sustainability

MDPI

Article

Spatiotemporal Distribution and Health Risk of $PM_{2.5}$ in the Urban Belt Around the Tarim Basin, China

Liang Guo [1,*], Qian Sun [2], Guizhen Gao [2], Yueqi Zhao [2] and Lunan Chen [1]

[1] College of Resources and Environment, Xinjiang Agricultural University, Urumqi 830052, China; chenlunan7411@163.com

[2] College of Forestry and Landscape Architecture, Xinjiang Agricultural University, Urumqi 830052, China; sq061@163.com (Q.S.); gaoguizhen1984@163.com (G.G.); 15146442758@163.com (Y.Z.)

* Correspondence: guoliang@xjau.edu.cn

Abstract

The Taklamakan Desert, the primary source of sand and dust in China, has a significant impact on $PM_{2.5}$ concentrations in the urban belt around the Tarim Basin. In this study, moderate-resolution imaging spectroradiometer (MODIS) images and machine learning algorithms were used to simulate $PM_{2.5}$ concentrations and analyze deaths from ischemic heart disease (IHD), chronic obstructive pulmonary disease (COPD), lung cancer (LC), and stroke (STK) attributed to $PM_{2.5}$. From 2015 to 2024, the aerosol optical depth (AOD) showed a decreasing trend, and the spatial distribution pattern gradually increased from west to east. The counties and cities with higher average $PM_{2.5}$ concentrations during the 2024 dust-prone season were Awati County, Alar City, Moyu County, and Aksu City, with average concentrations of 89.7 $\mu g/m^3$, 88.8 $\mu g/m^3$, 86.0 $\mu g/m^3$, and 84.1 $\mu g/m^3$, respectively. $PM_{2.5}$ concentrations during the non-dust season were significantly lower (approximately 50%) than those during the dust-prone season. Among the four major fatal diseases, STK and IHD accounted for the majority of attributable deaths, with proportions exceeding 80% of the total. The results show that $PM_{2.5}$ concentrations can be accurately monitored over a large scale, addressing the shortage of fixed monitoring stations, and providing a theoretical basis for sustainable development of ecological environment, health-risk management and control in the urban belt around the Tarim Basin.

Keywords: machine learning; AOD; $PM_{2.5}$; health risks; environmental sustainability

1. Introduction

$PM_{2.5}$ refers to particulate matter with an aerodynamic diameter of 2.5 µm or less. Exposure to $PM_{2.5}$ pollution is associated with an increased risk of various cardiovascular and respiratory diseases [1,2]. Although the establishment of ground-based monitoring stations has improved the capacity to monitor $PM_{2.5}$, most of these stations are concentrated in bustling urban areas, and many people still live far from monitoring stations, such as in rural areas. Currently, the continued lack of $PM_{2.5}$ concentration data hinders effective air pollution assessment and the associated health risks at greater spatial scales.

$PM_{2.5}$ concentration data can be obtained using various approaches, including ground-based monitoring, satellite remote-sensing retrieval, and chemical transportation model simulations [3]. Although aerosol observations obtained by remote sensing cannot directly represent ground-level particulate matter concentrations, mathematical relationships between aerosol optical depth (AOD) and $PM_{2.5}$ can be established using ground-based $PM_{2.5}$

Check for updates

Academic Editor: Deming Han

Received: 4 July 2026
Revised: 12 August 2026
Accepted: 14 August 2026
Published: 15 August 2026

measurements [4]. Therefore, AOD is widely regarded as a useful proxy for $PM_{2.5}$ pollution levels. The moderate-resolution imaging spectroradiometer (MODIS) aboard NASA's Terra satellite can effectively retrieve AOD data. Over the years, MODIS AOD products have provided an important data basis for estimating $PM_{2.5}$ concentrations and assessing air quality by exploiting the relationship between AOD and $PM_{2.5}$. Validation across multiple temporal scales has demonstrated significant statistical associations between AOD and $PM_{2.5}$ in annual, seasonal, and monthly comparisons. Seasonal variations substantially affect the relationship between AOD and $PM_{2.5}$ [5]. In humid regions, both MODIS AOD and $PM_{2.5}$ concentrations exhibit pronounced seasonal variations [6]. At a local scale, such as within an industrial area, linear regression models can be developed to estimate surface-level $PM_{2.5}$ concentrations based on the relationships between daily and monthly mean $PM_{2.5}$ and AOD observations. These findings further demonstrate that AOD data can be used to retrieve $PM_{2.5}$ concentrations at regional and even finer spatial scales [7]. The advantages of remote sensing have become increasingly evident as the spatial scale of analysis expands. Long-term MODIS AOD records enable the estimation of $PM_{2.5}$ concentrations over extensive areas. However, before such estimations are performed, the relationships and variation patterns between AOD and $PM_{2.5}$ should be characterized as precisely as possible at different spatial scales to improve the robustness and reliability of AOD-based $PM_{2.5}$ estimates [8]. Improving the accuracy of satellite-based $PM_{2.5}$ estimations remains an important research objective because of missing satellite observations and the limited spatial resolution of remote-sensing products. A two-stage algorithm can effectively improve estimation accuracy [9]. By accounting for spatiotemporal heterogeneity and incorporating key meteorological variables, this approach can improve model accuracy and generate spatially continuous, gap-free $PM_{2.5}$ concentration estimates [10]. By comparison, although MCD19A2 data has a higher spatial resolution, its product release date was in 2018, which is relatively later. Additionally, the official data source does not directly provide aggregated monthly products, making MYD08 and MYD04 more practical and reliable for tracking monthly dynamic changes [11]. Moreover, the Dark Target (DT) and Deep Blue algorithms used in MYD08 have been iterated and optimized over multiple generations. The pattern in the long-term shift in global monthly averages has been fully quantified, and the trend analysis of monthly dynamic changes can be directly eliminated using validated deviation-correction methods to eliminate system errors. The reliability of time series analysis is higher, and the product validation system is more complete [12,13].

Machine learning algorithms offer distinct advantages for addressing classification and regression problems, and numerous studies have applied machine learning methods to air pollution research [14,15]. Machine learning is an important tool and effective strategy for data mining. Given their inherent strength, machine learning algorithms hold considerable promise for $PM_{2.5}$ concentration estimation. Cubist, random forest (RF), extreme gradient boosting (XGBoost), and back-propagation (BP) neural networks are popular nonlinear rule-based machine learning techniques. When combined with satellite remote-sensing data, they have considerable value in developing $PM_{2.5}$ estimation models [16]. Conventional machine learning algorithms can be used to develop effective models for estimating $PM_{2.5}$ concentrations from AOD, often achieving strong predictive performance [17,18]. As climatic and environmental conditions vary among study areas, a common methodological approach is to compare multiple machine learning algorithms for $PM_{2.5}$ prediction and select the best-performing model [19,20]. Combining machine learning models with average and K-fold cross-validation can better capture PM_{10} and $PM_{2.5}$ mass concentration point changes, thereby improving model evaluation reliability and methodological rigor [21,22]. To explore $PM_{2.5}$ concentration changes more comprehensively, the estimation models incorporate not only AOD data, but also meteorological factors, land use, and other relevant predictors [23,24]. Machine learning algorithm models are still under development. Stacking ensemble models that integrate base learners, such as AdaBoost, XGBoost, and RF, can overcome some of the limitations associated with individual models and further improve estimation performance [25].

Numerous studies have reported significant associations between $PM_{2.5}$ exposure and respiratory diseases, chronic obstructive pulmonary disease (COPD), cardiovascular and cerebrovascular diseases, and lung cancer (LC) [26,27]. In a meta-analysis of 196 studies, Orellano et al. found a positive association between $PM_{2.5}$ exposure and all-cause mortality (RR = 1.007, 95% CI: 1.004–1.009) [28]. Associations have also been reported between $PM_{2.5}$ exposure and mortality attributable to cardiovascular disease (CVD), cerebrovascular disease, chronic kidney disease, COPD, dementia, type 2 diabetes, hypertension, LC, and pneumonia [29,30]. In particular, cardiovascular mortality was significantly associated with ambient particulate matter concentrations. Significant associations between $PM_{2.5}$ exposure and mortality from most disease categories were also observed during the colder months [31]. Long-term exposure to $PM_{2.5}$ is associated with increased mortality, and $PM_{2.5}$-attributable mortality generally increases with increasing exposure levels [32]. A decade ago, the Global Burden of Disease study identified ambient $PM_{2.5}$ exposure as the fifth leading mortality risk factor worldwide. Between 1990 and 2015, the number of deaths attributable to $PM_{2.5}$ exposure increased substantially, accounting for 7.6% of the total global death toll, nearly 60% of which occurred in East and South Asia [33]. In Asian countries, such as South Korea and Thailand, particulate matter pollution represents an important environmental health risk, particularly to the respiratory system, and has become a major public health concern in many densely populated urban areas. $PM_{2.5}$-attributable deaths due to IHD, CHD, LC, and cerebrovascular disease have reached substantial levels [34,35]. During the novel coronavirus pneumonia epidemic, $PM_{2.5}$ concentrations exhibited marked temporal variations, which may have resulted in differing population health impacts, particularly relevant to CVD outcomes [36]. Particulate matter pollution is a major public health concern in India that adversely affects respiratory health and places considerable pressure on healthcare resources. It has also been associated with COPD, asthma, LC, and pulmonary fibrosis [37].

$PM_{2.5}$ pollution has also adversely affected population health in China [38–42]. Fine $PM_{2.5}$, characterized by small particle size, large specific surface area, and high adsorption capacity, is increasingly recognized as an independent yet insufficiently understood environmental risk factor for CVD [43]. A time-series study conducted in Ningbo, Zhejiang Province, China, found that $PM_{2.5}$ and PM_{10} concentrations were significantly associated with the risk of non-accidental mortality during the warm season, although the incidence was higher in women and the elderly aged over 65 years [44]. A substantial number of premature deaths in China are attributed to ambient $PM_{2.5}$ exposure. Premature deaths attributable to IHD, cerebrovascular disease, COPD, and LC account for a relatively large proportion of mortalities. Implementing a more stringent $PM_{2.5}$ standard of $<15\ \mu g/m^3$ could substantially reduce premature mortality [45]. Following the implementation of a series of national air pollution prevention and control action programs, China experienced marked improvements in ambient air quality, accompanied by a progressive decrease in the number of premature deaths attributable to $PM_{2.5}$ [46].

The urban belt around the Tarim Basin in Xinjiang, China, supports a large population distributed among geographically dispersed oasis settlements. However, ground-based monitoring networks in this region are sparse, and the area is frequently affected by severe dust events. The regional-scale spatiotemporal distribution of $PM_{2.5}$ concentrations and the associated population health risks remain poorly understood. Therefore, this study employed remote-sensing data and machine learning algorithms to estimate $PM_{2.5}$

concentrations, thereby providing a new approach for evaluating regional-scale $PM_{2.5}$ pollution, sustainable development of ecological environment, and associated population health risks in the urban belt around the Tarim Basin.

2. Materials and Methods

2.1. The Study Area

As shown in Figure 1, the urban belt around the Tarim Basin in Xinjiang, China, is located south of the Tianshan Mountains and north of the Kunlun Mountains in the Xinjiang Uygur Autonomous Region. The geographical location is between E 74°50′–90°10′, N 34°30′–44°30′, and the area is approximately 1.08×10^6 km^2. As the center of the basin is covered by the Taklamakan Desert, towns, farmlands, and traffic lines are forced to lie around the edge of the basin, forming a typical circular spatial pattern. Administrative divisions include the Bayingolin Mongol Autonomous, Aksu, Kashi, Kizilsu Kirghiz Autonomous, and Hotan Prefectures. Owing to climatic, topographical, and precipitation limitations, the oases are scattered in the desert and the residents of the urban belt around the Tarim Basin inhabit the Gobi region. The region has a large temperature difference, sufficient sunshine, droughts, minimal rain, high evaporation, and a temperate continental arid climate. The Taklamakan Desert, in the center of the Tarim Basin, is the largest desert in China and the second largest mobile desert in the world, with an area of approximately 3.3×10^5 km^2. It is the primary source of dust in China and is highly susceptible to dusty weather conditions.

