(\lambda)$, and the relative strength S of the clearance factors under different modes will be evaluated quantitatively and used to screen the effect factors involved in the construction of the mode classifier.

$$S_{F_i} = \frac{A_{F_i}}{\sqrt{\lambda_{F_i}}} \times \frac{E(\lambda_{F_i})}{\sum_k^n E(\lambda_{F_k})} \tag{16}$$

A_{F_i} in the above equation is the loading of the initial factor i on the principal component.

The removal mode-constrained component analysis of concomitant factor enables the extraction of effective and stable effect factors that map mode characteristics, reducing the redundancy and sparsity of the factor information and improving the accuracy of the clearance effect prediction.

3.2.4. Removal Mode Prediction Based on Factorization Machines

Factorization machines (FM) are a classification method with a good learning ability for sparse data, solving the sparse information generated by the one-hot coding of category features, while taking into account the two-two correlation features between the features to meet the correlation and between the environmental factors in the rainfall process [42].

We use the factors filtered by the factor strengths calculated in Section 3.2.3, normalized by features (scaling all factors to between −1 and 1) with label coding to form the feature vector $\overrightarrow{X_i}$, which is used to build the factor decomposer $\hat{y}$ of the time-series effects model.

$$\begin{cases} \hat{y} = w_0 + \sum\limits_{i=1}^{n} w_i + \overrightarrow{X_i} + \sum\limits_{i=1}^{n-1} \sum\limits_{j=i+1}^{n} v_i, v_j \overrightarrow{X_i} \overrightarrow{X_j} \\ w_{i,j} = v_i v_j \end{cases} \tag{17}$$

Assuming that the sustained rainfall time-series sample has n features, v_i and v_j in the above equation are the implicit vectors of the feature matrix decomposition, and $w_{i,j}$ is the interrelationship between the two features i, j.

Further construction of a classifier oriented to the removal mode: the loss function loss Equation (18) is designed using the logistic regression theory, the $\hat{y}$ is mapped into different classes by the step function sigmoid and the logistic loss is used as the criterion for optimization.

$$\begin{cases} loss(\hat{y}, y) = \sum\limits_{i=1}^{m} -ln\sigma(\hat{y}^{(i)}, y^{(i)}) \\ \sigma(x) = \frac{1}{1-e^{-x}} \end{cases} \tag{18}$$

We can use the factor classifier constructed from the mode information to achieve a prediction of the removal effect that will result from a sustained rainfall process based on the rainfall forecast information.

4. Experimental and Analysis Section

This paper uses the historical observation data of Nanjing from January 2016 to January 2020 as $Data_{train}$ for the model analysis and uses the data from January 2020 to December of the same year $Data_{test}$ to verify the feasibility and effectiveness of the method of this paper.

4.1. Construct the Sustained Rainfall Time-Series Sample

Firstly, all of the datasets were processed in a uniform manner.

A sliding window was designed according to Section 3.2.2 to extract the sustained rainfall process, with all of the rainfall durations greater than 1 h, where the general residence time of particulate matter in the air and a window threshold of Δt was set to 3 h, rainfall with an interruption of no more than 3 h was considered as a complete sustained process (starting moment t' and ending moment t''), with a total of 427 sustained rainfalls collected.

While calculating the removal in each concomitant factors for each process, including the rainfall total, the rainfall duration and initial $PM_{2.5}$ concentration are shown according to Table 3, with seasonal and diurnal values used for labeling; further combining Equation (3) to calculate the removal effect indicators, and using these to classify the above rainfall samples into six modes according to the removal mode classification rules in Table 2, the result is shown in Table 4.

Table 4. A sample table of the sustained rainfall time series in Nanjing.

