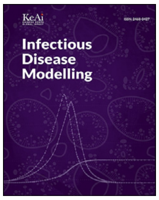
Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm



Confounding amplifies the effect of environmental factors on COVID-19

Zihan Hao [a], Shujuan Hu [a, *], Jianping Huang [b], Jiaxuan Hu [a], Zhen Zhang [a], Han Li [b], Wei Yan [b]

[a] *College of Atmospheric Sciences, Lanzhou University, Lanzhoum, 730000, China*
[b] *Collaborative Innovation Center for Western Ecological Safety, College of Atmospheric Sciences, Lanzhou University, Lanzhou, 730000, China*

ARTICLE INFO

Article history:
Received 15 March 2024
Received in revised form 26 May 2024
Accepted 16 June 2024
Available online 21 June 2024
Handling Editor: Dr Yijun Lou

Keywords:
COVID-19
Environmental factors
Causal analysis
Double machine learning

ABSTRACT

The global COVID-19 pandemic has severely impacted human health and socioeconomic development, posing an enormous public health challenge. Extensive research has been conducted into the relationship between environmental factors and the transmission of COVID-19. However, numerous factors influence the development of pandemic outbreaks, and the presence of confounding effects on the mechanism of action complicates the assessment of the role of environmental factors in the spread of COVID-19. Direct estimation of the role of environmental factors without removing the confounding effects will be biased. To overcome this critical problem, we developed a Double Machine Learning (DML) causal model to estimate the debiased causal effects of the influencing factors in the COVID-19 outbreaks in Chinese cities. Comparative experiments revealed that the traditional multiple linear regression model overestimated the impact of environmental factors. Environmental factors are not the dominant cause of widespread outbreaks in China in 2022. In addition, by further analyzing the causal effects of environmental factors, it was verified that there is significant heterogeneity in the role of environmental factors. The causal effect of environmental factors on COVID-19 changes with the regional environment. It is therefore recommended that when exploring the mechanisms by which environmental factors influence the spread of epidemics, confounding factors must be handled carefully in order to obtain clean quantitative results. This study offers a more precise representation of the impact of environmental factors on the spread of the COVID-19 pandemic, as well as a framework for more accurately quantifying the factors influencing the outbreak.

1. Introduction

Since the first case of the coronavirus disease 2019 (COVID-19) was reported in 2019, the outbreak has spread rapidly across countries. As of October 4, 2023, the World Health Organization reported 771, 151, 224 confirmed cases of COVID-19

* Corresponding author.
E-mail address: hushuju@lzu.edu.cn (S. Hu).

Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2024.06.005

worldwide, resulting in 6,960,783 deaths (WHO, 2023). The substantial harm caused by the outbreak has raised awareness regarding the criticality of timely alert, prevention and control of epidemics (Huang et al., 2020, 2021). Currently, the severity and fatality of COVID-19 has significantly decreased (Marziano et al., 2023). However, the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causing COVID-19 pandemic is constantly evolving, and the number of infections continues to rise, presenting challenges in the prevention, treatment, and diagnosis (Jacobs et al., 2023; Malik et al., 2022; Markov et al., 2023). With the rise of globalization, the potential risk for future infectious disease outbreaks has also risen significantly. It is crucial to have a comprehensive understanding of COVID-19 transmission mechanisms for our future control of novel infectious diseases.

The primary mode of transmission of the COVID-19 has been identified as inhalation of respiratory droplets from infected individuals and aerosols containing the virus, as well as direct contact with contaminated surfaces (Wang et al., 2021; Weaver et al., 2022). Given that both the survival and transmission of SARS-CoV-2 in aerosols are intricately linked to the environment (Weaver et al., 2022), it is imperative to attain a thorough and precise understanding of the causal association between COVID-19 and environmental factors. Recently, researchers have made many efforts to clarify the role of the environment in the spread of COVID-19 that mainly focus on the meteorological conditions (such as temperature, humidity, wind speed, and ultraviolet radiation) and air pollution elements (including PM2.5, PM10, nitrogen dioxide, and sulfur dioxide) (Ali & Islam, 2020; Liu et al., 2021; Paraskevis et al., 2021; Xu et al., 2021). They commonly used correlation analysis, regression methods and machine learning methods to quantify the impact of environmental factors on COVID-19 outbreaks. In the correlation analysis, Pearson correlation (Yang et al., 2021), Spearman rank correlation and Kendall correlation (Bashir et al., 2020; Zhang et al., 2020) were used to calculate the correlation coefficients between environmental factors and COVID-19. Regression models such as multivariate stepwise regression (Yang et al., 2021), multiple linear logistic regression (Wang, Dong, et al., 2021), Loess regression (Poirier et al., 2020), and multivariate Poisson regression (Jiang et al., 2020) were used to quantify the contributions of environmental factors. For the machine learning models, Random Forest (RF) were used to determine that PM2.5 was the main factor affecting the transmission of COVID-19 in the United States (Milicevic et al., 2021). Gradient Boosting Decision Trees (GBDT) and Ridge regression methods were used to analyze the impact of altitude, air pollutants and NPIs on the transmission process (Han et al., 2022). Shallow Perceptron Neural Network (SSLPNN) and Gaussian Process Regression models were used to analyze the relationship between COVID-19 cases and environmental factors in five regions of Asia (Ahmad et al., 2020). Thus, these current studies have focused primarily on establishing correlations and have not analyzed the causal relationship between environmental factors and the COVID-19 pandemic. In addition, despite extensive quantitative study, there were no consistent conclusions. The contributions of environmental factors on the transmission of COVID-19 exhibits significant variation from one region to another (Ford et al., 2022; Qi et al., 2020; Yang et al., 2021). For instance, some studies demonstrate a negative correlation between temperature and COVID-19 (Wang, Dong, et al., 2021; Poirier et al., 2020; Yin et al., 2022), whereas others reach the opposite conclusion (Xie & Zhu, 2020; Bashir et al., 2020; Sun et al., 2022) or even propose that temperature is not linked to COVID-19 (Yao et al., 2020). Among air pollutants, PM2.5 and PM10 have been observed to exhibit a positive association with COVID-19 in certain studies (Cao et al., 2021; Zhu et al., 2020). However, contrasting results have emerged, revealing a negative correlation between PM10 and COVID-19 outbreaks in one investigation (Jiang et al., 2020) and insignificance in another (Li et al., 2020). The diversity of environmental effects prompted an acknowledgment of the intricate interplay between the environment and the transmission of COVID-19. Therefore, it is imperative to establish statistical models to more finely quantify causality and comprehensively explore changes in environmental effects. Because of the myriad and interconnected factors influencing COVID-19, accurately quantifying the influence coefficients is challenged by the presence of confounding factors. Direct estimation of the target variable's contribution through regression models or machine learning methods can be subject to regularization bias if confounders are present (Chernozhukov et al., 2018). Also, overfitting of the model can lead to serious bias in the quantifying estimation. Recently the Double Machine Learning (DML) models for causal inference have been proposed to eliminate the bias introduced by the commonly used regression models or machine learning methods. It achieves debiasing by two key operations: (1) using orthogonalization to remove the role of confounding factors, thus overcoming regularization bias; (2) using cross-fitting to overcome bias caused by overfitting (Chernozhukov et al., 2018). The DML modeling framework is flexible enough to allow the use of a variety of machine learning methods as function estimators, facilitating the implementation of the fitting of nonlinear relationships. Currently, DML model methods are commonly used for estimating causal effects in socio-economic problems, such as estimating the direct versus indirect effects of health insurance on health status (Colangelo & Lee, 2020; Farbmacher et al., 2022). In this study, we constructed a DML model to calculate unbiased estimates of the causal effects of environmental factors on the spread of COVID-19 by using data on COVID-19 outbreaks in 2022 in various administrative regions of China. Then, the changing patterns of environmental effects were quantitatively analyzed. Our aim is to gain a better understanding of the causal relationship between environmental factors and the spread of COVID-19.