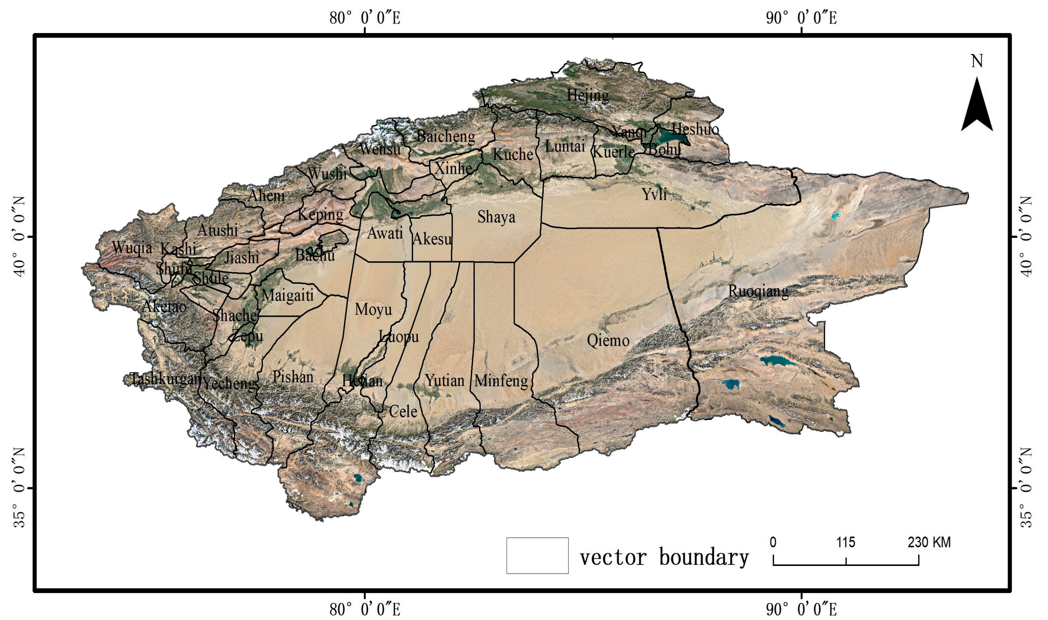

Figure 1. Geographical location of the study area.

2.2. Data Source

2.2.1. MODIS AOD Data

MODIS is a key sensor on the Terra and Aqua satellites. It is an optical remote-sensing instrument that combines maps. It is used for long-term global observations of the surface, biosphere, atmosphere, and oceans. The remote-sensing image (MYD08_M3 V6.1) used in this study was derived from the NASA official website (https://www.earthdata.nasa.gov/) (accessed on 16 June 2025) and is a global-scale atmospheric aerosol product with a time resolution of 1 month and a spatial resolution of 1 × 1°. The wavelength of the AOD product is 550 nm (in the visible light band). The AOD data of November and December 2024 were supplemented by MYD04_3K due to the lack of MYD08 products. MYD04 is the average AOD value over a 3 × 3 km grid, making it suitable for aerosol climatology research at global and regional scales. This product can be used to study clouds, water vapor, aerosols, trace gases, land surfaces, and ocean properties. To quantitatively evaluate the applicability of MYD data in arid areas, we set up two groups for comparison: one including the dust-prone season and the other the non-dust season. We took April as the dust-prone season and October as the typical month representing the non-dust season, respectively. Therefore, we selected April and October in 2018 and 2024 as the verification groups for comparison, and converted the daily AOD data from MCD19A2 into monthly averages, which were then compared and fitted to the spatial distribution of MYD08 data for the same period. The spatial differentiation characteristics of the two showed a high degree of consistency in the urban belt around the Tarim Basin in Xinjiang. All datasets were resampled to a grid size of 0.25°, and their spatial trends were compared (Text S1; Table S1).

Simultaneously, given that many researchers have conducted time-series cross-validation on the above data, it has been shown that the outlier-screening mechanism for MYD08 monthly scale synthesis is more complete and can produce quality-weighted monthly products with stronger stability in monthly means [47,48]. It further demonstrates the strong applicability of MYD08 data to typical arid areas, such as the urban belt around the Tarim Basin in Xinjiang.

In this study, the monthly mean AOD data from January 2015 to December 2024 were obtained. Owing to the lack of AOD values in some study areas, we used resampling and Kriging interpolation methods, setting the grid size to 0.25°, which matched the spatial resolution of meteorological parameters and enabled spatial-scale assimilation of AOD with all meteorological factors. The original storage value range of the MYD08 product is −100–5000, and the data conversion must be completed by using the officially specified 0.001 scaling factor. The negative value represents invalid pixels, such as cloud coverage and surface reflection anomaly [49].

2.2.2. $PM_{2.5}$ Measured Data

In this study, the $PM_{2.5}$ measured data were obtained from the China Environmental Monitoring Station National Urban Air Quality Real-Time Publishing Platform (https://www.cnemc.cn/sssj/) (accessed on 30 June 2025). The data type was the $PM_{2.5}$ 24-h average concentration, data duration was from January 2015 to December 2024, and time resolution was one day. To match the AOD data, the $PM_{2.5}$ 24-h average concentration was processed as the monthly average in this study. $PM_{2.5}$ measured data were used for training and verification of the machine learning algorithms.

2.2.3. Meteorological Data

Fully considering the influence of meteorological data, such as temperature, wind speed, wind direction, and air humidity, on $PM_{2.5}$ can effectively improve the accuracy of the $PM_{2.5}$–AOD correlation model [50]. In this study, some local meteorological factor data during the study period of 2015 and 2024 were used. These data were published by the European Center for Medium-Range Weather Forecasts. Based on international authoritative meteorological inversion data, they were assimilated, corrected, and downscaled using measured data from local meteorological stations. Monthly meteorological data were selected as the predictors of the $PM_{2.5}$ concentration estimation model, and the specific parameters are listed in Table 1.

Table 1. Meteorological data parameter statistics.

Meteorological Factor	Data Source	Temporal Resolution	Spatial Resolution	Data Profile
Air temperature (°C)	https://www.ecmwf.int (accessed on 23 June 2025)	/h	0.25	The temperature in the louver box at 1.5–2 m above the ground.
Air humidity (%)		/h	0.25	The air humidity is 1.25~2 m above the ground.
Atmospheric pressure (hPa)		/h	0.25	Refers to the air pressure value in this area.
Precipitation (mm/h)		/h	0.25	The depth at which liquid or solid (melted) water, falling from the sky to the ground, accumulates on a horizontal plane without evaporation, infiltration or loss.
Surface wind speed (m/s)		/h	0.25	The wind speed is about 10 m above the ground.

Hourly air temperature (°C), air humidity (%), atmospheric pressure (hPa), precipitation (mm/h), and surface wind speed (m/s) data were used to calculate daily and monthly means. Using ArcGIS software 10.5, the remote sensing data of meteorological factors were projected, transformed, and masked. The raster data was clipped to the study area, laying the foundation for the spatial analysis of the follow-up data. Using SPSS 26.0, the measured surface temperature, precipitation, and wind speed data were used to fit the GCOM-W1 satellite table data linearly. We calculated the trend in the measured meteorological and satellite data at corresponding points and found that over 85% of the grids showed consistent trend values (increase or decrease). The satellite data was mathematically corrected using the measured meteorological data. To avoid further increasing data heterogeneity, a 2% linear stretch was applied to the fitted results; this stretch treats the lowest and highest 2% of pixel values in satellite data as outliers and directly discards them. After removing the outliers at both ends, the pixel values within the remaining 2–98% range were linearly mapped to the full display range, avoiding compression of the effective signal's display range [51]. This operation method is beneficial for removing outliers, suppressing noise, and enhancing contrast. Resampling all meteorological factor data to 0.25° not only supports large-scale research in the study area but also aligns with data scales such as AOD, establishing a consistent spatial scale for the construction of subsequent mathematical models.

2.2.4. Population Density Data

Population density data were derived from the WorldPop R2025A v1 dataset with a resolution of 1 × 1 km, developed by the University of Southampton. The dataset used top-down and interpolation methods to predict population data at two time points, 2015 and 2030, to obtain a complete annual population forecast from 2015 to 2030. The population was decomposed using the RF areal measurement method, and grid population density data covering 242 countries and regions from 2015 to 2030 were constructed [52]. In this study, the resolution of the population density data was adjusted for consistency with the AOD data.

2.3. Methods

2.3.1. K-Fold Cross-Validation

K-fold cross-validation is the most commonly used method for estimating the likelihood that machine learning results are due to chance, and its performance is usually better than that of traditional hypothesis testing. To ensure that the training and testing sets have sufficient samples, data partitioning based on K-fold cross-validation is necessary. Essentially, cross-validation involves dividing the data into k subsets (folds), and training and validation are performed k times. During iterative training and verification, the number ranges from 1 to K, and K rounds of iterations are performed. In each iteration, one k subset is used as the validation set, and the remaining $k - 1$ subsets are used for training. The k determines how many folds the data set is divided into. As the number of folds increases, the number of training subsets will also increase [53]. Merging the K-round test results, usually by averaging them, yields the final model performance index [54]. Here, K = 5 was used as a classic cross-validation fold.

2.3.2. Machine Learning Algorithms Simulate $PM2_{.5}$ Concentrations

Machine learning is an important part of data science and is widely used to analyze nonlinear ecological and environmental problems [55]. Machine learning uses known input and corresponding output data as training sets to construct models such that computers can automatically learn and improve optimization from existing data to classify or regress new data. In this study, AOD and $PM_{2.5}$ measured data from monitoring stations were trained using a BP neural network, RF, support vector machine (SVM) and convolutional neural network (CNN) algorithms, and the algorithm with the best fitting effect in arid areas was selected to simulate the spatial distribution of $PM_{2.5}$ concentrations in the urban belt around the Tarim Basin. We selected AOD and five key meteorological factors (air temperature, air humidity, surface wind speed, atmospheric pressure, and precipitation) as independent variables, and $PM_{2.5}$ concentration as the dependent variable for training and regression prediction.

There are 11 monitoring stations, and the station information is shown in Table 2. Due to the close proximity of some monitoring stations, $PM_{2.5}$ concentrations in each region were arithmetic-averaged using data from five monitoring stations: Bayingolin, Aksu, Kashi, Hotan, and Kizilsu. The observation data from the monitoring stations are the 24-h daily averages from 1 January 2015 to 31 December 2024. Due to inconsistencies in the construction times of monitoring stations, daily operations and quality control, equipment failures, and other factors, data may be missing or discontinuous. To ensure the accuracy of the results, this study used only monthly averages of continuous $PM_{2.5}$ concentrations, AOD, and meteorological data to construct a dataset. Finally, a total of 194 AOD–meteorology–$PM_{2.5}$ datasets were obtained to ensure that each sample contains the monthly averages of $PM_{2.5}$ concentration, AOD, and meteorological parameters for a monitoring station over the entire month.

The BP neural network is a widely used artificial neural network that can provide good simulation and prediction of nonlinear relationships, including input, hidden, and output layers. The learning process consists of two steps: signal forward propagation and error BP. In forward propagation, the input samples are introduced from the input layer and processed from each hidden layer to the output layer. If the actual output of the output layer does not match the expected output, it is transferred to the error BP stage. Error BP is used to transmit the output error to the input layer-by-layer through the hidden layer in some form and to allocate the error to all units of each layer to obtain the error signal of each layer unit. This error signal is used as the basis for correcting the weight of each unit. The weight adjustment process of each signal forward propagation and error BP layer was performed periodically. The continuous adjustment of weights is the network learning and training process. This process continues until the network output error is reduced to an acceptable level or until the preset number of learning iterations is reached [56].

Table 2. Monitoring Station information.

Station	Prefecture	Number of Data Sets	Longitude	Latitude
1#	Bayingolin Mongolian Autonomous Prefecture	50	86.147	41.754
2#			86.197	41.717
3#			86.236	41.717
4#	Aksu Prefecture	50	80.292	41.169
5#			80.261	41.155
6#	Kashi Prefecture	25	76.007	39.447
7#			75.976	39.469
8#	Hotan Prefecture	44	79.924	37.088
9#			79.913	37.113
10#	Kizilsu Kirghiz Autonomous Prefecture	25	76.145	39.714
11#			76.166	39.716

RF is an ensemble classifier developed from the regression decision tree proposed by Breiman [57]. The bootstrap resampling technique is used to randomly select a certain number of samples from the original training sample set, generate a new training sample set, and use these self-help sample sets to construct multiple classification trees and form an RF. For new data samples, the RFs determine reliability of the classification results based on the voting status of the classification tree. RFs are constructed by integrating multiple decision trees, each based on independently extracted samples. RF comprises multiple decision trees. Its core concept is to aggregate multiple weak classifiers to improve decision-making ability. The model must adjust fewer parameters and has a strong generalization ability [58]. In the regression problem, the predicted value of the RF model is the mean of the output results of all the decision trees.