$\vec{V}$	$[t'-n, t''+n]$	P	D	$\overline{T}$	$\overline{H}$	W	C_1	S	K	$[RF, RS, RP]$	**M**
$\vec{V_1}$	2016/01/04/18:00-2016/01/05/07:00	18.0	12	8.21	4.91	4	276.33	4	2	[0.96, 0.31, 0.95]	M_P
$\vec{V_2}$	2016/01/10/21:00-2016/01/11/08:00	7.0	12	6.04	3.67	3	150.67	4	2	[0.19, −0.03, 0.22]	M_T
… …	… …						… …				… …
$\vec{V_{426}}$	2020/10/15/16:00-2016/10/16/17:00	17.6	24	14.02	2.31	2	15.33	3	2	[0.93, 0.83, 0.61]	M_P
$\vec{V_{427}}$	2020/10/21/05:00-2020/10/21/10:00	3.8	6	17.38	4.75	2	38.33	3	0	[0.08, −0.02, 0.11]	M_T

The findings show 85 times of M_T, 191 times of M_P, M_D 80 times, M_A 85 times, M_R 12 times and M_L 14 times. By plotting the numerical distribution of concomitant factors for different modes (Figure 6), the removal mechanism of sustained rainfall in Nanjing was explored.

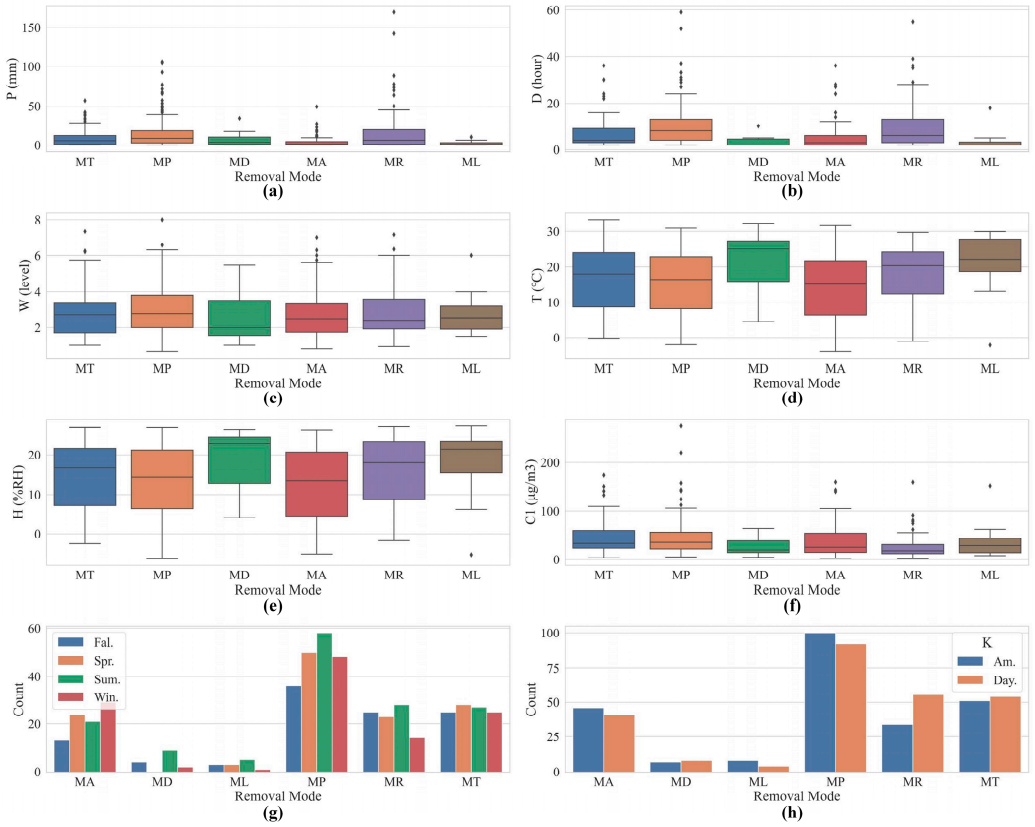

Figure 6. Distribution of the numerical distribution of concomitant factors for different removal modes. (**a**) Rainfall Total P and removal modes; (**b**) Rainfall Duration D and removal modes; (**c**) Wind Power W and removal modes; (**d**) Temperature $\overline{T}$ and removal modes; (**e**) Humidity $\overline{H}$ and removal modes; (**f**) Initial $PM_{2.5}$ C_1 and removal modes; (**g**) Seasonal factor S and removal modes; (**h**) Day and night factor K and removal modes.

4.2. Principal Component Analysis of the Removal Mode

Due to the large differences in the values of the effectors, the rainfall samples were first normalized by the eigenvector $\vec{V}$, and the high-dimensional feature samples of each type of mode were further subjected to the principal component analysis. According to the criteria for selecting the principal components, in this case, when the eigenvalues of the components were all greater than 1 and the cumulative contribution rate was 85%, there were the four types of extracted components. The results are shown in Table 5.