The organization of the subsequent sections of the paper is as follows. Section 2 describes the study area and data, and illustrates the model theory used in this paper. Section 3 is the results section. A preliminary analysis of the data, initially presented in Section 3.1, indicates the potential for confounding effects between environmental factors and the spread of COVID-19. Section 3.2 presents an debiased estimation of the causal effect of environmental factors on the spread of COVID-19 using a double machine learning model. It also compares the results of multiple linear regression with those of the double machine learning causal model. Section 3.3 further discusses the heterogeneity of environmental causal effects and analyzes their trends. Finally, the discussion and conclusion are in Section 4.

2. Materials & methods

2.1. Study area and data

We collected the number of daily new asymptomatic COVID-19 cases in 4 municipalities and 333 prefecture-level administrative districts in mainland China from 1 January to November 30, 2022 (notably, China issued a notice on December 7, 2022 to stop full-scale nucleic acid testing and announced on December 13, 2022 to stop reporting the number of daily infections). In addition, to ensure a sample size of data for each outbreak, we screened all outbreak sequences lasting more than 45 days in 2022, resulting in time-series data for a total of 261 outbreaks.

In evaluating the progression of an outbreak, commonly utilized indicators include the number of effective reproductions and the number of new cases. Given that China conducted nucleic acid testing on all residents during the outbreak, the number of daily new cases can be an acceptable and more direct proxy indicator of transmissibility. According to data published by the National Health Commission, China categorizes COVID-19 infections into asymptomatic and confirmed cases (newly confirmed cases include individuals who have progressed from asymptomatic to confirmed cases). China's asymptomatic and confirmed cases both showed two peaks in the first half (April—May) and second half (November—December) of 2022, with asymptomatic cases generally outnumbering confirmed cases (Fig. 1 (a)), accompanied by regionally widespread outbreaks (Fig. 1(b and c)). The number of cases was predominantly asymptomatic, and the correlation coefficient between new confirmed cases and new asymptomatic cases was 0.888. The data on new asymptomatic cases provided the majority of the information regarding the development of the outbreak. Consequently, the number of new asymptomatic cases was employed as the dependent variable in the proposed model, representing the spread of the epidemic, denoted as Y. Furthermore, the data was smoothed over a 14-day period in order to reduce the impact of random errors.

Daily meteorological factors include 2 m air temperature (TEMP), 2 m dewpoint temperature (D), 10 m wind speed (W), and downward ultraviolet radiation at the surface (UV), where the 10 m wind speed is calculated by taking the square root of the sum of the squared 10 m u-component and v-component of horizontal wind. The meteorological data were obtained by averaging over the region using ERA5 reanalysis data with a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$(Hersbach et al., 2023). ERA5 is the fifth-generation atmospheric reanalysis of global climate data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Daily air pollution factors include concentrations of PM_{10}, $PM_{2.5}$, sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3) obtained from the China National Environmental Monitoring Center (China Environmental Monitoring Center). Data on population mobility factors are obtained from the Baidu Migration Index (Baidu Migration), which includes the daily in-migration size index (MI), out-migration size index (MO), work travel intensity index (WI), and dining & leisure travel intensity index (DI) for each region. Finally, the various control policies implemented by governments in response to outbreaks are also influential factors that cannot be ignored. The Oxford Covid-19 Government

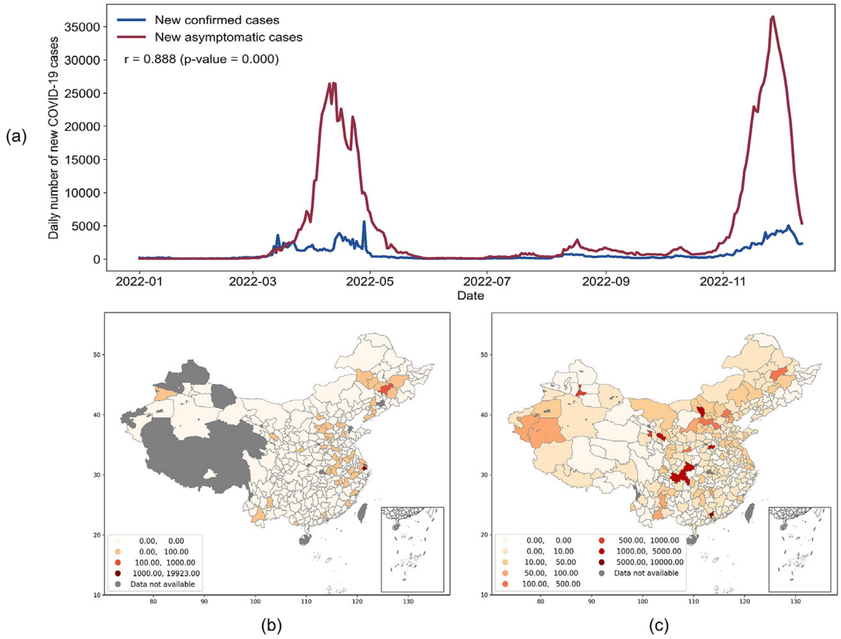


Fig. 1. Descriptive analysis of COVID-19 data. (a): Daily number of new cases in the COVID-19 epidemic in China in 2022. The blue line indicates new confirmed cases, whereas the red line denotes new asymptomatic cases. The Spearman correlation coefficient between the new confirmed cases and new asymptomatic cases is 0.888. (b) Distribution of new asymptomatic cases in mainland China on April 15, 2022. (c): Distribution of new asymptomatic cases in mainland China on November 15, 2022.

Response Tracker (OxCGRT) collects information on the response measures taken by governments and calculates a stringency index (Hale et al., 2021). The stringency indices (SI) are calculated for containment and closure policies and health system policies. At the same time, policies were differentiated by vaccine, and the index was calculated separately for vaccinated and unvaccinated people, and then the average of the two was calculated to obtain the stringency index. However, the stringency index dataset is daily data for provincial administrative districts. We assume that local governments in the same province follow the same outbreak control policies. In general, the data used in this study and the basic information of the data are summarized in Table 1.

2.2. Double machine learning

We define the dependent variable as Y and the influence factors as $X = (X_1, X_2, \ldots, X_P)$. When calculating the effect (denoted by θ) of a variable T (often called the treatment variable) in X on Y, there may exist some variables that affect Y while also having an effect on T. Such variables are called confounding factors (X^c). We can describe this problem using the partial linear regression (PLR) model (Robinson, 1988), as shown in Equations (1) and (2). The effect of confounding factors on Y is denoted as $g(X^c)$, and the effect of confounding factors on T is $m(X^c)$.

$$Y = \theta T + g\left(X^c\right) + U, E\left[U\middle|X^c, T\right] = 0, \tag{1}$$

$$T = m\left(X^c\right) + V, E\left[V\middle|X^c\right] = 0, \tag{2}$$

where U and V are the residual terms, $E[U|X^c, T]$ is the conditional expectation of U given X^c and T, $E[V|X^c]$ is the conditional expectation of V given X^c.