An SVM was developed based on statistical learning theory [59]. Compared to traditional large sample modeling methods, SVM has a unique advantage in small-sample machine learning. SVM regression is a machine learning model that uses interval maximization loss and kernel functions. Compared with traditional neural networks, it avoids dimensional disasters and overfitting and improves generalization ability. In the regression problem, the SVM attempts to find a function point in the input space and classification hyperplane to minimize differences between the predicted and actual outputs. The independent variables are normalized so that the data are normalized to between $[-1, 1]$. When SVM was used in this study, the Radial Basis Function (RBF) was introduced to map the samples to high-dimensional space. The RBF kernel function is expressed as:

$$\min_{\mathrm{w,b},\xi} \frac{1}{2} w^T \mathrm{w} + \mathrm{C} \sum\nolimits_{\mathrm{i}=1}^{i} \varepsilon_i$$

$$\mathrm{K}\mathrm{x}_i, \mathrm{x}_j = \exp\left(-\gamma \left| x_i - \mathrm{x}_j \right|^2\right)$$

where C is the penalty parameter, which controls the penalty applied to misclassified samples. ε is the loss function parameter, which represents the allowable error tolerance, and ζ is the relaxation factor, which determines the tolerance of the model to classification errors. γ is the reciprocal of the influence radius of the support vector selected by the model [60,61].

The unique bionic design of CNN enables efficient extraction of the rich features of the target object with low computational complexity. The core of its architecture includes multiple levels: the data input layer, which introduces the information to be processed; the convolution layer performs the key steps of feature mapping; and the pooling layer reduces the number of parameters by down-sampling and improves the spatial invariance of the features. Finally, the fully connected layer integrates all features and outputs prediction results or classification decisions. The parameters of the machine learning algorithm are shown in Table 3.

Table 3. Parameters of machine learning algorithm.

Algorithm	Parameter	Value	Algorithm	Parameter	Value
BP	inputnum	2	RF	trees	100
	hiddennum	5		leaf	5
	outputnum	2		n_estimators	100
	net.trainParam.epochs	1000		max_features	3
	net.trainParam.lr	0.01		min_samples_split	2
	net.trainParam.goal	0.00001		min_samples_leaf	1
SVM	kernel	RBF	CNN	Kernel Size	3
	c	30		pool_size	2
	g	1		dropout	0.5
	epsilon	0.1		batch size	64
				learning rate	0.01

2.3.3. Accuracy Evaluation

In this study we used the coefficient of determination (R^2), Mean Absolute Error (MAE), root mean square error (RMSE), and normalized root mean square error (NRMSE) to assess model accuracy. The closer R^2 is to 1 and the closer the RMSE and NRMSE are to 0, the better the model. MAE is used to measure the average deviation between the predicted value and the real value of the model. The smaller the value, the more accurate the prediction of the model and the better the fitting effect.

MAE measures the average deviation between predicted and actual values, with a lower MAE indicating better performance [62]. Normalized Root Mean Square Error (NRMSE) is the normalized calculation of RMSE. The purpose is to eliminate the influence of dimensionality so that errors between data sets of different scales can be compared [63]. The smaller the NRMSE, the more accurate the model prediction, and the larger the NRMSE, the more significant the error. NRMSE does not have a fixed and unique formula. In practice, it depends on the choice of the normalized denominator. For example, normalization can be based on range, mean, or variance/standard deviation. The denominator is the difference between the maximum and minimum real values. The result is expressed as a percentage, and the range is [0, +∞).

The accuracy of NRMSE in describing the verification model is generally given a significantly different boundary. When NRMSE < 10%, there is almost no difference in the model; 10% ≤ NRMSE < 20%, the model difference is small; 20% ≤ NRMSE < 30%, the model difference is moderate; NRMSE ≥ 30%, the model is quite different.

The calculation formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}$$

n is the total number of samples, y_i and $\hat{y}_i$ are actual and predicted values, $\overline{y}_i$ is the average of the actual values, and $y_{\max}$ and $y_{\min}$ are the maximum and minimum of the measured values, respectively.

2.3.4. Population Density Data $PM_{2.5}$ Health-Risk Assessment

In this study, the integrated exposure-response (IER) model developed by Burnett et al. [64] was used to calculate the health effects of $PM_{2.5}$ concentrations in the urban belt around the Tarim Basin. The IER model was first proposed in the GBD2010 report, and has subsequently been widely used in related research on the health effects of $PM_{2.5}$ [65–67]. Four major fatal diseases (IHD, COPD, LC and STK) were selected as the health end-points. The relative risk (RR) was calculated as follows:

$$RR_{IER}(Z) = \begin{cases} 1, & Z \le Z_{cf} \\ 1 + \alpha\left\{1 - exp\left[-\gamma\left(Z - Z_{cf}\right)^{\delta}\right]\right\}, & Z > Z_{cf} \end{cases}$$

where Z is the $PM_{2.5}$ concentration, and Z_{cf} is the threshold corresponding to different health endpoints. When the concentration was less than or equal to the threshold, the RR was considered to be 1, and no additional health risks were generated. When the concentration exceeded the threshold, the RR increased. α, γ and δ are the exposure-response curve distribution parameters fitted for different health endpoints.

In this study, the attribution score was used to estimate the number of attributable deaths Mor_{ij} in southern Xinjiang grid cell i owing to specific disease j. The time-weighted normalization was used to calculate the annual disease mortality attributable to $PM_{2.5}$.

$$\begin{aligned} Mor_{ij} = y_j \times Pop_i \times AF_j &= y_i \times \tfrac{T_D}{T} \times Pop_i \times AF_{jD} + y_i \times \tfrac{T_N}{T} \times Pop_i \times AF_{jN} \\ &= y_i \times Pop_i \times \left[\tfrac{T_D}{T} \times \left(\tfrac{RR_{jD(Z_i)}-1}{RR_{jD(Z_i)}}\right) + \tfrac{T_N}{T} \times \left(\tfrac{RR_{jN(Z_i)}-1}{RR_{jN(Z_i)}}\right)\right] \end{aligned}$$

where y_j is the baseline mortality of specific disease j and derived from the results of Zhou et al. [68], where the unit of y_j is death rate per 100,000 people. Attributable mortality in 2015 and 2024 was calculated using the same baseline mortality, and the baseline mortality did not change over time, geography and population. Pop_i is the population of grid unit i; AF_j is the attributable fraction (AF) of specific disease j, which is used to measure the contribution of risk factors to disease or mortality; and $RR_{j(Zi)}$ is the RR of specific disease j in the grid with a $PM_{2.5}$ concentration of C_i. D and N denote the dust-prone season and non-dust season, respectively. T_D and T_N represent the number of days of the dust-prone season and non-dust season, respectively, and T is the total number of days of the year, $T = T_D + T_N$.

The parameters of the IER model and the mortality of specific diseases are shown in Table 4.

Since the calculated time period of the gridded relative risk in the dust-prone and non-dust seasons is only six months, this study adjusted the relative risk weight of each grid by normalization, so that the sum of the estimated values of the attributable deaths in the two periods is equal to the sum of the years. This ensured the accuracy of subsequent disease-attributable mortality estimates.

Table 4. Information on the parameters of the IER model and mortality of specific diseases [67–69].

Major Fatal Diseases	α	γ	δ	Z_{cf}	y_j
COPD	18.3	0.000932	0.682	7.17	0.000766
LC	159	0.000119	0.725	7.24	0.0005305
IHD	0.84	0.0724	0.544	6.96	0.001795
STK	1.01	0.0164	1.14	7.37	0.00251

The first-order error propagation [70] was used to calculate the uncertainties of health risk in this study, in which the uncertainty results are presented using the confidence factors (Cf_{out}) that span the 95% confidence interval (CI).

3. Results

3.1. Spatiotemporal Distribution of AOD in the Urban Belt Around the Tarim Basin

Considering the influence of the Taklamakan Desert, a climate characterized by low precipitation and high evaporation, and the basin topography, the Tarim Basin, located south of the Tianshan Mountains, is highly susceptible to sandstorms. Dust storms in the Tarim Basin generally occur from March to August annually; therefore, this study categorized the period from March to August as the dust-prone season and from September to February of the following year as the non-dust season. AOD represents the integral value of the extinction coefficient in the vertical direction of the atmosphere, which is closely related to the total concentration of aerosols in the vertical direction of the troposphere [71,72]. The AOD characterizes the degree of air pollution on the surface. We conducted spatial analysis of the average AOD values in the study area from 2015 to 2024, as illustrated in Figure 2.

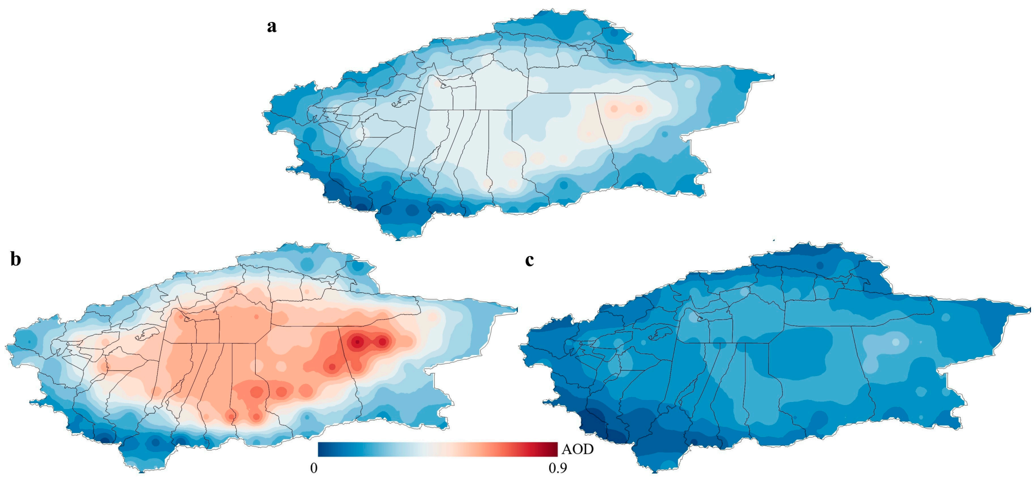

Figure 2. The average AOD value in the urban belt around the Tarim Basin from 2015 to 2024. (**a**) Average value of AOD in ten years; (**b**) average AOD value of dust-prone season over a decade; (**c**) average AOD value of non-dust season over a decade.

Over the previous decade, the average AOD values in the urban belt around the Tarim Basin exhibited a spatial distribution characteristic of low at the edge of the basin and high at the center. Furthermore, the average AOD values gradually increased from west to east within the basin, with higher values occurring in the southern (Minfeng and Yutian Counties in Hotan Prefecture) and eastern regions (Qiemo and Ruoqiang Counties in Bayingolin Mongolian Autonomous Prefecture). This is because the region south of the Tianshan Mountains in Xinjiang is influenced by prevailing westerly winds, and the Tarim Basin slopes from west to east, with a slight northward incline. The average AOD value during the dust-prone season was 0.447, which is significantly higher than the average AOD value of 0.171 during the non-dust season. Moreover, the average AOD values in various counties and cities during the dust-prone season were 2–3 times higher than those during the non-dust season, highlighting the significant impact of dust conditions on the urban belt around the Tarim Basin.

The counties and cities in the urban belt around the Tarim Basin, Alar City, Shaya County, Awati County, and Aksu City, had the highest average AOD values of 0.424, 0.416, 0.408, and 0.395, respectively. The average AOD in Tashkurgan Tajik Autonomous County was 0.135, which was the lowest among all the counties and cities. The spatial distribution of the average AOD during the dust-prone season and non-dust season over the previous decade was similar to the 10-year average AOD. The counties and cities with the highest average AOD values during the dust-prone season were Shaya County, Alar City, and Awati County, with AOD values of 0.615, 0.612, and 0.596, respectively. Tashkurgan Tajik Autonomous County had the lowest AOD in the urban belt around the Tarim Basin, with an AOD value of 0.202. Over the previous decade, the maximum AOD values during the dust-prone season and non-dust season in the Tarim Basin were 0.856 and 0.312, respectively, both of which occurred in Ruoqiang County. Maximum AOD values for the entire basin often occur in this region because Ruoqiang County is located east of the Tarim Basin and is influenced by westerly winds.

As shown in Figure 3, from 2015 to 2021, the average AOD value during the dust-prone season in the urban belt around the Tarim Basin fluctuated but showed an overall decreasing trend, with a significant decrease in 2021. However, after 2021, the average AOD value increased significantly. During the non-dust season, the average AOD remained relatively stable at approximately 0.17.