Table 5. Explanatory table for the principal component analysis of removal mode.

Component	Eigenvalue (λ)	Contribution of Variance $E(\lambda)$	Cumulative Contribution (%)
Principal component 1	2.774	35.67	35.67
Principal component 2	2.101	27.56	63.26
Principal component 3	1.662	16.28	79.54
Principal component 4	1.079	11.46	91.00
Principal component 5	0.662	6.061	97.06
Principal component 6	0.156	2.94	100.00

Based on the principal component eigenvalues and variance contribution rates, the factor strengths of different modes were calculated Equation (16) and the strengths of the effectors were ranked, and the total rainfall P, rainfall duration D, initial $PM_{2.5}$ concentration C_1 and wind speed scale W were selected to participate in the construction of the mode classifier (Table 6).

Table 6. Concomitant factor intensity of the sustained rainfall time series sample with removal mode.

	M_T	M_P	M_D	M_A	M_R	M_L
P	0.223	0.339	0.232	0.250	0.306	0.421
D	0.214	0.294	0.453	0.321	0.276	0.559
$\overline{T}$	0.127	0.168	−0.045	−0.174	−0.164	−0.103
$\overline{H}$	0.079	0.176	−0.137	−0.076	−0.037	0.218
W	0.263	0.132	0.232	0.271	0.386	0.372
C_1	0.193	0.266	0.411	0.263	0.442	0.421
S	0.074	0.135	0.033	0.173	−0.154	0.032
K	0.031	0.095	0.127	0.041	−0.076	0.093

4.3. Predict the Removal Mode and Phenomenon

We combine the above removal concomitant factor strengths with the factorization machines in Section 3.2.4 to build a classifier for predicting the removal mode and phenomenon: the total rainfall P, rainfall duration D, initial $PM_{2.5}$ concentration C_1 and wind speed scale W in $Data_{train}$ form the feature vector $\vec{X}$ and the classifier $\hat{y}$ in Equation (17), and the loss function in Equation (18) is combined to train the classifier according to the stochastic gradient descent (SGD) method. Figure 7 shows the effect of the removal modes classification, where $\langle C_{1,}, W \rangle$ and $\langle P, D \rangle$ are the factors that form the modal length of the vector, which is used to present the modes distribution based on the effector features; the classifier designed in this paper can classify the sustained precipitation process into the six types described according to the effectors.

To validate the removal effect classifier, all of the sustained rainfall extracted from January 2020 to December of the same year in Nanjing were classified using the samples in $Data_{test}$, and the accuracy of the classification results was evaluated by plotting the receiver operating characteristic (ROC) curve and the curve area, area under the curve (AUC), to

evaluate the accuracy of the classification results (Figure 8), where the true positive rate (*TPR*) and false positive rate (*FPR*) of the tested samples were calculated as follows:

$$\begin{cases} TPR = \frac{TP}{TP+FN} \\ FPR = \frac{TP}{FP+TN} \end{cases} \quad (19)$$