The traditional approach is to model T together with X^c for Y without considering the relationship between the confounding factors and the treatment variable. For example, the estimation of $\widehat{\theta T} + \widehat{g}(X)$ is obtained by directly fitting Equation (1), using multiple linear regression or other machine learning methods. If the number of samples is n, then the estimate $\widehat{\theta}$ for θ is:

$$\widehat{\theta} = \left(\frac{1}{n}\sum_{i\in n} T_i^2\right)^{-1} \frac{1}{n}\sum_{i\in n} T_i\left(Y_i - \widehat{g}\left(X_i^c\right)\right), \tag{3}$$

Assuming that the true θ is θ_0, the estimated bias is:

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \left(\frac{1}{n}\sum_{i\in n} T_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i\in n} T_i U_i + \left(\frac{1}{n}\sum_{i\in n} T_i^2\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i\in n} T_i\left(g(X_i) - \widehat{g}\left(X_i^c\right)\right). \tag{4}$$

The source of the bias is the overfitting of the model and the regularization bias introduced by the estimation $\widehat{g}(X_i^c)$ (Chernozhukov et al., 2018). Since T is related to X^c, $\frac{1}{\sqrt{n}}\sum\limits_{i\in N} T_i(g(X_i) - \widehat{g}(X_i^c))$ is nonzero. Thus, the estimator $\widehat{\theta}$ typically has a convergence rate of less than $1/\sqrt{n}$.

Table 1
Overview of data used in the study.

	Data name	Unit	Timescale
Epidemic data	New asymptomatic cases	persons	Daily
Meteorological data	2 m air temperature	° C	Daily
	2 m dewpoint temperature	° C	
	10 m horizontal wind speed	m/s	
	downward ultraviolet radiation at the surface	KJ/m^2	
Air pollution data	PM10	μg/m^3	Daily
	PM2.5	μg/m^3	
	SO_2	μg/m^3	
	NO_2	μg/m^3	
	CO	mg/m^3	
	O_3	μg/m^3	
Population mobility	in-migration size index	/	Daily
	out-migration size index	/	
	work travel intensity index	/	
	dining & leisure travel intensity index	/	
Government control policy	stringency index	/	Daily

The Double Machine Learning (DML) algorithm (Chernozhukov et al., 2018) overcome regularization bias by orthogonalization. By using a machine learning method to fit Equation (2), the obtained residual $V = T - m(X^c)$ is orthogonal to X^c, and the effect of X^c on T is eliminated from T. At the same time, the effect of X^c on Y is also removed from Y. Finally, a regression model of the residual of T and the residual of Y is established:

$$Y - g\left(X^c\right) = \theta V + U. \tag{5}$$

In the DML model, $\hat{m}(X^c)$ and $\hat{g}(X^c)$ are fitted by machine learning model respectively. At this point, the estimate for θ is:

$$\breve{\theta} = \left(\frac{1}{n}\sum_{i\in n}\hat{V}_i T_i\right)^{-1}\frac{1}{n}\sum_{i\in n}\hat{V}_i\left(Y_i - \hat{g}\left(X_i^c\right)\right), \tag{6}$$

Since V and X^c are orthogonal, it is easy to get that $\frac{1}{\sqrt{n}}\sum\limits_{i\in n}\widehat{V}_i(g(X_i) - \widehat{g}(X_i^c))$ tends to zero, thereby avoiding the regularization bias in the estimation error.

Additionally, the cross-fitting operation using in DML model can effectively avoid the bias caused by over-fitting (Chernozhukov et al., 2018). Specifically, the total sample is divided into two sub datasets. First, one sub dataset is selected as the training set for fitting $g(X^c)$ and $m(X^c)$, The other sub dataset is used as the test set to compute θ after obtaining the residual of T and the residual of Y based on the trained model. Subsequently, the two sub datasets are exchanged, and the aforementioned steps are repeated in order to obtain a new estimation of θ. The average of the two estimates ($\widehat{\theta}_{mean}$) is calculated as the final estimate of θ.

2.3. Modeling configurations for quantifying environmental effects on COVID-19

The DML model framework we established to calculate the causal effects of environmental factors on COVID-19 is shown in Fig. 2 (a). In the model framework, Random Forest model (Breiman, 2001) was employed to fit functions $g(X^c)$ and $m(X^c)$ with three-fold cross validation. In addition to environmental factors, population movements can lead to the spread of epidemics and increase the risk of large-scale outbreaks (Jia et al., 2020; Li et al., 2023; Zheng et al., 2020). Therefore, the variables input into the model that affect epidemic transmission include environmental factors and population movement factors. Moreover, we have introduced the first-order lag of daily new asymptomatic cases and time index as input variables to reflect the trends of the epidemic and environmental factors. The output variable was the daily count of new asymptomatic cases. In order to estimate the causal effect of environmental factors on the transmission of COVID-19, environmental factors were successively taken as treatment variables (T) in the model, and the remaining influence factors were confounding factors (X^c). For example, when estimating the causal effect of temperature, temperature will be used as a treatment variable (T), and influence factors other than temperature will be used as confounding factors (X^c). By replacing treatment variables and confounding factors, the model is trained to estimate the causal effects of all environmental factors. The code implementation of the above model is based on the DoubleML python package. (Bach et al., 2022).

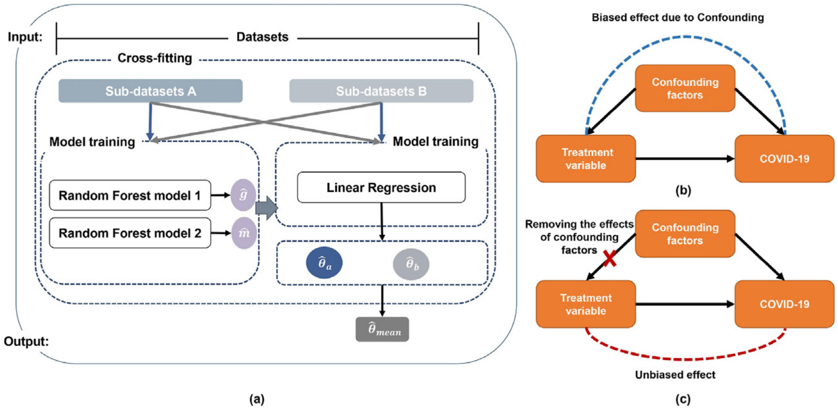


Fig. 2. Framework of calculated unbiased causal coefficients of environmental factors on COVID-19. (a): The DML model framework diagram. (b): Direct estimation of the role of the treatment variable will produce biased estimates due to the confounding effects. (c): Unbiased estimates of the role of the treatment variable are obtained by removing the effects of confounders.