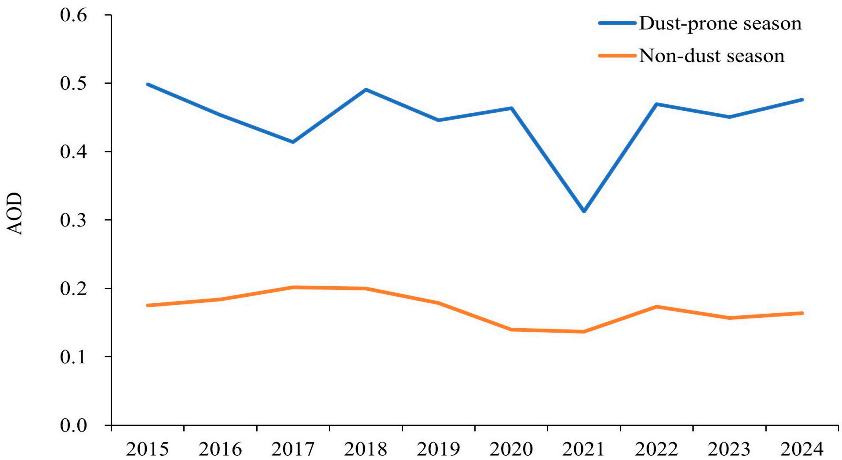

Figure 3. The temporal variation in AOD in the urban belt around the Tarim Basin during the dust-prone season and non-dust season from 2015 to 2024.

This study used the slope function to calculate AOD change trends in the study area. A positive slope value indicates an increasing trend in the AOD, and a negative value indicates a decreasing trend. To further analyze the statistical significance of the trend, the Mann–Kendall (MK) test was used. This method does not require the data to obey a specific distribution and is not sensitive to outliers. According to the absolute value of the standardized test statistic Z, when $|Z| > 1.645$, the trend reaches the 90% confidence level; when $|Z| > 1.96$, it reaches the 95% confidence level; and when $|Z| > 2.58$, it reaches the 99% confidence level [73,74].

As shown in Figure 4a, over the past decade, with the exception of Ruoqiang, Qiemo, Minfeng, and Yutian Counties, AOD levels in the urban belt around the Tarim Basin have generally shown a decreasing trend, indicating that dust conditions gradually decreased throughout the study area. The AOD change trend during the dust-prone season and non-dust season was consistent with the AOD changes over the previous decade, except for a significant increase at the border between Qiemo and Minfeng Counties during the dust-prone season. The Pishan-Hotan-Cele–Minfeng line in the southern Tarim Basin is a high-value area with frequent floating-dust events. Since the beginning of the 21st century, the floating frequency extreme value center has moved eastward to the Minfeng region, with a center value of 152 d/y [75]. As shown in Figure 5, the MK significance test was used to assess the long-term trend of AOD in the urban belt around the Tarim Basin in Xinjiang. The absolute value of the standardized test statistic Z was used to determine trend significance. The significance test results showed a zonal distribution feature that gradually decreased from south to north. The coverage area with $|Z| > 2.58$ accounted for over 80% of the total area, and the long-term variation trend of AOD in most regions of the study area passed the 99% confidence level significance test. The significant coverage area across different confidence levels showed a clear, decreasing gradient: the significant coverage area at the 99% confidence level > the significant coverage area at the 95% confidence level > the significant coverage area at the 90% confidence level. The Z statistics for the vast majority of regions were concentrated in the high-value range above 2.58, and only a few samples had Z values in the middle- and low-significance range of 1.64–2.58. The long-term change in AOD in the urban belt around the Tarim Basin is a global, highly robust process in a statistical sense. The core sand source area of the Taklamakan Desert is constantly affected by strong wind erosion, resulting in significant interannual fluctuations in dust aerosols. In recent years, the climate transition from warm-drying to warm-wetting in the region has further amplified interannual differences in dust weather events, thereby continuously strengthening the long-term trend signal in the AOD sequence. In recent decades, large-scale agricultural development and road traffic construction in the oasis area have continued to input anthropogenic aerosols into the atmosphere [76,77]. The annual increase in anthropogenic emissions and the annual fluctuations in natural dust have produced a superposition effect, making the monotonic trend characteristics of the AOD series more prominent and ultimately yielding a very high significance level in the MK test.

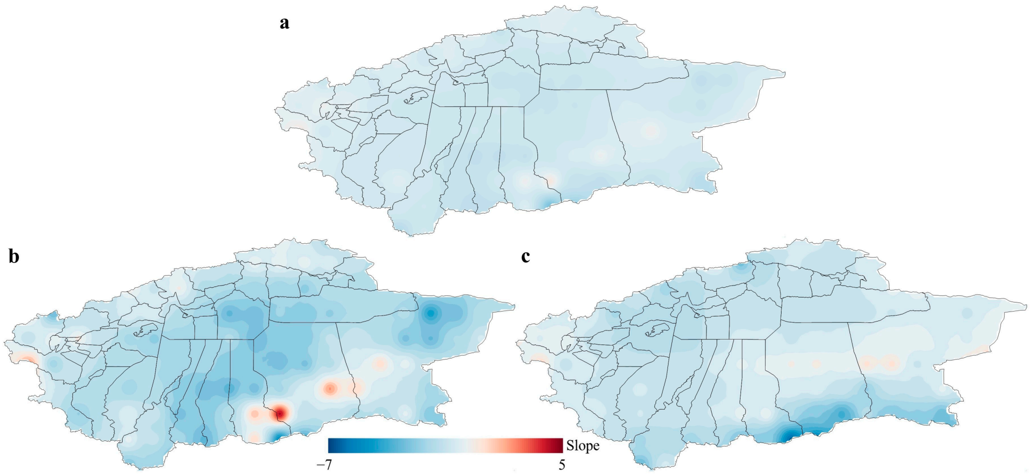

Figure 4. The trend of AOD change in the urban belt around the Tarim Basin from 2015 to 2024. (**a**) Trend of AOD change in ten years; (**b**) trend of AOD change in dust-prone season over a decade; (**c**) trend of AOD change in non-dust season over a decade.

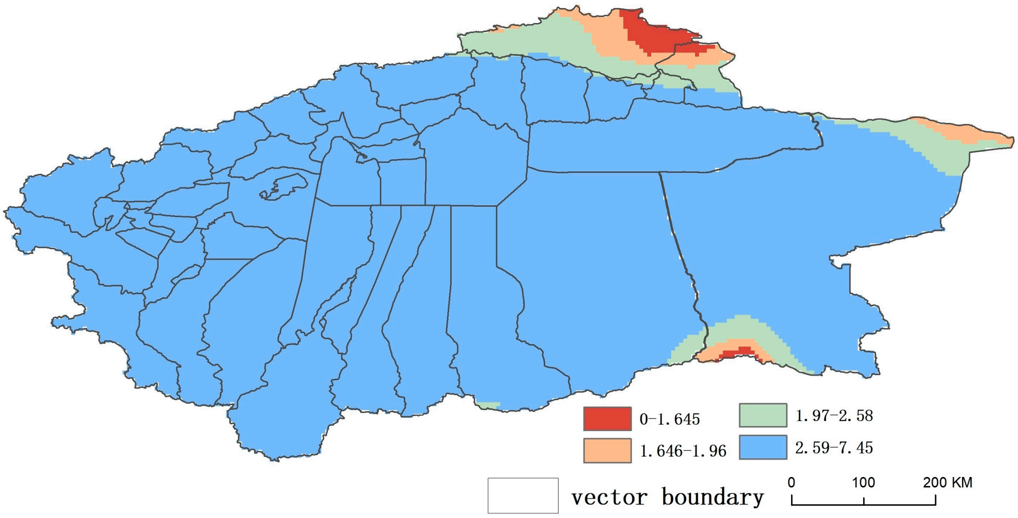

Figure 5. The results of the Mann–Kendall (MK) significance test.

The average AOD values of the urban belt around the Tarim Basin in 2015 and 2024 are shown in Figure 6. During the 2015 dust-prone season, the average AOD across the urban belt around the Tarim Basin was 0.499. Higher AOD values were concentrated in the Tarim Basin, particularly in the eastern and southern basins. The maximum AOD value of 0.953 was observed in Qiemo County. The AOD values at the edge of the basin were relatively low. Compared with the 2015 dust-prone season, the average AOD value decreased by 4.6% during the 2024 dust-prone season. Moreover, higher AOD values shifted eastward to Qiemo and Ruoqiang Counties in the eastern basin, in particular Ruoqiang County, where the maximum AOD reached 0.960.

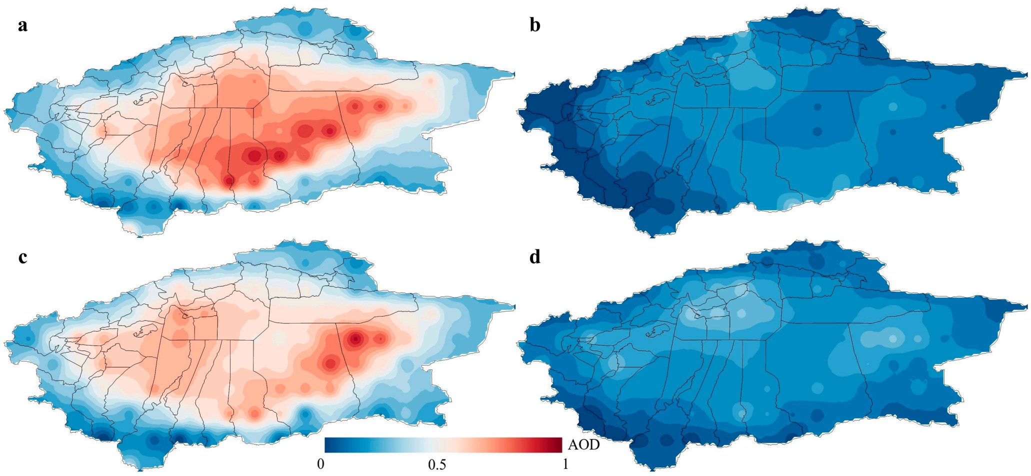

Figure 6. The average AOD values of the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

During the non-dust season, the difference in the average AOD values between 2015 and 2024 was relatively small, with values of 0.175 and 0.164, respectively. Higher AOD values were observed in the northern and southern basins in 2015. Higher values were observed in the northwestern and eastern basins in 2024. For example, from 2015 to 2024, the maximum AOD value during the non-dust season shifted from Minfeng County in Hotan Prefecture to Xinhe County in Aksu Prefecture.

3.2. Spatiotemporal Distribution of $PM_{2.5}$ Concentrations in the Urban Belt Around the Tarim Basin

In the construction of a $PM_{2.5}$-concentration estimation model, selection of appropriate estimation factors is an important prerequisite for its success. The selected estimation factors must have clear physical and chemical significance and have good application experience in construction of a $PM_{2.5}$ concentration model. Concurrently, to avoid the collinearity of variables between factors and reduce the risk of model over-fitting, the number of model estimation factors was controlled.

In this study, $PM_{2.5}$ measured data as a dependent variable, and AOD, air temperature, air humidity, surface wind speed, atmospheric pressure, and precipitation were selected as potential independent variables for the $PM_{2.5}$-concentration estimation model. The correlation between the above single independent variables and $PM_{2.5}$ concentration was analyzed. As shown in Figure 7, AOD, air temperature, and surface wind speed were positively correlated with $PM_{2.5}$ concentration, with correlation coefficients of 0.63, 0.51, and 0.69, respectively, suggesting a moderately strong positive correlation and indicating that the variables were highly aligned with the numerical change trend of $PM_{2.5}$ concentration. Air humidity was negatively correlated with $PM_{2.5}$ concentration, the variable with the strongest negative linear correlation among all variables. As air humidity increases, the $PM_{2.5}$ concentration shows a significant downward trend.

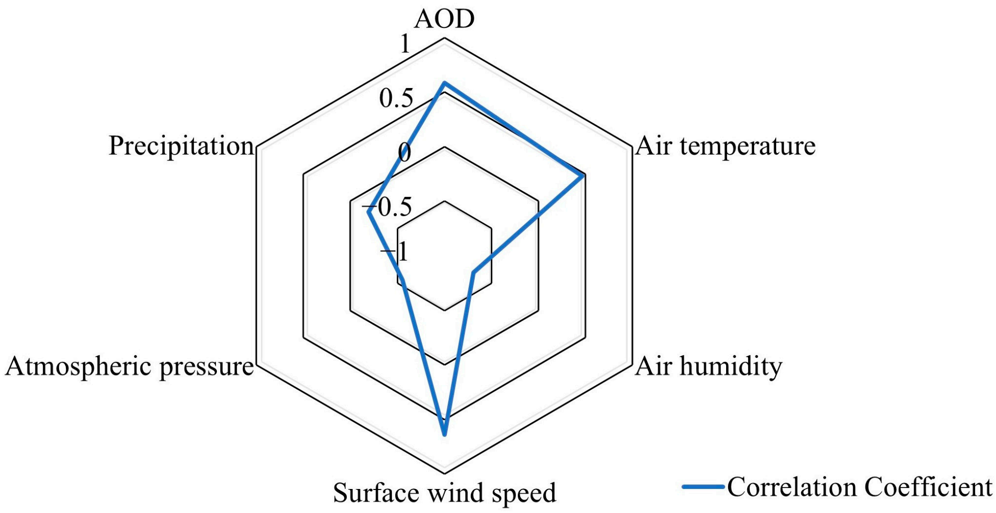

Figure 7. The correlation between meteorological factors and $PM_{2.5}$ concentration.