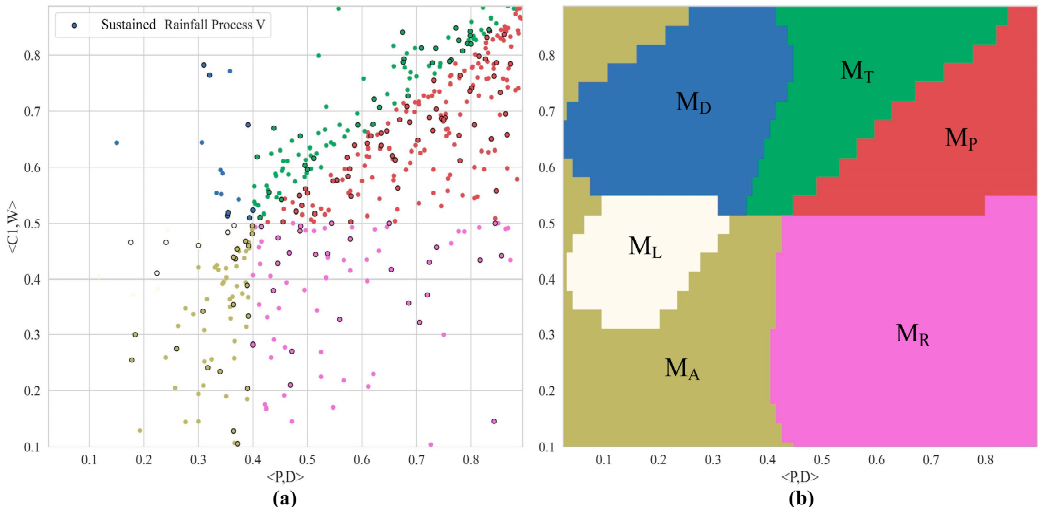

Figure 7. FM classifier result graph for removal modes, (**a**) Scatter plot of modes classification results; (**b**) Interpolated rendering of modes classification results.

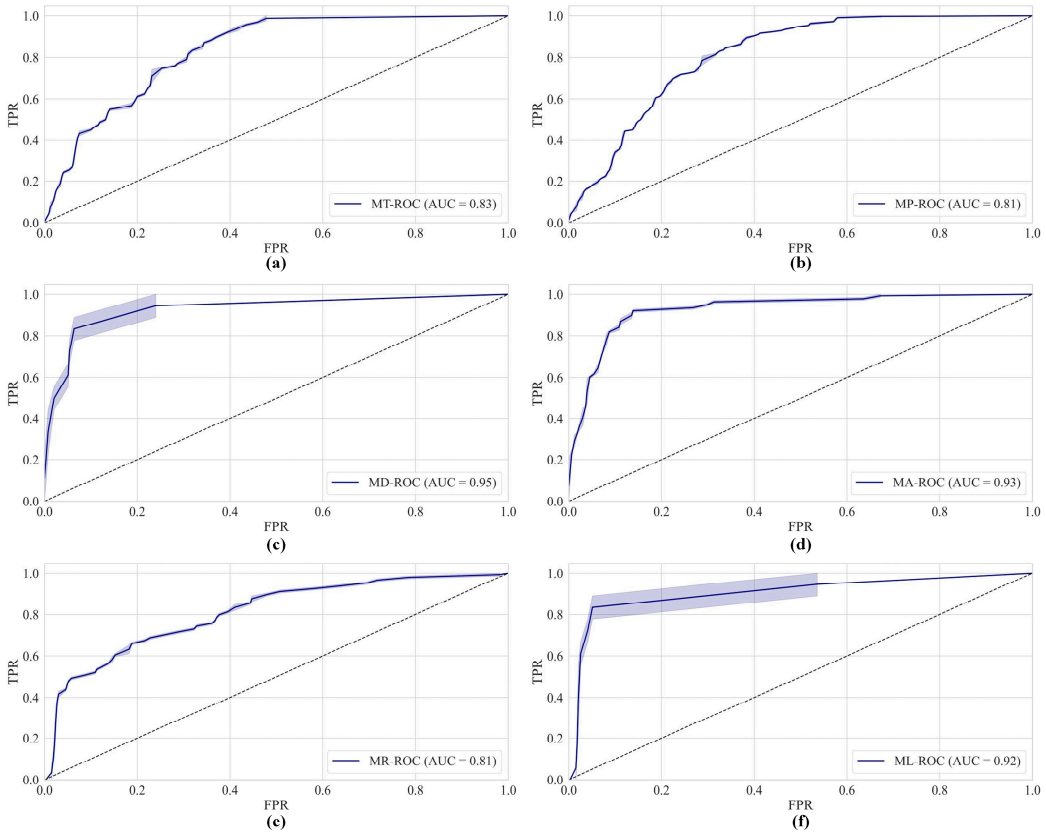

Figure 8. The ROC and AUC of removal modes classifier, (**a**) ROC of Totally Removal Mode; (**b**) ROC of Removal Mode; (**c**) ROC of Delayed Removal Mode; (**d**) ROC of Totally Ascend Mode; (**e**) ROC of Rebounding Ascend Mode; (**f**) ROC of Lasting Ascend Mode.

5. Discussion

In this paper, we mainly analyze the effect of sustained rainfall on $PM_{2.5}$ removal from the perspective of a microscopic temporal scale with historical observation data at the hourly scale. By combining previous research results, we propose a model analysis framework to quantitatively describe the removal effect from a more refined perspective for combining the mechanism of sustained rainfall effect on $PM_{2.5}$. The primary conclusions are summarized as follows.