3. Results

3.1. Confounding effect between environmental factors and COVID-19

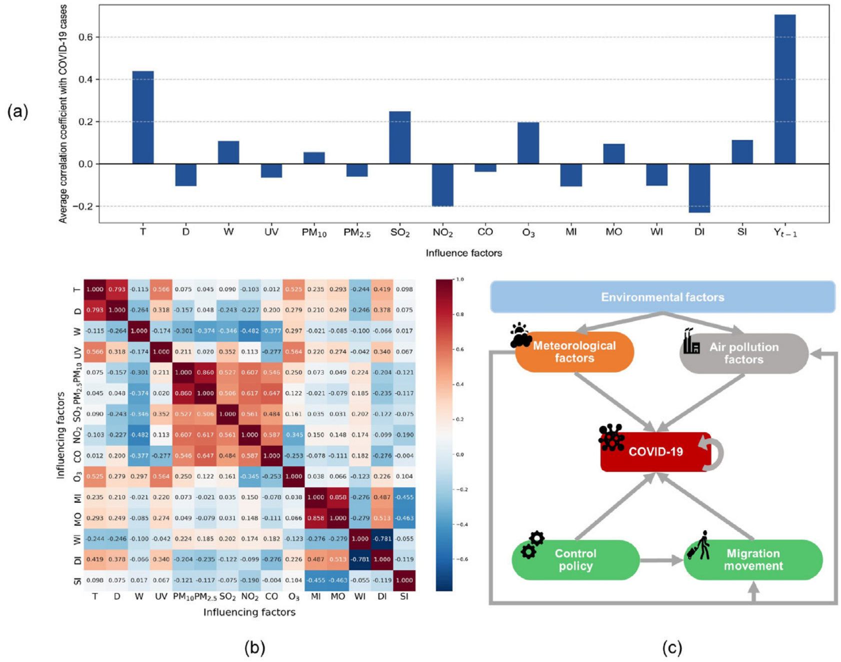
During the outbreak in each city, we computed the partial correlation coefficients between each influencing factor (including environmental factors, population mobility factors and stringency index) and the number of daily new asymptomatic cases of COVID-19. The average of all statistically significant results ($p \leq 0.05$) is illustrated in Fig. 3 (a). Meteorological factors, air pollution, population movement and stringency index are all correlated to some degree with the development of epidemics. Additionally, there is a strong correlation between the number of new asymptomatic infections reported on the current day and the number reported on the previous day (Y_{t-1}). Furthermore, Fig. 3 (b) displays the mean values of statistically significant correlation coefficients ($p \leq 0.05$) between influencing factors during the outbreaks in all cities. There is a significant correlation between the influencing factors. This indicates that there are confounding factors in calculating the effects of each factor.

The intricate interplay of various factors in COVID-19 transmission poses challenges for the precise evaluation of their contributions. There is a potential for bias when attempting to quantify the contribution of influencing factors. It is crucial to emphasize the use of a debiased model to obtain an unbiased estimate of the causal effect of environmental factors on COVID-19.

3.2. A debiased estimate of the environmental contribution to COVID-19

To assess causality, it is imperative to create a causality diagram and comprehensively consider the influences of the epidemic. Utilizing the correlations identified, we established the causal graph depicted in Fig. 3 (c). Subsequently, we constructed a double machine learning causal inference model based on this graph to verify the causal relationship between environmental factors and COVID-19.

We utilized multiple linear regression (MLR) to estimate the impact of the environment on COVID-19 and compared it with the causal effects obtained from the DML causal inference model. In order to ensure fairness, the input and output of the MLR model are consistent with the DML model. First, we compare the root-mean-square errors of the MLR model and the DML model on the test set. The results show that the DML model has a smaller error and it is reasonable to assume that the DML model estimates environmental effects more accurately (Fig. S1). Fig. 4 shows the distribution of the impact of environmental factors on many outbreaks estimated by the DML and MLR models. Only the results that passed the statistical significance test ($p \leq 0.05$) in both models are shown. From Fig. 4, we get a clear comparison of the estimation results for the contribution of environmental factors between the DML and MLR models. The estimates from both models indicate that



	T	D	W	UV	PM_{10}	$PM_{2.5}$	SO_2	NO_2	CO	O_3	MI	MO	WI	DI	SI
T	1.000	0.793	-0.115	0.566	0.075	0.045	0.090	-0.103	0.012	0.525	0.235	0.293	-0.244	0.419	0.098
D	0.793	1.000	-0.264	0.318	-0.157	0.048	-0.243	-0.227	0.200	0.279	0.210	0.249	-0.246	0.378	0.075
W	-0.115	-0.264	1.000	-0.174	-0.301	-0.374	-0.346	-0.482	-0.377	0.297	-0.021	-0.085	-0.100	-0.066	0.017
UV	0.566	0.318	-0.174	1.000	0.211	0.020	0.352	0.113	-0.277	0.564	0.220	0.274	-0.042	0.340	0.067
PM_{10}	0.075	-0.157	-0.301	0.211	1.000	0.860	0.527	0.607	0.546	0.250	0.073	0.049	0.224	-0.204	-0.121
$PM_{2.5}$	0.045	0.048	-0.374	0.020	0.860	1.000	0.506	0.617	0.647	0.122	-0.021	-0.079	0.185	-0.235	-0.117
SO_2	0.090	-0.243	-0.346	0.352	0.527	0.506	1.000	0.561	0.484	0.161	0.035	0.031	0.202	-0.122	-0.075
NO_2	-0.103	-0.227	-0.482	0.113	0.607	0.617	0.561	1.000	0.587	-0.345	0.150	0.148	0.174	-0.099	-0.190
CO	0.012	0.200	-0.377	-0.277	0.546	0.647	0.484	0.587	1.000	-0.253	-0.078	-0.111	0.182	-0.276	-0.004
O_3	0.525	0.279	0.297	0.564	0.250	0.122	0.161	-0.345	-0.253	1.000	0.038	0.066	-0.123	0.226	0.104
MI	0.235	0.210	-0.021	0.220	0.073	-0.021	0.035	0.150	-0.078	0.038	1.000	0.858	-0.276	0.487	-0.455
MO	0.293	0.249	-0.085	0.274	0.049	-0.079	0.031	0.148	-0.111	0.066	0.858	1.000	-0.279	0.513	-0.463
WI	-0.244	-0.246	-0.100	-0.042	0.224	0.185	0.202	0.174	0.182	-0.123	-0.276	-0.279	1.000	-0.781	-0.055
DI	0.419	0.378	-0.066	0.340	-0.204	-0.235	-0.122	-0.099	-0.276	0.226	0.487	0.513	-0.781	1.000	-0.119
SI	0.098	0.075	0.017	0.067	-0.121	-0.117	-0.075	-0.190	-0.004	0.104	-0.455	-0.463	-0.055	-0.119	1.000

Fig. 3. There are confounding factors in the relationship between environmental factors and COVID-19. (a): Average partial correlation coefficient between influencing factors and COVID-19. (b): Mean spearman correlation coefficients among influencing factors during the COVID-19 outbreak in China. (c): Causal graph of influencing factors and COVID-19.

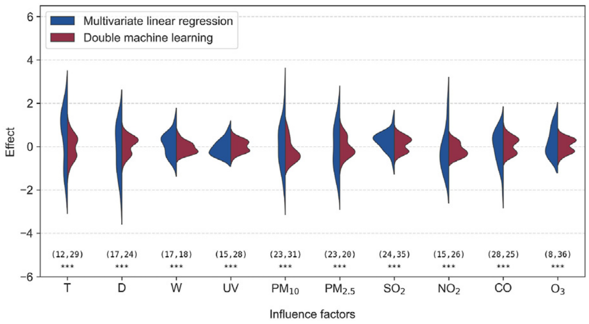


Fig. 4. Distribution of estimates of environmental factor contributions in MLR and DML models that passed statistical significance tests. *** stands for p-value ≤ 0.05, and the numbers in parentheses represent the amount of MLR and DML model estimates that pass the statistical test at the significance level of $\alpha = 0.05$, respectively.

environmental factors play varying roles in outbreaks across different regions and exhibit a roughly bimodal distribution. However, the estimates obtained from MLR exhibit greater dispersion. When confounding effects are present, the bias in the MLR estimates further exacerbates the heterogeneity of environmental contributions.