The influence of wind-speed change on $PM_{2.5}$ was more significant than that of air temperature; however, air temperature had a stronger indirect effect on $PM_{2.5}$ [78]. The impact of air pressure on $PM_{2.5}$, a complex nonlinear relationship, is an important meteorological factor affecting the $PM_{2.5}$ concentration [79]. When the pressure is low, the ascending motion of the atmosphere pressure is enhanced, which is conducive to pollutant diffusion, whereas high pressure can easily lead to pollutant accumulation. Precipitation has a scavenging effect on $PM_{2.5}$; however, the specific effect depends on the amount of precipitation and initial pollution status [80]. The correlation coefficient between precipitation and $PM_{2.5}$ concentration was only -0.17, indicating almost no correlation. The influence of numerical fluctuation on $PM_{2.5}$ concentration can be ignored. Therefore, in the $PM_{2.5}$ concentration model, the precipitation factor was excluded.

The experiment used a simple prediction model to minimize complexity, thereby improving the model's generalization and making the results more accurate [81]. To better evaluate applicability of the BP neural network, RF, SVM and CNN algorithms for simulating and predicting $PM_{2.5}$ concentrations in arid areas, this study used four indicators, the coefficient of determination (R^2), MAE, root mean square error (RMSE), and normalized root mean square error (NRMSE), to assess model accuracy.

For the selection of machine learning algorithms, this study prioritized higher R^2 values, followed by MAE, RMSE and NRMSE values. The verification results of the simulated and measured values for the three machine learning algorithms are listed in Table 5.

Table 5. Validation results of BP, RF, SVM and CNN algorithms.

Algorithms	R^2	MAE	RMSE	NRMSE (%)
BP	0.49 ± 0.04	29.2	36.5	15.2
RF	0.42 ± 0.04	29.5	37.3	15.8
SVM	0.70 ± 0.03	25.1	46.7	16.4
CNN	0.31 ± 0.05	47.4	54.5	83.0

Table 5 shows that the NRMSE of BP, RF and SVM do not significantly differ, all are in the range of 10–20%, indicating that the model difference is small. The MAE of SVM was the smallest among the four algorithms, and the $PM_{2.5}$ concentration magnitude, which served as the denominator, was relatively large. The MAE value of SVM was 25.1, and the RMSE value was 1.8 times that of MAE. There was a deviation between the predicted and true values for each sample, indicating the presence of outliers. The MAE of the CNN model showed the most significant deviation, and the model's stability was poor. During the modeling and validation process, a 5-fold cross-validation method was employed. The training set is randomly and evenly divided into 5 subsets, with 4 subsets selected for model training each time, and the remaining 1 subset used to calculate RMSE and NRMSE. This method can guarantee that each sample is not only a modeling set, but also a validation set. The average of the five results is taken as the final verification value, which effectively reduces the contingency of the results caused by a single data splitting. The SVM model has the highest R^2 in the $PM_{2.5}$ prediction model, but it also has a relatively high RMSE. The classical cross-validation method will underestimate the prediction error in the spatial scenario; the spatial dependence can also lead to overoptimistic estimates of accuracy [82]. In the prediction of spatial data, especially when considering complex factors such as environment and ecology at a large spatial scale, pessimistic bias may arise in cases of randomness, clustering, or uneven distribution of sampling points, resulting in overestimated errors and a larger RMSE. Dust weather can cause $PM_{2.5}$ concentrations to surge from low levels within a few hours; nonetheless, the sudden, extremely high values are outliers. For complex issues related to environmental and ecological nonlinearity, the SVM model's R^2 was 0.7, which fell in the middle-to-upper range, and its trend-capture ability was acceptable.

However, the R^2 of SVM is significantly higher than that of BP, RF and CNN, indicating that the accuracy of SVM simulation results is higher and that SVM is more suitable for arid areas than BP, RF and CNN. Therefore, this study used an SVM machine learning algorithm to simulate $PM_{2.5}$ concentrations in the urban belt around the Tarim Basin. The spatial distribution of the simulation results is shown in Figure 8.

Owing to the positive correlation between $PM_{2.5}$ concentration and AOD value, the spatial distribution of $PM_{2.5}$ concentrations during the dust-prone season and non-dust season in the urban belt around the Tarim Basin was similar to those of the AOD. As shown in Figure 8, higher $PM_{2.5}$ concentrations during the 2015 dust-prone season occurred in Minfeng, Yutian, Qiemo, and Ruoqiang Counties in the southern and eastern Tarim Basin. The maximum $PM_{2.5}$ concentration in these four counties exceeded 100 μg/m^3. By 2024, the counties with maximum $PM_{2.5}$ concentrations exceeding 100 μg/m^3 during the dust-prone season included Ruoqiang, Qiemo, and Minfeng Counties, with higher $PM_{2.5}$ concentrations shifting eastward. From the county and city perspective, Shaya, Luopu, Yutian, and Minfeng Counties had the highest $PM_{2.5}$ concentrations among all counties and cities during the 2015 dust-prone season, with $PM_{2.5}$ concentrations of 90.7 μg/m^3, 90.4 μg/m^3, 89.7 μg/m^3, and 89.1 μg/m^3, respectively. During the 2024 dust-prone season, the top four in terms of $PM_{2.5}$ concentration were replaced by Awati County, Alar City, Moyu County, and Aksu City, with $PM_{2.5}$ concentrations of 89.7 μg/m^3, 88.8 μg/m^3, 86.0 μg/m^3, and 84.1 μg/m^3, respectively.

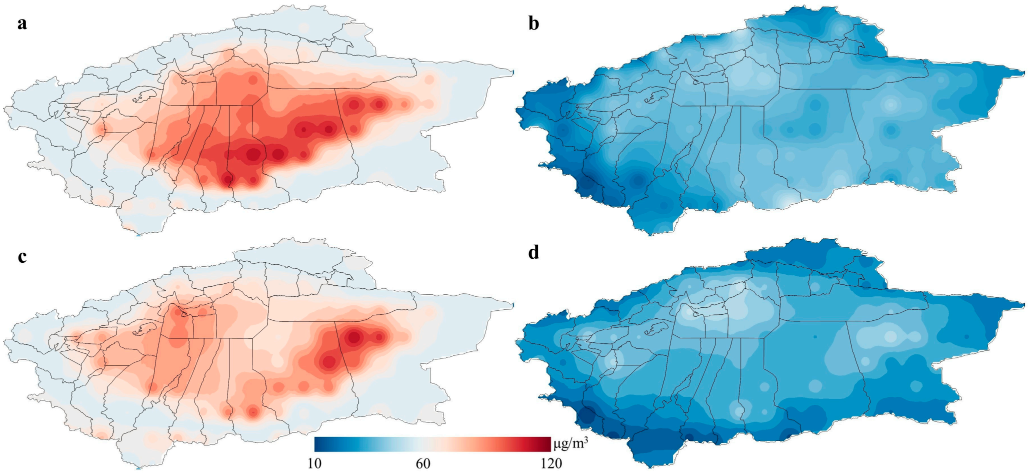

Figure 8. $PM_{2.5}$ concentration in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

The $PM_{2.5}$ values in the urban belt around the Tarim Basin during the non-dust season were approximately half of those during the dust-prone season. In the 2015 and 2024 non-dust season, Xinhe and Shaya Counties and Alar and Aksu Cities had the highest $PM_{2.5}$ concentrations, with respective values of 42.3 μg/m^3, 40.3 μg/m^3, 39.4 μg/m^3, and 38.0 μg/m^3 in 2015, and 45.9 μg/m^3, 42.0 μg/m^3, 47.0 μg/m^3, and 42.2 μg/m^3 in 2024. The $PM_{2.5}$ concentrations in other counties and cities during these two periods were respectively between 18.1–37.9 μg/m^3 and 20.0–41.6 μg/m^3.

3.3. Health Risks Attributed to $PM_{2.5}$ Comparison Between 2015 and 2024

In this study, four major fatal diseases, IHD, COPD, LC, and STK, were selected as health endpoints of the urban belt around the Tarim Basin. The spatial distribution of deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin is shown in Figure 9. As the number of deaths attributed to $PM_{2.5}$ is the superposition of population and RR, its spatial distribution is similar to that of the $PM_{2.5}$ concentration; however, there were no attributed deaths in the uninhabited desert and Gobi regions. Therefore, the spatial characteristics exhibited a distribution pattern around the basin and were concentrated in the oases. The number of deaths attributed to $PM_{2.5}$ was 21,363 (95% CI: 16,169–28,226) and 15,201 (95% CI: 11,505–20,084) during the dust-prone season and non-dust season in 2015, respectively. However, in 2024, they were 25,671 (95% CI: 19,429–33,918) and 18,800 (95% CI: 14,229–24,840), respectively, representing increases of 20.2% and 23.7% compared to those of 2015. Compared to the non-dust season, attributable deaths during the 2015 and 2024 dust-prone season increased by 40.5% and 36.5%, respectively, highlighting the significant impact of $PM_{2.5}$ during dust-prone season on human health.

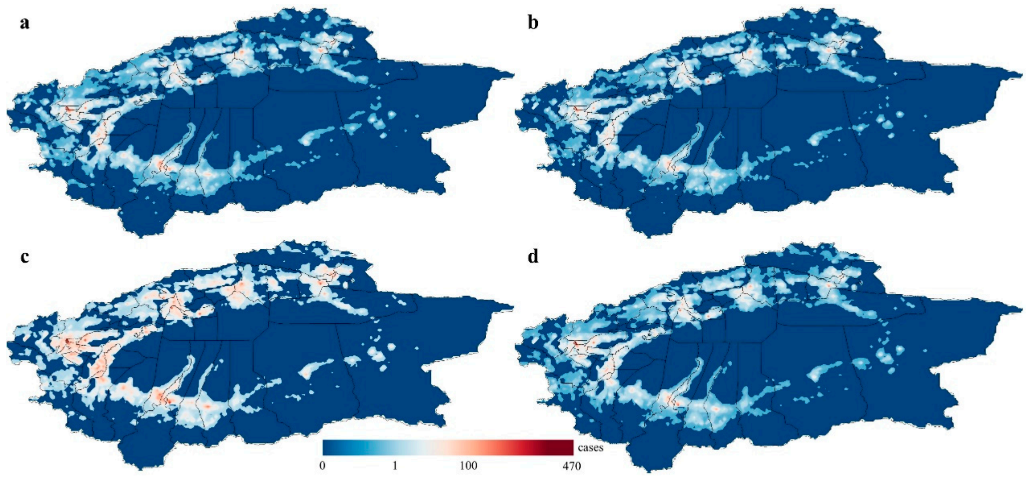

Figure 9. The spatial distribution of deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

From the perspective of administrative divisions, the counties and cities with higher health risks attributed to $PM_{2.5}$ in 2015 and 2024 include Kuche City, and Shule, Jiashi, Shache, and Moyu Counties. In 2015, the attributable deaths in Kuche City and Shule and Shache Counties during the dust-prone season and non-dust season were 1092 (95% CI: 826–1443) and 867 (95% CI: 656–1146), 1221 (95% CI: 924–1613) and 835 (95% CI: 632–1103), and 1587 (95% CI: 1201–2097) and 1145 (95% CI: 867–1513), respectively. The attributable deaths in the three cities during the dust-prone season were 26.0%, 46.2%, and 38.6% higher than those during the non-dust season, respectively. In the 2024 non-dust season, the attributable deaths in Shule, Jiashi, Shache, and Moyu Counties were 1008 (95% CI: 763–1332), 1007 (95% CI: 762–1331), 1369 (95% CI: 1036–1809), and 883 (95% CI: 668–1167), respectively. The dust-prone season deaths were 32.4%, 30.5%, 29.0%, and 53.3% higher than those in the non-dust season.