In this paper, we use hourly scale observations for proposing models to quantitatively express sustained rainfall processes with intermittent duration and relative complexity. It is able to provide data boundaries for studying the role of rainfall removal. Moreover, we consider a large number of environmental influences and construct a concomitant factor model F, which can improve the accuracy and information dimension of the analysis. Based on the above considerations, we conclusively propose the removal modes for a quantitative description of the removal phenomenon.

- M_T, the $PM_{2.5}$ concentration change has a continuous decreasing trend during the rainfall process, which has good improvement of the air quality for a period of time after the precipitation.
- M_P, the $PM_{2.5}$ concentration change is due to the fact that when the removal of particulate pollutants by prolonged precipitation reaches its limit [29], a small portion of the particulate matter does not completely settle to the ground and floats into the air again, thus showing a slight rebound of the concentration values.
- M_D, $PM_{2.5}$ concentrations continue to rise during rainfall, but drop sharply after the end and are lower than the average concentration values before it.
- M_A, $PM_{2.5}$ concentration changes in a continuous upward trend when the rainfall duration is too short or small; the humid air will make the suspended pollutants expand, which is more likely to cause the accumulation of pollutants and make the $PM_{2.5}$ concentration rise.
- M_R, due to the longer duration of the process, there is often a short gap or the secondary precipitation is weak precipitation and other phenomena, which will cause a serious concentration rebound, making the concentration of particulate matter higher than before the precipitation.
- M_L, $PM_{2.5}$ concentrations continue to rise without rebound during rainfall, and the rise tends to scale off after the end, eventually making the $PM_{2.5}$ concentrations rise.

The method in this paper is able to classify the proposed model by historical observation data. The results show that of the sustained rainfall processes occurring in Nanjing from 2016 to 2020, only 85 were able to provide complete removal of $PM_{2.5}$, 63.4% of the precipitation processes resulted in $PM_{2.5}$ rebound and up to 177 sustained rainfall processes ultimately led to elevated $PM_{2.5}$ concentrations.

Based on the above quantitative modeling framework, we construct a classifier in combination with the model identification method, considering it for the accurate forecasting of future air quality in short periods. The accuracy evaluation results of the model show that the ROC of our constructed classifier performs well, and the AUC refers to more than 0.85, showing the reasonableness and effectiveness of the method in this paper.

Due to the limited data acquisition, more years of hourly and environmental data for $PM_{2.5}$ are lacking in this paper. Therefore, it lacks the samples of the lasting ascent mode and the delayed removal mode. Future studies are expected to obtain more hourly temporal observations for the purpose of removal modes construction and acquisition, ultimately to improve the accuracy of the prediction classifiers.

6. Conclusions

Rainfall is an effective way to remove major air pollutants such as $PM_{2.5}$. However, most studies on the relationship between rainfall data and $PM_{2.5}$ concentrations have only focused on the changes in the air quality under the influence of long-time span rainfall,

ignoring the effects of single rainfall processes that lead to increases or rebound changes in the $PM_{2.5}$ concentrations, in addition to the effect of wet deposition.

Therefore, based on the definition and generalization of the sustained rainfall process and its effects, this paper uses a time-series statistical method to extract and calculate the single sustained rainfall process and its removal effect factors based on microscopic time-scale observation data; in this process, the effect evaluation index is specifically proposed to quantitatively describe the degree of the removal effect, so as to establish a time-series effect model of $PM_{2.5}$ concentration removal by rainfall. The potential, deep-seated effect of rainfall processes on $PM_{2.5}$ concentrations is explored. The model is further combined with pattern recognition theory to design an effect pattern classifier for the sample characteristics of the rainfall process, and finally realize the micro-temporal prediction of air quality after a single rainfall. Using the hourly observation data of Nanjing from 2016 to 2020, a total of 427 sustained rainfall processes were collected using this micro-temporal time-series effects model, and the rate of process, rate of rebound and rate of final were calculated and classified into six types of modes: 85 totally removal mode, 191 partly removal mode, 14 delayed removal mode, 85 rebounding sscend mode and 12 lasting ascent mode. The classifier was constructed based on the factors, indicators and model categories, and the ROC evaluation index showed that the classifier has good performance and is capable of quantitatively predicting future $PM_{2.5}$ concentration decreases, increases and rebound effects using easily accessible rainfall and $PM_{2.5}$ concentration forecast information, with a view to providing decision-making information for future regional ambient air quality forecasting and refined control. For the acquisition of hour-by-hour $PM_{2.5}$ concentration forecasts, further investigation of the finer variation characteristics within rainfall periods is required on the basis of this study.