Furthermore, we use slope plots to compare the estimates of the two models in the same region. Fig. 5 displays the results for each environmental factor separately. Blue is employed to signify instances where MLR estimates surpass DML estimates, while red is utilized when MLR estimates are lower than DML estimates. From the overall view of the figure, the predominance of blue tends to highlight areas where environmental factors play a positive role, whereas red is more prevalent in regions where environmental factors exhibit a negative impact. Therefore, the MLR is somewhat overestimated for both positive and negative contributions to environmental factors.

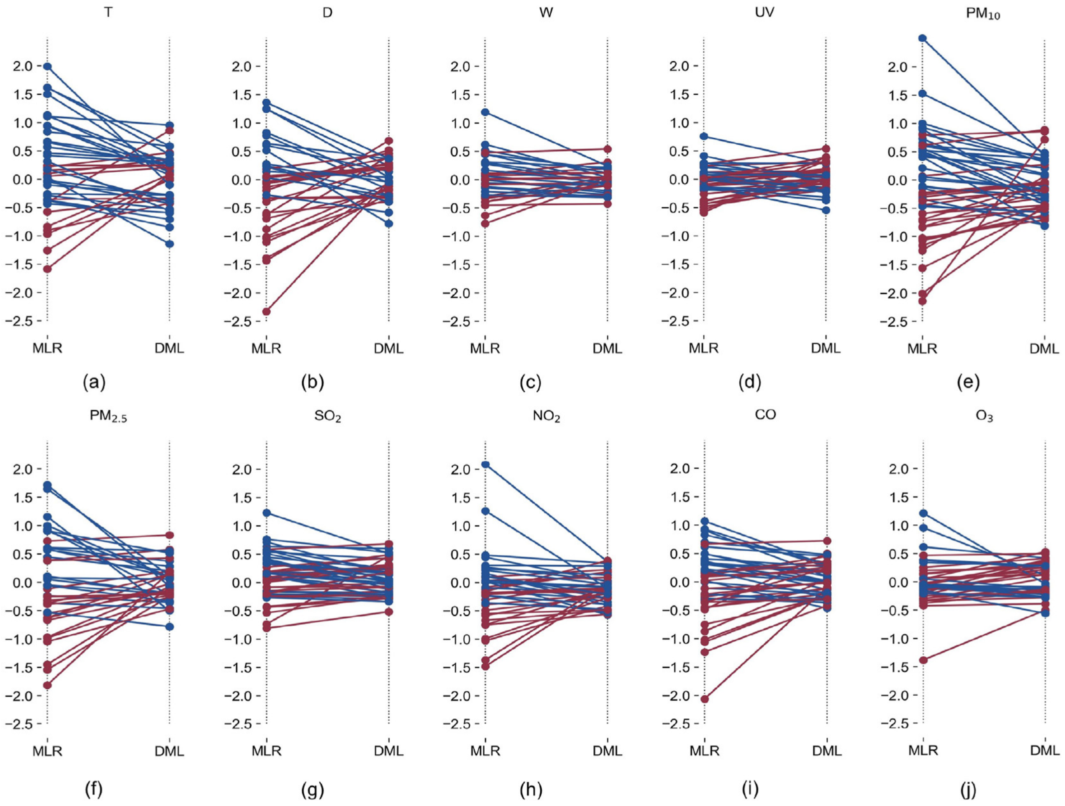


Fig. 5. A slope graph comparing the discrepancies in estimated contributions of environmental factors to COVID-19 between MLR and DML models for the same region. The subplots respectively display the differences in results between the two models for different environmental factors. In the graphs, blue indicates instances where MLR model results surpass those of the DML model, while red signifies cases where MLR results are lower than those of the DML model.

The results of the statistical hypothesis tests further illustrate the differences between the MLR and DML model results (Table 2). First, we used one-tailed F-tests to compare the variance of the results from the two models. The results for all environmental factors rejected the null hypothesis at the $\alpha = 0.05$ significance level that the variance of the MLR model results is greater than the variance of the DML model results. As a result, the estimates from the MLR model are more discrete, confirming that the MLR model increases heterogeneity in the role of environmental factors. We also tested whether the means of the results of the two models were significantly different. In the presence of positive and negative numbers, the mean test may not show differences in the means due to offsetting problems. To better understand and test the differences between the results of the two models, we classified the positive and negative effects of environmental factors based on the estimates of the DML model. Based on the results of the Shapiro-Wilk test (Table S1), the data did not satisfy normal distribution, so the nonparametric test Wilcoxon signed-rank test was chosen to test whether the results of MLR were significantly greater than those of DML. At the significance level of $\alpha = 0.05$, all environmental factors passed the significance test in either positive or negative effects except for CO, indicating the overestimation of the role of the environment by the MLR model. The results for CO did not pass the significance test, indicating that there was no significant difference between the two models in estimating the role of CO on COVID-19, which may be due to the weaker effect of confounders on CO. In addition, the sample sizes for all statistical tests are shown in Table S2.

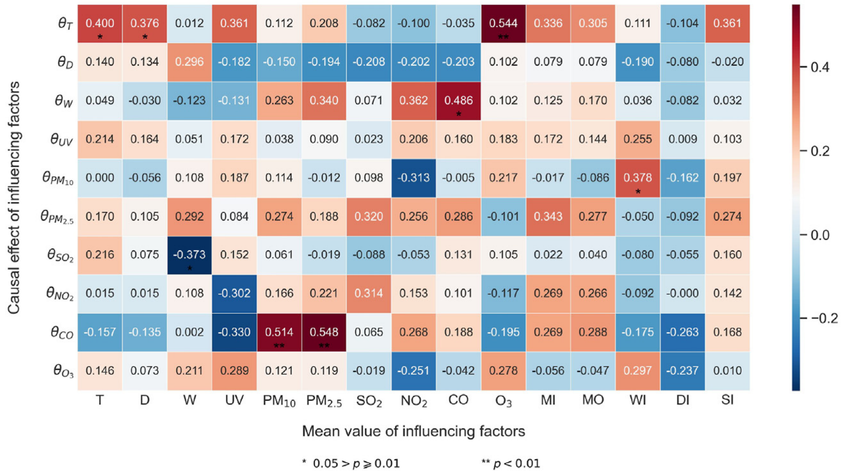
3.3. Heterogeneity in the causal effects of environmental factors on COVID-19

The estimation results in Section 3.2 have shown that the contribution of environmental factors to COVID-19 transmission varies by region. In this section, we will further analyze the causes of this heterogeneity and the changing rules of the role of environmental factors. Fig. 6 shows the spearman correlation coefficient between the mean values of influencing factors and environmental causal effects at the time of urban outbreaks. The causal effect of temperature on COVID-19 (θ_T) showed a significant correlation with local mean temperature, mean dew point temperature, and mean O_3 concentrations. Likewise, the causal effect of W (θ_W) showed a significant correlation with the local mean CO concentration. In the case of PM_{10}, its causal effects ($\theta_{PM_{10}}$) were notably correlated with local mean WI. The causal effect of SO_2 (θ_{SO_2}) showed a significant correlation with the local mean wind speed. Furthermore, the causal effect of CO (θ_{CO}) exhibited a significant correlation with the local mean PM_{10} concentration and mean $PM_{2.5}$ concentration. Correlation analyses reveal environmental disparities as contributors to causal effects heterogeneity.