Figures 10–13 show the spatial distributions of IHD, COPD, LC, and STK deaths attributable to $PM_{2.5}$ in the urban belt around the Tarim Basin, respectively. Among the four major diseases, STK and IHD were predominant in both the dust-prone season and non-dust season, accounting for >80% of the total deaths. In the 2024 dust-prone season, the attributable deaths to STK, IHD, COPD, and LC were 14,741 (95% CI: 11,867–18,310), 6811 (95% CI: 5133–9038), 2214 (95% CI: 1613–3038) and 1903 (95% CI: 1416–2557), respectively, accounting for 57.4%, 26.5%, 8.63%, and 7.41% of the total deaths at the four health endpoints. Population is an important factor affecting the number of attributable deaths. Shache County is the most populous county in the urban belt around the Tarim Basin. In 2015 and 2024, attributable deaths from the four major diseases in the dust-prone season and non-dust season accounted for 7.43%, 7.53%, 6.88%, and 7.28% of the total, respectively, ranking highest in the urban belt around the Tarim Basin. During the 2015 dust-prone season, the attributable deaths from IHD, COPD, LC and STK in Shache County were 421 (95% CI: 317–559), 140 (95% CI: 102–192), 120 (95% CI: 89–161), and 905 (95% CI: 729–1124), respectively. $PM_{2.5}$ concentrations during the dust-prone season in Shache County in 2024 increased by 3.87% compared to that in 2015, and the population increased by 10.5%. In the context of the dual increase in $PM_{2.5}$ concentrations and population, the number of attributable deaths during the 2024 dust-prone season increased to 468 (95% CI: 353–621), 157 (95% CI: 114–215), 135 (95% CI: 100–181), and 1006 (95% CI: 810–1250), respectively, with an average increase of 11.5%.

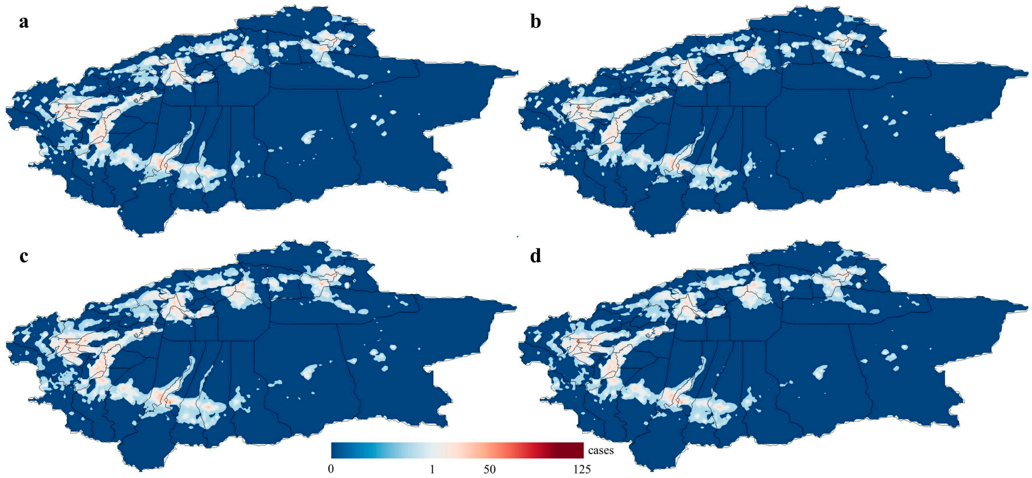

Figure 10. The spatial distribution of IHD deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

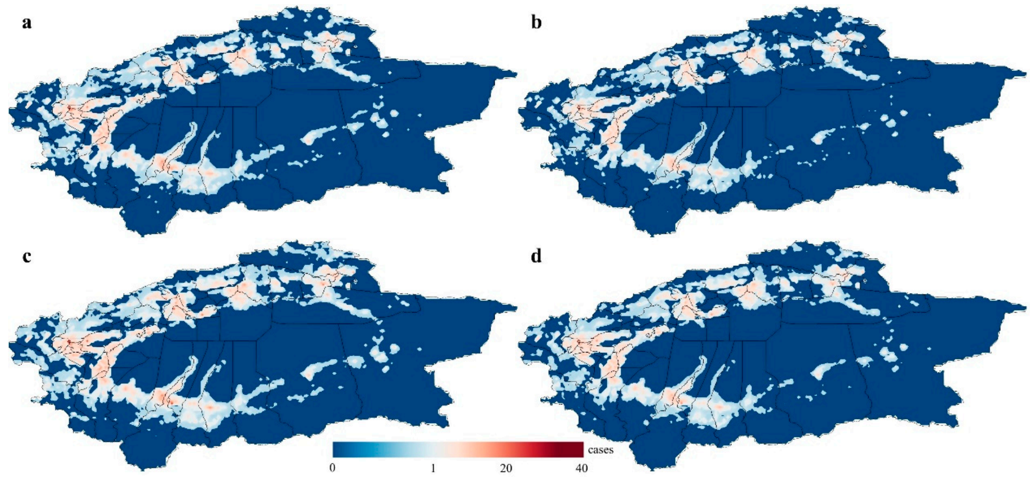

Figure 11. The spatial distribution of COPD deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

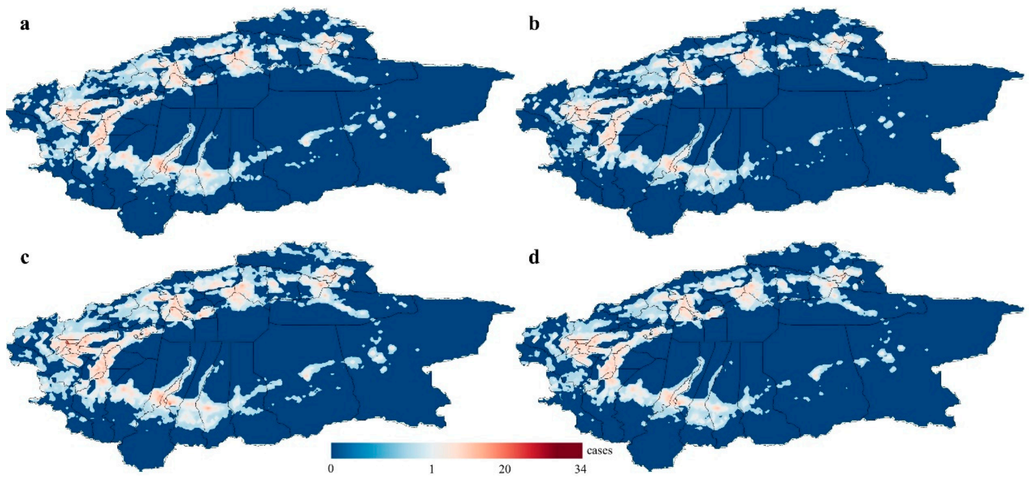

Figure 12. The spatial distribution of LC deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

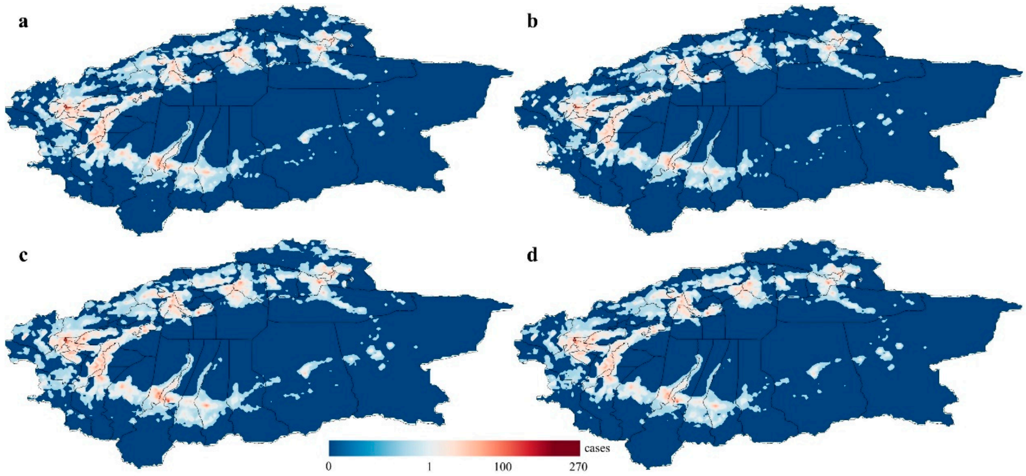

Figure 13. The spatial distribution of STK deaths attributed to $PM_{2.5}$ in the urban belt around the Tarim Basin. (**a**) Dust-prone season in 2015; (**b**) non-dust season in 2015; (**c**) dust-prone season in 2024; (**d**) non-dust season in 2024.

4. Discussion

4.1. Machine Learning Algorithm Uncertainty Analysis in $PM_{2.5}$ Concentration Simulations

Remote-sensing technology can overcome the shortcomings of fixed monitoring sites and limited scale. The prominent advantages of machine learning algorithms in handling complex and nonlinear problems have proven to be applicable for simulating $PM_{2.5}$ concentrations. However, some factors lead to uncertainty in machine learning algorithmic $PM_{2.5}$ concentration simulation results.

First is the selection of the machine learning algorithm. Numerous machine learning algorithms are available for simulation; however, different methods exhibit certain differences in practical research, and applicability of the algorithm to the sample size must also be considered. For example, the SVM method can yield more accurate results than other algorithms for small-sample training sets [83]. Second, the uncertainty of the remote-sensing data itself is one of the sources affecting the uncertainty of $PM_{2.5}$ simulation results. High-precision remote-sensing data can ensure the quality of simulation results. The source and accuracy of meteorological factor datasets, such as air temperature, air humidity, precipitation, and surface wind speed, which are closely related to $PM_{2.5}$, also cause uncertainty in the simulation results. Despite the findings, our study has objective limitations. Although machine learning algorithms often employ k-fold cross-validation for model validation, it may yield overly optimistic error estimates when the data exhibit spatial or temporal autocorrelation [84]. Hence, this method cannot fully test the transferability of space or time. Finally, the training set is a crucial component for constructing machine learning models. Given the limited training set coverage, the simulation results tended to perform well within the scope of the training set; however, the uncertainty increased in areas not covered by the training set. Therefore, the range of values in the training set can significantly affect the final simulation accuracy.

4.2. Uncertainty Analysis of Health Risk Caused by $PM_{2.5}$

The health risks caused by $PM_{2.5}$ are influenced by several factors. Uncertainty of $PM_{2.5}$ concentration data input by the model can affect IER model health-risk results. $PM_{2.5}$ is a mixture of coal-fired sulfate, traffic-derived organic carbon, and secondary nitrate, and its unit concentration toxicity may vary by several times. However, in health-risk assessments, the total mass concentration is often regarded as a “homogeneous toxicity,” which introduces a substantial chemical composition heterogeneity bias. The parameters of the

IER model were derived from epidemiological data, and the use of limited epidemiological data to estimate the model parameters can introduce parameter uncertainty. In addition, the lack of localized parameters in the model and population differences between the study area and other regions can also lead to uncertainty in the results. After exposure to $PM_{2.5}$, the health damage to the population is not immediately apparent; it only becomes apparent after a delay [85]. Therefore, there may be some uncertainty in the mortality of various diseases calculated by the IER model for 2015 and 2024.

4.3. The Significance of High-Resolution $PM_{2.5}$ Products for Grassroots Public Health Management

The significant impact of $PM_{2.5}$ on human health is widely recognized, and estimating health risks requires accurate $PM_{2.5}$ concentration data, along with appropriate local parameters [86,87]. How to obtain high-precision $PM_{2.5}$ concentration data remains an urgent problem. The urban belt area around the Tarim Basin is extensive, and the oasis settlements are scattered, which conflicts with the small number and sparse distribution of monitoring stations. National monitoring stations are often located in urban centers and cannot capture the $PM_{2.5}$ concentration distribution in suburban and rural areas. In future work, the results of this study can help grasp the overall distribution of $PM_{2.5}$ concentrations in the urban belt around the Tarim Basin and help grassroots disease control institutions conduct health risk assessment and policy formulation.

5. Conclusions

The Tarim Basin is located in the Taklamakan Desert in China and is significantly affected by sandstorms. Owing to the influence of the prevailing westerly winds and the terrain of the Tarim Basin, the spatial distribution of AOD showed an increasing pattern from west to east. Except for individual regions, the AOD showed a decreasing trend from 2015 to 2024 in the urban belt around the Tarim Basin.

Based on the well-performing SVM simulation of $PM_{2.5}$ concentrations, the spatial distribution of $PM_{2.5}$ concentrations during the dust-prone season and non-dust season in the urban belt around the Tarim Basin was consistent with the AOD, with $PM_{2.5}$ values during the non-dust season being significantly lower than those during the dust-prone season. Higher $PM_{2.5}$ concentrations during the dust-prone season occurred in Minfeng, Qiemo, and Ruoqiang Counties in the southern and eastern Tarim Basin, showing an eastward-shifting trend.

The mortality attributed to $PM_{2.5}$ was estimated in this study. Through the comparative analysis between 2015 and 2024, the results showed the impact of $PM_{2.5}$ on human health was significant during the dust-prone season, with an increase of over 36% in attributable deaths compared to the non-dust season. Among the four major fatal diseases, namely IHD, COPD, LC and STK, STK and IHD played a dominant role, accounting for >80% of the total attributable deaths. Given the characteristics of IHD and STK, $PM_{2.5}$ concentrations and health risks typically vary substantially with age, and the age-specificity of these diseases can be explored in depth in subsequent studies.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/su18168373/s1, Text S1: Adaptability evaluation of MYD data; Table S1: Statistical comparison of MYD08 and MCD19A2 data in typical months of dust-prone season and non-dust season.