Author Contributions: Conceptualization, X.X.; methodology, X.X. and T.W.; software, T.W.; validation, T.W. and X.X.; formal analysis, X.X. and T.W.; investigation, T.W.; resources, X.X.; data curation, B.X.; writing—original draft preparation, T.W. and X.X.; writing—review and editing, X.X. and T.W.; visualization, T.W.; supervision, X.X. and T.L.; project administration, B.X.; funding acquisition, B.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (41971166) and LZJTU EP (201806).

Institutional Review Board Statement: Ethical review and approval were waived because necessary permissions were obtained from the local governments to which the schools are affiliated for this study. Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Informed Consent Statement: Studies not involving human (Not applicable).

Data Availability Statement: Data available on request due to ethical restrictions. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to not having participants consent to share their anonymized data with a third party.

Acknowledgments: We express our sincere appreciation to Qing Zhu from the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University and the anonymous reviewer for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wu, H.; Gao, X. Multimodal Data Based Regression to Monitor Air Pollutant Emission in Factories. *Sustainability* **2021**, *13*, 2663. [CrossRef]
2. Wang, Z.; Liang, L.; Wang, X. Spatiotemporal evolution of $PM_{2.5}$ concentrations in urban agglomerations of China. *J. Geogr. Sci.* **2021**, *31*, 878–898. [CrossRef]
3. Pui, D.; Chen, S.C.; Zuo, Z. $PM_{2.5}$ in China: Measurements, sources, visibility and health effects, and mitigation. *Particuology* **2014**, *13*, 1–26. [CrossRef]