To analyze the changing trend of environmental factors' causal effect on COVID-19, regression analysis was performed on the statistically significant correlations mentioned above. Fig. 7 illustrates the results of the regression coefficients passing the significance test (p-value $\leq$0.05). Under different temperature and humidity conditions, significant differences in the relationship between temperature and COVID-19 transmission were observed. The causal effect of temperature on COVID-19 showed a tendency to increase with rising temperature and humidity. On average, the causal effect of temperature on COVID-19 increases by about 0.04 and 0.03 per degree increase in mean temperature and mean dew point temperature, respectively (Fig. 7(a and b)). In regions where temperatures fall below 10 °C, the correlation between temperature and COVID-19 was primarily negative. Within the temperature range of 10–20 °C, a mixed pattern emerges, with both negative and positive effects observed, suggesting the involvement of additional influencing factors. Areas with temperatures above 20 °C turned out to show mainly positive causality. In summary, characterizing the impact of temperature and humidity on COVID-19 as strictly positive or negative proves insufficient. Instead, its effects exhibit diversity, contingent upon the specific environmental conditions. Temperature has a significant impact on the transmission of COVID-19 in both hot, humid climates and cold, dry climates. Changes in temperature and humidity, whether they increase or decrease, do not guarantee containment of the outbreak.

Table 2
Statistical tests of the results of MLR and DML models.

Environmental factor	F-test		Wilcoxon signed-rank test (+)		Wilcoxon signed-rank test (−)	
	statistic	p	statistic	p	statistic	p
T	6.62	0.00	5029.00	0.01	4695.00	0.59
D	8.15	0.00	4101.00	0.14	3677.00	0.00
W	2.95	0.00	5469.00	0.00	3761.00	0.06
UV	3.27	0.00	5855.00	0.01	2665.00	0.00
PM10	3.86	0.00	5834.00	0.03	2856.00	0.02
PM2.5	6.16	0.00	5274.00	0.01	3377.00	0.02
SO2	2.35	0.00	5413.00	0.00	3335.00	0.01
NO2	4.39	0.00	6552.00	0.07	2023.00	0.00
CO	4.73	0.00	3448.00	0.57	4540.00	0.11
O3	2.68	0.00	5605.00	0.01	2489.00	0.00
F-test			H0: $\sigma_{MLR} \leq \sigma_{DML}$; H1: $\sigma_{MLR} > \sigma_{DML}$			
Wilcoxon signed-rank test			H0: $\lvert\mu_{MLR}\rvert \leq \lvert\mu_{DML}\rvert$; H1: $\lvert\mu_{MLR}\rvert > \lvert\mu_{DML}\rvert$			



Causal effect of influencing factors	T	D	W	UV	PM_{10}	$PM_{2.5}$	SO_2	NO_2	CO	O_3	MI	MO	WI	DI	SI
θ_T	0.400 *	0.376 *	0.012	0.361	0.112	0.208	-0.082	-0.100	-0.035	0.544 **	0.336	0.305	0.111	-0.104	0.361
θ_D	0.140	0.134	0.296	-0.182	-0.150	-0.194	-0.208	-0.202	-0.203	0.102	0.079	0.079	-0.190	-0.080	-0.020
θ_W	0.049	-0.030	-0.123	-0.131	0.263	0.340	0.071	0.362	0.486 *	0.102	0.125	0.170	0.036	-0.082	0.032
θ_{UV}	0.214	0.164	0.051	0.172	0.038	0.090	0.023	0.206	0.160	0.183	0.172	0.144	0.255	0.009	0.103
$\theta_{PM_{10}}$	0.000	-0.056	0.108	0.187	0.114	-0.012	0.098	-0.313	-0.005	0.217	-0.017	-0.086	0.378 *	-0.162	0.197
$\theta_{PM_{2.5}}$	0.170	0.105	0.292	0.084	0.274	0.188	0.320	0.256	0.286	-0.101	0.343	0.277	-0.050	-0.092	0.274
θ_{SO_2}	0.216	0.075	-0.373 *	0.152	0.061	-0.019	-0.088	-0.053	0.131	0.105	0.022	0.040	-0.080	-0.055	0.160
θ_{NO_2}	0.015	0.015	0.108	-0.302	0.166	0.221	0.314	0.153	0.101	-0.117	0.269	0.266	-0.092	-0.000	0.142
θ_{CO}	-0.157	-0.135	0.002	-0.330	0.514 **	0.548 **	0.065	0.268	0.188	-0.195	0.269	0.288	-0.175	-0.263	0.168
θ_{O_3}	0.146	0.073	0.211	0.289	0.121	0.119	-0.019	-0.251	-0.042	0.278	-0.056	-0.047	0.297	-0.237	0.010

* $0.05 > p \geqslant 0.01$ ** $p < 0.01$

Fig. 6. Spearman correlation coefficient between environmental causal effects and regional environmental conditions.

The role of air pollution varies across different environments and is influenced by interactions between air pollutants. With the elevation of PM_{10} and $PM_{2.5}$ concentration levels, the causal effect coefficients of CO on COVID-19 increase at average rates of 0.01. (Fig. 7(d and e)). In summary, an elevation in airborne pollutant concentrations increases the contribution of pollutants to the spread of an outbreak. The environment in areas with more severe air pollution is more favorable to the spread of epidemics. Moreover, there is an interplay between meteorological conditions and air pollution. In this study, it was