Author Contributions: Data curation, L.G. and Q.S.; Methodology, L.G., Q.S. and G.G.; Software, L.G. and Q.S.; Writing—original draft, L.G.; Supervision, L.G. and Q.S.; Visualization, Y.Z. and L.C.; Funding acquisition, L.G. and Q.S.; Writing—review and editing, L.G. and Q.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grants 42567015 and 42467012).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in MODIS at https://www.earthdata.nasa.gov/ (accessed on 16 June 2025); the $PM_{2.5}$ concentration data presented in this study are openly available in the China Environmental Monitoring Station National Urban Air Quality Real-Time Publishing Platform at https://www.cnemc.cn/sssj/ (accessed on 30 June 2025); the population density data presented in this study are openly available in the WorldPop R2025A v1 dataset at 10.5258/SOTON/WP00845 (accessed on 15 March 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Southerland, V.; Brauer, M.; Mohegh, A.; Hammer, M.; Van Donkelaar, A.; Martin, R.V.; Apte, J.; Anenberg, S. Global urban temporal trends in fine particulate matter (PM2.5) and attributable health burdens: Estimates from global datasets. *Lancet Planet. Health* **2022**, *6*, e139–e146. [CrossRef] [PubMed]
2. Bowe, B.; Xie, Y.; Li, T.; Yan, Y.; Xian, H.; Al-Aly, Z. The 2016 global and national burden of diabetes mellitus attributable to PM2.5 air pollution. *Lancet Planet. Health* **2018**, *2*, e301–e312. [CrossRef] [PubMed]
3. Handschuh, J.; Erbertseder, T.; Baier, F. On the added value of satellite AOD for the investigation of ground-level PM2.5 variability. *Atmos. Environ.* **2024**, *331*, 120601. [CrossRef]
4. Yang, Z.; Zdanski, C.; Farkas, D.; Bang, J.; Williams, H. Evaluation of Aerosol Optical Depth (AOD) and PM2.5 associations for air quality assessment. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100396. [CrossRef]
5. Guo, J.; Xia, F.; Zhang, Y.; Liu, H.; Li, J.; Lou, M.; He, J.; Yan, Y.; Wang, F.; Min, M.; et al. Impact of diurnal variability and meteorological factors on the PM2.5-AOD relationship: Implications for PM2.5 remote sensing. *Environ. Pollut.* **2017**, *221*, 94–104. [CrossRef] [PubMed]
6. Wang, X.; Qin, L.; Wang, D.; Wang, B. Study on the correlation between aerodynamic thickness of airborne aerosol and PM2.5 in Chongqing Region. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2018; Volume 108, p. 042001.
7. Shao, P.; Xin, J.; An, J.; Kong, L.; Wang, B.; Wang, J.; Wang, Y.; Wu, D. The empirical relationship between PM2.5 and AOD in Nanjing of the Yangtze River Delta. *Atmos. Pollut. Res.* **2017**, *8*, 233–243. [CrossRef]
8. Yang, Q.; Yuan, Q.; Yue, L.; Li, T.; Shen, H.; Zhang, L. The relationships between PM2.5 and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. *Environ. Pollut.* **2019**, *248*, 526–535. [CrossRef] [PubMed]
9. Cordero, J. A two-stage algorithm to estimate ground-level PM2.5 concentration levels in Madrid (Spain) from AOD satellite data and surface proxies. *Atmos. Pollut. Res.* **2025**, *16*, 102678. [CrossRef]
10. Chen, Q.; Shao, K.; Zhang, S. Enhanced PM2.5 estimation across China: An AOD-independent two-stage approach incorporating improved spatiotemporal heterogeneity representations. *J. Environ. Manag.* **2024**, *368*, 122107. [CrossRef] [PubMed]
11. Wang, J.; Li, X. Influence of industrial sustainability transition on air quality in a typical resource-exhausted city. *Heliyon* **2024**, *10*, e25138. [CrossRef] [PubMed]
12. Levy, R.; Mattoo, S.; Sawyer, V.; Shi, Y.; Colarco, P.; Lyapustin, A.; Wang, Y.; Remer, L. Exploring systematic offsets between aerosol products from the two MODIS sensors. *Atmos. Meas. Tech.* **2018**, *11*, 4073–4092. [CrossRef] [PubMed]
13. Jeong, M.; Hsu, N.; Sayer, A.; Bettenhausen, C. Validation and Uncertainty Estimates for MODIS Collection 6 "Deep Blue" Aerosol Data. *J. Geophys. Res. Atmos.* **2013**, *118*, 7864–7872. [CrossRef]
14. Adams, M.; Massey, F.; Chastko, K.; Cupini, C. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. *Atmos. Environ.* **2020**, *230*, 117479. [CrossRef]
15. Yang, J.; Yan, R.; Nong, M.; Liao, J.; Li, F.; Sun, W. PM2.5 concentrations forecasting in Beijing through deep learning with different inputs, model structures and forecast time. *Atmos. Pollut. Res.* **2021**, *12*, 101168. [CrossRef]
16. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [CrossRef]
17. Chen, G.; Li, S.; Knibbs, L.; Hamm, N.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M.; Guo, Y. A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52–60. [CrossRef] [PubMed]

18. Sukprasert, S.; Thanasutives, P.; Araki, S.; Uranishi, K.; Matsuo, T.; Shimadera, H. A comparative study on CMAQ-based predictors and reanalysis AOD for improving PM2.5 estimation in Thailand using machine learning. *Atmos. Environ.* **2026**, *375*, 122006. [CrossRef]
19. Mallet, M. Meteorological normalisation of PM10 using machine learning reveals distinct increases of nearby source emissions in the Australian mining town of Moranbah. *Atmos. Pollut. Res.* **2021**, *12*, 23–35. [CrossRef] [PubMed]
20. Analitis, A.; Barratt, B.; Green, D.; Beddows, A.; Samoli, E.; Schwartz, J.; Katsouyanni, K. Prediction of PM2.5 concentrations at the locations of monitoring sites measuring PM10 and NOX, using generalized additive models and machine learning methods: A case study in London. *Atmos. Environ.* **2020**, *240*, 117757. [CrossRef]
21. Stafoggia, M.; Bellander, T.; Bucci, S.; Davoli, M.; De Hoogh, K.; De' Donato, F.; Gariazzo, C.; Lyapustin, A.; Michelozzi, P.; Renzi, M.; et al. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170–179. [CrossRef] [PubMed]
22. Bagheri, H. A machine learning-based framework for high resolution mapping of PM2.5 in Tehran, Iran, using MAIAC AOD data. *Adv. Space Res.* **2022**, *69*, 3333–3349. [CrossRef]
23. Hu, X.; Belle, J.; Meng, X.; Wildani, A.; Waller, L.; Strickland, M.; Liu, Y. Estimating PM2.5 concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol.* **2017**, *51*, 6936–6944. [CrossRef] [PubMed]
24. Amiri, Z.; Shahne, M. Modeling PM2.5 concentration in Tehran using satellite-based Aerosol optical depth (AOD) and machine learning: Assessing input contributions and prediction accuracy. *Remote Sens. Appl. Soc. Environ.* **2025**, *38*, 101549. [CrossRef]
25. Chen, J.; Yin, J.; Zang, L.; Zhang, T.; Zhao, M. Stacking machine learning model for estimating hourly PM2.5 in China based on Himawari 8 aerosol optical depth data. *Sci. Total Environ.* **2019**, *697*, 134021. [CrossRef] [PubMed]
26. Crouse, D.; Peters, P.; Donkelaar, A.; Goldberg, M.; Villeneuve, P.; Brion, O.; Khan, S.; Atari, D.; Jerrett, M.; Pope, C.; et al. Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: A Canadian national-level cohort study. *Environ. Health Perspect.* **2012**, *120*, 708. [CrossRef] [PubMed]
27. Sousa, S.; Pires, J.; Martins, E.; Fortes, J.; Alvim-Ferraz, M.; Martins, F. Short-term effects of air pollution on respiratory morbidity at Rio de Janeiro—Part II: Health assessment. *Environ. Int.* **2012**, *43*, 1–5. [CrossRef] [PubMed]
28. Orellano, P.; Reynoso, J.; Quaranta, N.; Bardach, A.; Ciapponi, A. Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis. *Environ. Int.* **2020**, *142*, 105876. [CrossRef] [PubMed]
29. Bowe, B.; Xie, Y.; Yan, Y.; Al-Aly, Z. Burden of cause-specific mortality associated with PM2.5 air pollution in the United States. *JAMA Netw. Open* **2019**, *2*, e1915834. [CrossRef] [PubMed]
30. Kim, D.; Song, M.; Yuk, J.; Park, J.; Choi, J.; Lee, K. Immunosuppression induced by co-exposure to allergens and PM2.5 increases the risk of lung disease upon viral protein challenge. *Environ. Int.* **2025**, *203*, 109722. [CrossRef] [PubMed]
31. Talbott, E.; Rager, J.; Benson, S.; Brink, L.; Bilonick, R.; Wu, C. A case-crossover analysis of the impact of PM2.5 on cardiovascular disease hospitalizations for selected CDC tracking states. *Environ. Res.* **2014**, *134*, 455–465. [CrossRef] [PubMed]
32. Kloog, I.; Ridgway, B.; Koutrakis, P.; Coull, B.; Schwartz, J. Long- and short-term exposure to PM2.5 and mortality: Using novel exposure models. *Epidemiology* **2013**, *24*, 555–561. [PubMed]
33. Cohen, A.; Brauer, M.; Burnett, R.; Anderson, H.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **2017**, *389*, 1907–1918. [CrossRef] [PubMed]
34. Han, C.; Kim, S.; Lim, Y.; Bae, H.; Hong, Y. Spatial and temporal trends of number of deaths attributable to ambient PM2.5 in the Korea. *J. Korean Med. Sci.* **2018**, *33*, e193. [CrossRef] [PubMed]
35. Sukchinda, S.; Meesang, W.; Saneha, S. Linking Changes in Meteorology and Particulate Matter (PM10 and PM2.5) in Ambient air to Respiratory Disease Hospitalizations in Urban Areas. *Case Stud. Chem. Environ. Eng.* **2026**, *13*, 101399. [CrossRef]
36. Shin, H.; Lee, H.; Song, J.; Park, S.; Kim, J.; Park, S.; Kim, H.; Chung, J.; Kim, K.; Park, S. Short-term PM2.5 exposure among 3-year cancer survivors with incident cardiovascular disease before and during the COVID-19 pandemic period. *Atmos. Pollut. Res.* **2025**, *16*, 102574. [CrossRef]
37. Joshi, D.; Negi, P.; Devi, S.; Lohani, H.; Kumar, R.; Gupta, M.; Ming, L. Fine particulate matter (PM2.5, PM10): A silent catalyst for chronic lung diseases in India; a comprehensive review. *Environ. Chall.* **2025**, *20*, 101215. [CrossRef]
38. Chen, L.; Shi, M.; Gao, S.; Li, S.; Mao, J.; Zhang, H.; Sun, Y.; Bai, Z.; Wang, Z. Assessment of population exposure to PM2.5 for mortality in China and its public health benefit based on BenMAP. *Environ. Pollut.* **2017**, *221*, 311–317. [CrossRef] [PubMed]
39. Li, G.; Xue, M.; Zeng, Q.; Cai, Y.; Pan, X.; Meng, Q. Association between fine ambient particulate matter and daily total mortality: An analysis from 160 communities of China. *Sci. Total Environ.* **2017**, *599*, 108–113. [CrossRef] [PubMed]
40. Li, J.; Liu, H.; Lv, Z.; Zhao, R.; Deng, F.; Wang, C.; Qin, A.; Yang, X. Estimation of PM2.5 mortality burden in China with new exposure estimation and local concentration-response function. *Environ. Pollut.* **2018**, *243*, 1710–1718. [CrossRef] [PubMed]
41. Liu, H.; Tian, Y.; Cao, Y.; Song, J.; Huang, C.; Xiang, X.; Li, M.; Hu, Y. Fine particulate air pollution and hospital admissions and readmissions for acute myocardial infarction in 26 Chinese cities. *Chemosphere* **2018**, *192*, 282–288. [CrossRef] [PubMed]