4. Amrita, T. Study of Ambient Air Quality Trends and Analysis of Contributing Factors in Bengaluru, India. *Orient. J. Chem.* **2017**, *33*, 1051–1056.
5. Kao, Y.H.; Lin, C.W.; Chiang, J.K. Predictive Meteorological Factors for Elevated $PM_{2.5}$ Levels at an Air Monitoring Station Near a Petrochemical Complex in Yunlin County, Taiwan. *Open J. Air Pollut.* **2019**, *8*, 1–17. [CrossRef]
6. Preethi, B.; Kripalani, R.H.; Kumar, K.K. Indian summer monsoon rainfall variability in global coupled ocean-atmospheric models. *Clim. Dyn.* **2010**, *35*, 1521–1539. [CrossRef]
7. Neal, C.; Reynolds, B.; Neal, M.; Pugh, B.; Hill, L.; Wickham, H. Long-term changes in the water quality of rainfall, cloud water and stream water for moorland, forested and clear-felled catchments at Plynlimon, mid-Wales. *Hydrol. Earth Syst. Sci.* **1999**, *5*, 459–476. [CrossRef]
8. Liu, Z.; Shen, L.; Yan, C.; Du, J.; Li, Y.; Zhao, H. Analysis of the Influence of Precipitation and Wind on $PM_{2.5}$ and PM10 in the Atmosphere. *Adv. Meteorol.* **2020**, *2020*, 1–13. [CrossRef]
9. Yamamoto, T.; Hori, K.; Tatebayashi, J. An experimental investigation of the PM adhesion characteristics in a fluidized bed type PM removal device. *Powder Technol.* **2016**, *289*, 31–36. [CrossRef]
10. Qin, S.; Liu, F.; Wang, J.; Sun, B. Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models. *Atmos. Environ.* **2014**, *98*, 665–675. [CrossRef]
11. Fu, S.M.; Sun, J.H.; Ling, J.; Wang, H.J.; Zhang, Y.C. Scale interactions in sustaining persistent torrential rainfall events during the Mei-yu season. *J. Geophys. Res. Atmos.* **2016**, *121*, 12856–12876. [CrossRef]
12. Liu, Z.Z.; Yan, Z.X.; Duan, J.; Qiu, Z.H. Infiltration regulation and stability analysis of soil slope under sustained and small intensity rainfall. *J. Cent. South Univ.* **2013**, *20*, 2519–2527. [CrossRef]
13. Das, A.V.; Basu, S. Temporal trend of microsporidial keratoconjunctivitis and correlation with environmental and air pollution factors in India. *Indian J. Ophthalmol.* **2021**, *69*, 1089.
14. Sharma, R.C.; Sharma, N. Statistical Investigation of Effect of Rainfall on Air Pollutants in the Atmosphere, Haryana State, Northern India. *Am. J. Environ. Prot.* **2018**, *6*, 14–21.
15. Shaibu, V.O.; Weli, V.E. Relationship between $PM_{2.5}$ and Climate Variability in Niger Delta, Nigeria. *Am. J. Environ. Prot.* **2017**, *5*, 20–24. [CrossRef]
16. Jassim, M.S.; Coskuner, G.; Munir, S. Temporal analysis of air pollution and its relationship with meteorological parameters in Bahrain. *Arab. J. Geosci.* **2018**, *11*, 1–15. [CrossRef]
17. Witkowska, A.; Lewandowska, A.U. Water soluble organic carbon in aerosols (PM1, PM2.5, PM10) and various precipitation forms (rain, snow, mixed) over the southern Baltic Sea station. *Sci. Total Environ.* **2016**, *573*, 337–346. [CrossRef]
18. Wei, L.A.; Yza, B.; Ping, L.C. Numerical simulation of the influence of major meteorological elements on the concentration of air pollutants during rainfall over Sichuan Basin of China. *Atmos. Pollut. Res.* **2020**, *11*, 2036–2048.
19. Kapwata, T.; Wright, C.Y.; du Preez, D.J.; Kunene, Z.; Mathee, A.; Ikeda, T.; Landmandi, W.; Maharajj, R.; Sweijdk, N.; Minakawa, N.; et al. Exploring rural hospital admissions for diarrhoeal disease, malaria, pneumonia, and asthma in relation to temperature, rainfall and air pollution using wavelet transform analysis. *Sci. Total Environ.* **2021**, *2021*, 148207.
20. Zhu, T.; Shang, J.; Zhao, D.F. The roles of heterogeneous chemical processes in the formation of an air pollution complex and gray haze. *Sci. China-Chem.* **2011**, *54*, 145–153. [CrossRef]
21. Liu, F.; Tan, Q.; Jiang, X.; Yang, F.; Jiang, W. Effects of relative humidity and $PM_{2.5}$ chemical compositions on visibility impairment in Chengdu, China. *J. Environ. Sci.* **2019**, *86*, 15–23. [CrossRef]
22. Guan, P.; Zhang, H.; Zhang, Z.; Chen, H.; Li, Y. Assessment of Emission Reduction and Meteorological Change in $PM_{2.5}$ and Transport Flux in Typical Cities Cluster during 2013–2017. *Sustainability* **2021**, *13*, 5685. [CrossRef]
23. Zhang, C.; Pei, H.; Jia, Y.; Bi, Y.; Lei, G. Effects of air quality and vegetation on algal bloom early warning systems in large lakes in the middle–lower Yangtze River basin. *Environ. Pollut.* **2021**, *2012*, 117455. [CrossRef]
24. Amelija, D.; Nenad, Z.; Emina, M.; Jasmina, R.; Miomir, R.; Ljiljana, Z. The effect of pollutant emission from district heating systems on the correlation between air quality and health risk. *Therm. Sci.* **2011**, *15*, 293–310. [CrossRef]
25. Li, Y.; Zhu, K. Spatial dependence and heterogeneity in the location processes of new high-tech firms in Nanjing, China. *Pap. Reg. Sci.* **2017**, *96*, 519–536. [CrossRef]
26. Zhai, K.; Gao, X.; Zhang, Y.; Wu, M. Perceived Sustainable Urbanization Based on Geographically Hierarchical Data Structures in Nanjing, China. *Sustainability* **2019**, *11*, 2289. [CrossRef]
27. Chen, K.; Yan, Y.; Hu, Z. Influence of Air Pollutants on Fog Formation in Urban Environment of Nanjing, China. *Procedia Eng.* **2011**, *24*, 654–657. [CrossRef]
28. Wang, G.; Huang, L.; Gao, S.; Song, T. Characterization of water-soluble species of PM10 and PM2.5 aerosols in urban area in nanjing, china. *Atmos. Environ.* **2002**, *36*, 1299–1307. [CrossRef]
29. Chen, T.; He, J.; Lu, X.; She, J.; Guan, Z. Spatial and temporal variations of $PM_{2.5}$ and its relation to meteorological factors in the urban area of nanjing, china. *Int. J. Environ. Res. Public Health* **2016**, *13*, 921. [CrossRef]
30. Zou, X.; Qian, Y.; Zhang, S. Spatiotemporal Variations of $PM_{2.5}$ Concentration and Relationship with Other Criteria Pollutants in Nanjing, China. *Nat. Environ. Pollut. Technol.* **2018**, *17*, 499–505.
31. Guijo-Rubio, D.; Duran-Rosal, A.M.; Gutierrez, P.A.; Troncoso, A.; Hervas, M.C. Time-Series Clustering Based on the Characterization of Segment Typologies. *IEEE Trans. Cybern.* **2020**, *99*, 1–14. [CrossRef] [PubMed]