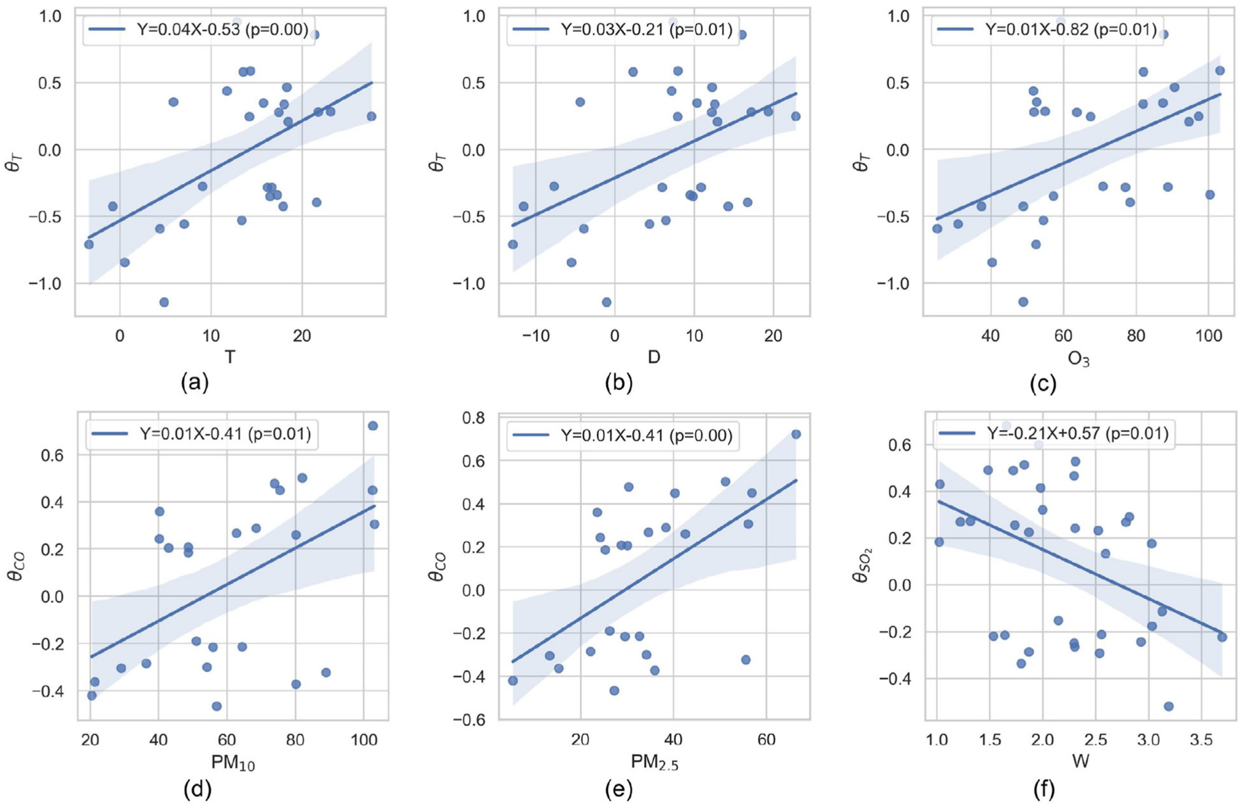


Fig. 7. Trends in the causal effects of environmental factors on COVID-19. (a–c): Changing trend of T's causal effect on COVID-19 at different regional mean temperatures, mean dew point temperature, and average O_3 concentration respectively; (d–e): Changing trend of CO's causal effect on COVID-19 at different regional average PM_{10} concentration and average $PM_{2.5}$ concentration respectively; (f): Changing trend of SO_2's causal effect on COVID-19 at different regional average wind speed.

shown that an increase in wind speed was beneficial in reducing the effect of SO_2 on the development of the outbreak (Fig. 7 (f)). As wind speed increases, the causal effect of SO_2 on COVID-19 decreases by 0.21 on average.

4. Discussion & conclusion

4.1. Complex nonlinear relationship between environment and COVID-19 transmission

It is unsuitable to solely attribute environmental factors as having a negative or positive correlation with COVID-19 transmission. The connection between the two is complex and non-linear. In different regional environments, the mechanisms by which environmental factors affect COVID-19 transmission will change. During the spread of the pandemic, there is an interaction between meteorological factors and air pollution. Similarly, Zhang et al. (2020) also highlighted the correlation between ambient temperature and air pollution on epidemic transmission. The combination of different factors complicates the transmission mechanism of COVID-19. Epidemics are possible in various environments. In fact, although COVID-19 tends to have more frequent outbreaks in winter, there is no evidence to suggest that the pandemic disappears with the rise in temperature during the summer. On the contrary, this study suggests that in regions with higher average temperatures, the increase in temperature favors the development of the pandemic. Other research suggests that high temperatures and heat waves increase the risk of COVID-19 transmission (Lian et al., 2023). Multiple evidences suggest that the spread of COVID-19 is a highly non-linear process. The role of environmental factors in COVID-19 changes dynamically. This also demonstrates the need for outbreak prevention and control. It is not feasible to rely on changes in the environment to contain the outbreak.

In addition, the role of environmental factors passed significance testing in a small proportion of the 261 outbreaks that we analyzed. This suggests that environmental factors were not the dominant cause of the widespread outbreak in China in 2022. Moreover, it is crucial to acknowledge that variations in the contributions of factors do not necessarily align with disparities in regional environments concerning COVID-19 transmission. The correlation between regional environments and the causal effects of environmental factors is also constrained. The intricate mechanisms driving the transmission of COVID-19 cannot be solely ascribed to environmental variations. Indeed, the influencing factors in the spread of COVID-19 are highly intricate. Regional social economic development level significantly affected the severity of COVID-19 (McGowan & Bambra, 2022). The interplay of these pivotal factors also shapes the role of the environment in the dissemination of the pandemic. It is imperative to distinguish the mechanisms of pandemic transmission in diverse scenarios. When forecasting the progression of the pandemic, due consideration must be given to disparities in environmental and socio-economic factors.

4.2. Limitations and future works

Our study currently has several limitations. The use of daily cases as a metric for the transmissibility of COVID-19 presents certain limitations. While this measure provides a direct and immediate representation of the number of daily cases, it does not account for the complex dynamics of transmission within a population. A more specialized metric often employed to estimate viral transmissibility is the effective reproduction number, which represents the average number of secondary cases produced by a primary case. The effective reproduction number should be used in place of daily cases as an indicator of the spread of the epidemic in future analyses. This would allow for a more robust interpretation of the virus's transmissibility over time and across different population dynamics.

Outbreaks can be influenced by a variety of factors, including regional vaccination status, demographics, socioeconomic conditions, and cultural practices (McGowan & Bambra, 2022). Additional factors need to be considered to more accurately quantify transmission mechanisms. The heterogeneity of the role of environmental factors also needs to be further explored by including more factors. In the future, the introduction of regionally differentiated roles of environmental factors in outbreak simulation and forecasting models will help to more accurately characterize outbreak transmission mechanisms, identify high-risk areas and populations, reduce the likelihood of major outbreaks, and minimize public health and economic impacts. In addition, our study is a time series analysis that does not fully account for spatial correlations. In the future, spatial-temporal models should be developed to analyze the role of environmental factors.

4.3. Conclusion

Globalization processes and climate change increase the risk of future pandemics. There is a growing need to study how the environment affects the spread of epidemics. Our study rigorously analyzes the causal relationship between environmental factors and COVID-19 transmission in 261 outbreaks across diverse Chinese cities in 2022, utilizing a robust double machine learning causal inference model. Through a comparative analysis with the outcomes of the traditional multiple linear regression model and an exploration of the heterogeneity in the causal effect of environmental factors, the following primary conclusions were drawn: (1) Environmental factors exhibit a confounding influence on the transmission of COVID-19. (2) In the presence of confounding effect, it is biased to directly employ the multiple linear regression model to estimate the effect of environmental factors. This bias results in an overestimation of the role of environmental factors to some extent. (3) The impact of environmental factors on COVID-19 transmission is heterogeneous, exhibiting variations across different regions. Both hot and humid, as well as cold and dry environments, can contribute to the risk of COVID-19 transmission. The higher the pollutant concentration, the more pronounced the impact on the spread of the epidemic. Additionally, meteorological factors and air pollution factors interact in the spread of the epidemic. (4) Ultimately, we posit that the relationship between the environment and COVID-19 is intricate and nonlinear. Successful prevention and prediction of future epidemics necessitate the consideration of regional environmental differences.

CRediT authorship contribution statement

Zihan Hao: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Data curation. **Shujuan Hu:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Jianping Huang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Jiaxuan Hu:** Writing – review & editing, Data curation. **Zhen Zhang:** Writing – review & editing, Data curation. **Han Li:** Writing – review & editing. **Wei Yan:** Writing – review & editing.

Declaration of competing interest

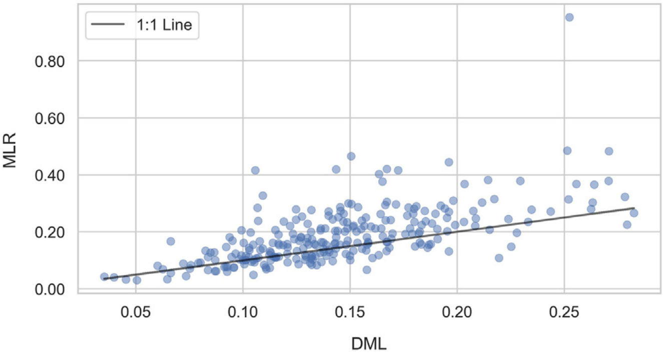
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Self-supporting Program of Guangzhou Laboratory (SRPG22-007), the National Key Research and Development Program of China (2023YFC3503400), and the Gansu Province Intellectual Property Project under Grant (22ZSCQD02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2024.06.005.