42. Liu, C.; Chen, R.; Sera, F.; Vicedo-Cabrera, A.; Guo, Y.; Tong, S.; Coelho, M.; Saldiva, P.; Lavigne, E.; Matus, P.; et al. Ambient particulate air pollution and daily mortality in 652 cities. *N. Engl. J. Med.* **2019**, *381*, 705–715. [CrossRef] [PubMed]
43. Kong, S.; Zhang, K.; Sang, J.; Zhang, J.; Kan, C.; Sun, X.; Hou, N. Micro- and Nanoplastics and PM2.5 in Cardiovascular Disease: Emerging Mechanisms, Impacts, and Therapeutic Insights. *Food Chem. Toxicol.* **2026**, *211*, 115982. [CrossRef] [PubMed]
44. Li, D.; Wang, J.; Zhang, Z.; Shen, P.; Zheng, P.; Jin, M.; Lu, H.; Lin, H.; Chen, K. Association between short-term exposure to ambient air pollution and daily mortality: A time-series study in Eastern China. *Environ. Sci. Pollut. Res.* **2018**, *25*, 16135–16143. [CrossRef] [PubMed]
45. Hu, J.; Huang, L.; Chen, M.; Liao, H.; Zhang, H.; Wang, S.; Zhang, Q.; Ying, Q. Premature mortality attributable to particulate matter in China: Source contributions and responses to reductions. *Environ. Sci. Technol.* **2017**, *51*, 9950–9959. [CrossRef] [PubMed]
46. Zou, B.; You, J.; Lin, Y.; Duan, X.; Zhao, X.; Fang, X.; Campen, M.; Li, S. Air pollution intervention and life-saving effect in China. *Environ. Int.* **2019**, *125*, 529–541. [CrossRef] [PubMed]
47. Li, X. Improved Understanding of Trends, Variations, and Causes of Atmospheric Aerosols Using Ground Measurements, Satellite Observations, and Atmospheric Chemistry Modeling. Ph.D. Thesis, University of Connecticut, Storrs, CT, USA, 2019.
48. Xu, Q.; Chen, X.; Rupakheti, D.; Dong, J.; Tang, L.; Kang, S. Evaluation and Comparison of Spatio-Temporal Relationship between Multiple Satel-lite Aerosol Optical Depth (AOD) and Near-Surface PM2.5 Concentration over China. *Remote Sens.* **2022**, *14*, 5841. [CrossRef]
49. Lyapustin, A.; Wang, Y. *MODIS Multi-Angle Implementation of Atmospheric Correction (MAIAC) Data User's Guide*; NASA: Greenbelt, MD, USA, 2018; Volume 6, pp. 1–19.
50. Rodriguez-Villamizar, L.; Rojas, Y.; Grisales, S.; Mangones, S.; Cáceres, J.; Agudelo-Castañeda, D.; Herrera, V.; Marín, D.; Jiménez, J.; Belalcázar-Ceron, L.; et al. Intra-urban variability of long-term exposure to PM2.5 and NO2 in five cities in Colombia. *Environ. Sci. Pollut. Res.* **2024**, *31*, 3207–3221. [CrossRef] [PubMed]
51. Tiwari, K.; Chatterjee, C.; Gontia, N.; Jena, S. Remote Sensing & GIS Applications [Course Content]. Agrimoon. Available online: https://agrimoon.com/wp-content/uploads/Remote-Sensing-and-GIS-Application.pdf (accessed on 7 August 2026).
52. Bondarenko, M.; Priyatikanto, R.; Tejedor-Garavito, N.; Zhang, W.; McKeen, T.; Cunningham, A.; Woods, T.; Hilton, J.; Cihan, D.; Nosatiuk, B.; et al. *The Spatial Distribution of Population in 2015–2030 at a Resolution of 30 Arc (Approximately 1 km at the Equator) R2025A, Version v1*; Global Demographic Data Project-Funded by The Bill and Melinda Gates Foundation (INV-045237); WorldPop-School of Geography and Environmental Science, University of Southampton: Southampton, UK, 2025. [CrossRef]
53. Kjeldsberg, F.; Munim, Z.; Bustgaard, M.; Bhagat, S.; Lindroos, E.; Haavardtun, P. Sensitivity of predictive performance assessment accuracy in varying k-fold cross validation. In *International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning*; Springer Nature: Cham, Switzerland, 2024; pp. 71–82.
54. Abedin, T.; Xu, H.; Uddin, S. The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models. *Sci. Rep.* **2026**, *16*, 6084. [CrossRef] [PubMed]
55. Sun, Q.; Guo, L.; Gao, G.; Baidurela, A.; Li, L.; Song, T. Construction and verification of thermal infrared hyperspectral mathematical model for soil organic matter in arid areas. *J. Soils Sediments* **2026**, *26*, 16. [CrossRef]
56. Wang, X.; Zhang, F.; Ding, J.; Kung, H.; Latif, A.; Johnson, V. Estimation of soil salt content (SSC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR), Northwest China, based on a Bootstrap-BP neural network model and optimal spectral indices. *Sci. Total Environ.* **2018**, *615*, 918–930. [CrossRef] [PubMed]
57. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
58. Lin, X.; Wu, M.; Shao, X.; Li, G.; Hong, Y. Water turbidity dynamics using random forest in the Yangtze River Delta Region, China. *Sci. Total Environ.* **2023**, *903*, 166511. [CrossRef] [PubMed]
59. Vapnik, V. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **2019**, *10*, 988–999.
60. Martínez-Álvarez, F.; Troncoso, A.; Asencio-Cortés, G.; Riquelme, J. A survey on data mining techniques applied to electricity-related time series forecasting. *Energies* **2015**, *8*, 13162–13193. [CrossRef]
61. Eslavath, R.; Mummadi, U. Enhancing Android Malware Detection: A Grid-Tuned Two-Layered Stacking Approach. *SSRG Int. J. Electron. Commun. Eng.* **2024**, *11*, 253–269. [CrossRef]
62. Villavicencio, C.; Macrohon, J.; Inbaraj, X.; Jeng, J.; Hsieh, J. Covid-19 prediction applying supervised machine learning algorithms with comparative analysis using weka. *Algorithms* **2021**, *14*, 201. [CrossRef]
63. Leite, D.; Weigang, L.; Taco, P. Artificial intelligence perfecting autonomous distress locator: Defusing false positives. In *XXII Simpósio de Transporte Aéreo*; Universidade de Brasília: Brasília, Brazil, 2026. [CrossRef]
64. Burnett, R.; Pope, I.I.I.C.; Ezzati, M.; Olives, C.; Lim, S.; Mehta, S.; Shin, H.; Singh, G.; Hubbell, B.; Brauer, M.; et al. An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **2014**, *122*, 397. [CrossRef] [PubMed]
65. Lelieveld, J.; Evans, J.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **2015**, *525*, 367–371. [CrossRef] [PubMed]

66. Apte, J.; Marshall, J.; Cohen, A.; Brauer, M. Addressing global mortality from ambient PM2.5. *Environ. Sci. Technol.* **2015**, *49*, 8057–8066. [CrossRef] [PubMed]
67. Maji, K.; Dikshit, A.; Arora, M.; Deshpande, A. Estimating premature mortality attributable to PM2.5 exposure and benefit of air pollution control policies in China for 2020. *Sci. Total Environ.* **2018**, *612*, 683–693. [CrossRef] [PubMed]
68. Zhou, M.; Wang, H.; Zhu, J.; Chen, W.; Wang, L.; Liu, S.; Li, Y.; Wang, L.; Liu, Y.; Yin, P.; et al. Cause-specific mortality for 240 causes in China during 1990–2013: A systematic subnational analysis for the Global Burden of Disease Study 2013. *Lancet* **2016**, *387*, 251–272. [CrossRef] [PubMed]
69. Xie, Y.; Dai, H.; Dong, H.; Hanaoka, T.; Masui, T. Economic impacts from PM2.5 pollution-related health effects in China: A provincial-level analysis. *Environ. Sci. Technol.* **2016**, *50*, 4836–4843. [CrossRef] [PubMed]
70. Huang, T.; Ling, Z.; Ma, J.; Macdonald, R.W.; Gao, H.; Tao, S.; Tian, C.; Song, S.; Jiang, W.; Chen, L.; et al. Human exposure to polychlorinated biphenyls embodied in global fish trade. *Nat. Food* **2020**, *1*, 292–300. [CrossRef]
71. Su, Y.; Chen, X.; Guo, J.; Yang, A. Spatio-temporal distribution of AOD and its response to regional energy consumption and air pollution factors in China. *Environ. Res.* **2025**, *290*, 123393. [CrossRef] [PubMed]
72. Li, K.; Bai, K.; Li, Z.; Guo, J.; Chang, N. Synergistic data fusion of multimodal AOD and air quality data for near real-time full coverage air pollution assessment. *J. Environ. Manag.* **2022**, *302*, 114121. [CrossRef] [PubMed]
73. Zhang, X.; Liu, C.; Jiang, Y. Spatiotemporal dynamics of terrestrial vegetation carbon sinks and its response to climate change in northwest China. *Front. Environ. Sci.* **2025**, *13*, 1550501. [CrossRef]
74. Xia, Z. Analysis of the Spatiotemporal Variation Characteristics and Driving Factors of Land Vegetation GPP in a Certain Region of Asia. *Open J. Ecol.* **2024**, *14*, 523–543.
75. Meng, L.; Zhao, T.; He, Q.; Yang, X.; Mamtimin, M.; Yang, F.; Zhou, C.; Huo, W.; Wang, M.; Pan, H.; et al. Climatic characteristics of floating dust and persistent floating dust over the Tarim basin in the recent 30 years. *Acta Meteorol. Sin.* **2022**, *80*, 322–333. (In Chinese) [CrossRef]
76. Meng, L.; Zhao, T.; He, Q.; Yang, X.; Mamtimin, A.; Wang, M.; Pan, H.; Huo, W.; Yang, F.; Zhou, C. Dust Radiative Effect Characteristics during a Typical Springtime Dust Storm with Persistent Floating Dust in the Tarim Basin, Northwest China. *Remote Sens.* **2022**, *14*, 1167. [CrossRef]
77. Aishan, T.; Sun, Y.; Halik, M.; Betz, F.; Yusup, A.; Rezhake, R. Spatiotemporal changes in fine particulate matter and ozone in the oasis city of Korla, northeastern Tarim Basin of China. *Sci. Rep.* **2024**, *14*, 12. [CrossRef] [PubMed]
78. Radzka, E.; Rymuza, K. The effect of meteorological conditions on PM10 and PM2.5 pollution of the air. *Rocz. Ochr. Sr.* **2019**, *21*, 611–628. [CrossRef] [PubMed]
79. Zhan, S. How Meteorological Factors Affect the Formation and Diffusion of Atmospheric Particles: A Case Study of Beijing. In *E3S Web of Conferences*; EDP Sciences: Les Ulis, France, 2025; Volume 666, p. 01012.
80. Fujino, R.; Miyamoto, Y. PM2.5 decrease with precipitation as revealed by single-point ground-based observation. *Atmos. Sci. Lett.* **2022**, *23*, e1088. [CrossRef]
81. Rahman, H.; Kwicklis, M.; Ottom, M.; Amornsriwatanakul, A.; Abdul-Mumin, K.; Rosenberg, M.; Dinov, I. Machine learning-based prediction of mental well-being using health behavior data from university students. *Bioengineering* **2023**, *10*, 575. [CrossRef] [PubMed]
82. Akpabio, A.; Minnitt, R. Improving grade estimation using machine learning: A comparative study of ordinary kriging against machine learning algorithms. *J. S. Afr. Inst. Min. Metall.* **2026**, *126*, 143–156. [CrossRef]
83. Ebrahimy, H.; Azadbakht, M. Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Comput. Geosci.* **2019**, *124*, 93–102. [CrossRef]
84. Filippelli, S.; Schleeweis, K.; Nelson, M.; Fekety, P.; Vogeler, J. Testing temporal transferability of remote sensing models for large area monitoring. *Sci. Remote Sens.* **2024**, *9*, 100119. [CrossRef]
85. *EPA-452/R-12-003*; Regulatory Impact Analysis for the Proposed Revisions to the National Ambient Air Quality Standards for Particulate Matter. US Environmental Protection Agency, Office of Air Quality Planning and Standards, Health and Environmental Impacts Division: Washington, DC, USA, 2012.
86. Chen, H.; Yan, Y.; Hu, D.; Peng, L.; Wang, C. PM2.5-bound heavy metals in a typical industrial city of Changzhi in North China: Pollution sources and health risk assessment. *Atmos. Environ.* **2024**, *321*, 120344. [CrossRef]
87. Chen, S.; Li, D.; Wu, X.; Chen, L.; Zhang, B.; Tan, Y.; Yu, D.; Niu, Y.; Duan, H.; Li, Q.; et al. Application of cell-based biological bioassays for health risk assessment of PM2.5 exposure in three megacities, China. *Environ. Int.* **2020**, *139*, 105703. [CrossRef] [PubMed]