32. Leles, M.C.; Sansão, J.P.H.; Mozelli, L.A.; Guimarães, H.N. Improving reconstruction of time-series based in singular spectrum analysis: A segmentation approach. *Digit. Signal Process.* **2017**, *77*, 63–76. [CrossRef]
33. Onuorah, C.U.; Leton, T.G.; Momoh, Y. Influence of meteorological parameters on particle pollution (PM2.5 and PM10) in the tropical climate of port harcourt, nigeria. *Arch. Curr. Res. Int.* **2019**, 1–12. [CrossRef]
34. Kayes, I.; Shahriar, S.A.; Hasan, K.; Akhter, M.; Salam, M.A. The relationships between meteorological parameters and air pollutants in an urban environment. *Glob. J. Environ. Sci. Manag.* **2019**, *5*, 265–278.
35. Uchiyama, R.; Okochi, H.; Ogata, H.; Katsumi, N.; Nakano, T. Characteristics of trace metal concentration and stable isotopic composition of hydrogen and oxygen in "urban-induced heavy rainfall" in downtown Tokyo, Japan; the implication of mineral/dust particles on the formation of summer heavy rainfall. *Atmos. Res.* **2018**, *217*, 73–80. [CrossRef]
36. Gautam, S.; Yadav, A.; Pillarisetti, A.; Smith, K.; Arora, N. Short-term introduction of air pollutants from fireworks during diwali in Rural Palwal, Haryana, India: A case study. *IOP Conf. Ser.: Earth Environ. Sci.* **2018**, *120*, 012009. [CrossRef]
37. Yang, W.; He, Z.; Huang, H.; Huang, J. A clustering framework to reveal the structural effect mechanisms of natural and social factors on $PM_{2.5}$ concentrations in China. *Sustainability* **2021**, *13*, 1428. [CrossRef]
38. Kashki, A.; Karami, M.; Zandi, R.; Roki, Z. Evaluation of the effect of geographical parameters on the formation of the land surface temperature by applying ols and gwr, a case study shiraz city, Iran. *Urban Clim.* **2021**, *37*, 100832. [CrossRef]
39. Zhang, Y.; Zhang, F.; Zhu, H.; Guo, P. An optimization-evaluation agricultural water planning approach based on interval linear fractional bi-level programming and iahp-topsis. *Water* **2019**, *11*, 1094. [CrossRef]
40. Kulanuwat, L.; Chantrapornchai, C.; Maleewong, M.; Wongchaisuwat, P.; Boonya-Aroonnet, S. Anomaly detection using a sliding window technique and data imputation with machine learning for hydrological time series. *Water* **2021**, *13*, 1862. [CrossRef]
41. Cai, T.T.; Ma, Z.; Wu, Y. Sparse pca: Optimal rates and adaptive estimation. *Ann. Stat.* **2012**, *41*, 3074–3110. [CrossRef]
42. Wang, S.; Li, C.; Zhao, K.; Chen, H. Learning to context-aware recommend with hierarchical factorization machines. *Inf. Sci.* **2017**, *409–410*, 121–138. [CrossRef]