References

Ahmad, F., Almuayqil, S. N., Humayun, M., Naseem, S., Khan, W. A., & Junaid, K. (2020). Prediction of COVID-19 cases using machine learning for effective public health management. *Computers, Materials & Continua, 66*, 2265–2282. https://doi.org/10.32604/cmc.2021.013067

Ali, N., & Islam, F. (2020). The effects of air pollution on COVID-19 infection and mortality—a review on recent evidence. *Frontiers in Public Health, 8*, Article 580057. https://doi.org/10.3389/fpubh.2020.580057

Bach, P., Chernozhukov, V., Kurz, M. S., & Spindler, M. (2022). DoubleML: An object-oriented implementation of double machine learning in Python. *Journal of Machine Learning Research, 23*(1), 2469–2474. https://doi.org/10.5555/3586589.3586642

Baidu Migration. Retrieved from https://qianxi.baidu.com/#/. Accessed May 20, 2024.

Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the Total Environment, 728*, Article 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Cao, H., Li, B., Gu, T., Liu, X., Meng, K., & Zhang, L. (2021). Associations of ambient air pollutants and meteorological factors with COVID-19 transmission in 31 Chinese provinces: A time series study. *Inquiry: The Journal of Health Care Organization, Provision, and Financing, 58*, Article 00469580211060259. https://doi.org/10.1177/00469580211060259

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

China Environmental Monitoring Center. Retrieved from https://www.cnemc.cn/sssj/. Accessed May 20, 2024.

Colangelo, K., & Lee, Y. Y. (2020). *Double debiased machine learning nonparametric inference with continuous treatments*. arXiv preprint arXiv:2004.03036.

Farbmacher, H., Huber, M., Lafférs, L., Langen, H., & Spindler, M. (2022). Causal mediation analysis with double machine learning. *The Econometrics Journal, 25*(2), 277–300. https://doi.org/10.1093/ectj/utac003

Ford, J. D., Zavaleta-Cortijo, C., Ainembabazi, T., Anza-Ramirez, C., Arotoma-Rojas, I., Bezerra, J., Chicmana-Zapata, V., Galappaththi, E. K., Hangula, M., Kazaana, C., & Lwasa, S. (2022). Interactions between climate and COVID-19. *The Lancet Planetary Health, 6*(10), e825–e833. https://doi.org/10.1016/S2542-5196(22)00174-7

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour, 5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Han, Y., Huang, J., Li, R., Shao, Q., Han, D., Luo, X., & Qiu, J. (2022). Impact analysis of environmental and social factors on early-stage COVID-19 transmission in China by machine learning. *Environmental Research, 208*, Article 112761. https://doi.org/10.1016/j.envres.2022.112761

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N., 2023. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47 (Accessed on 22 June 2024).

Huang, J., Liu, X., Zhang, L., Zhao, Y., Wang, D., Gao, J., Lian, X., & Liu, C. (2021). The oscillation-outbreaks characteristic of the COVID-19 pandemic. *National Science Review, 8*(8), nwab100. https://doi.org/10.1093/nsr/nwab100

Huang, J., Zhang, L., Liu, X., Wei, Y., Liu, C., Lian, X., Huang, Z., Chou, J., Liu, X., Li, X., Yang, K., Wang, J., Liang, H., Gu, Q., Du, P., & Zhang, T. (2020). Global prediction system for COVID-19 pandemic. *Science Bulletin, 65*(22), 1884–1887. https://doi.org/10.1016/j.scib.2020.08.002

Jacobs, J. L., Haidar, G., & Mellors, J. W. (2023). COVID-19: Challenges of viral variants. *Annual Review of Medicine, 74*, 31–53. https://doi.org/10.1146/annurev-med-042921-020956

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature, 582*(7812), 389–394. https://doi.org/10.1038/s41586-020-2284-y

Jiang, Y., Wu, X. J., & Guan, Y. J. (2020). Effect of ambient air pollutants and meteorological variables on COVID-19 incidence. *Infection Control & Hospital Epidemiology, 41*(9), 1011–1015. https://doi.org/10.1017/ice.2020.222

Li, H., Huang, J., Lian, X., Zhao, Y., Yan, W., Zhang, L., & Li, L. (2023). Impact of human mobility on the epidemic spread during holidays. *Infectious Disease Modelling, 8*(4), 1108–1116. https://doi.org/10.1016/j.idm.2023.10.001

Li, H., Xu, X. L., Dai, D. W., Huang, Z. Y., Ma, Z., & Guan, Y. J. (2020). Air pollution and temperature are associated with increased COVID-19 incidence: A time series study. *International Journal of Infectious Diseases, 97*, 278–282. https://doi.org/10.1016/j.ijid.2020.05.076

Lian, X., Huang, J., Li, H., He, Y., Ouyang, Z., Fu, S., Zhao, Y., Wang, D., Wang, R., & Guan, X. (2023). Heat waves accelerate the spread of infectious diseases. *Environmental Research, 231*, Article 116090. https://doi.org/10.1016/j.envres.2023.116090

Liu, X., Huang, J., Li, C., Zhao, Y., Wang, D., Huang, Z., & Yang, K. (2021). The role of seasonality in the spread of COVID-19 pandemic. *Environmental Research, 195*, Article 110874. https://doi.org/10.1016/j.envres.2022.112690

Malik, J. A., Ahmed, S., Mir, A., Shinde, M., Bender, O., Alshammari, F., Ansari, M., & Anwar, S. (2022). The SARS-CoV-2 mutations versus vaccine effectiveness: New opportunities to new challenges. *Journal of infection and public health, 15*(2), 228–240. https://doi.org/10.1016/j.jiph.2021.12.014

Markov, P. V., Ghafari, M., Beer, M., et al. (2023). The evolution of SARS-CoV-2. *Nature Reviews Microbiology, 21*(6), 361–379. https://doi.org/10.1038/s41579-023-00878-2

Marziano, V., Guzzetta, G., Menegale, F., Sacco, C., Petrone, D., Mateo Urdiales, A., Del Manso, M., Bella, A., Fabiani, M., Vescio, M. F., & Riccardo, F. (2023). Estimating SARS-CoV-2 infections and associated changes in COVID-19 severity and fatality. *Influenza and Other Respiratory Viruses, 17*(8), Article e13181. https://doi.org/10.1111/irv.13181

McGowan, V. J., & Bambra, C. (2022). COVID-19 mortality and deprivation: Pandemic, syndemic, and endemic health inequalities. *The Lancet Public Health, 7*(11), e966–e975. https://doi.org/10.1016/S2468-2667(22)00223-7

Milicevic, O., Salom, I., Rodic, A., Markovic, S., Tumbas, M., Zigic, D., Djordjevic, M., & Djordjevic, M. (2021). PM2. 5 as a major predictor of COVID-19 basic reproduction number in the USA. *Environmental Research, 201*, Article 111526. https://doi.org/10.1016/j.envres.2021.111526

Paraskevis, D., Kostaki, E. G., Alygizakis, N., Thomaidis, N. S., Cartalis, C., Tsiodras, S., & Dimopoulos, M. A. (2021). A review of the impact of weather and climate variables to COVID-19: In the absence of public health measures high temperatures cannot probably mitigate outbreaks. *Science of the Total Environment, 768*, Article 144578. https://doi.org/10.1016/j.scitotenv.2020.144578

Poirier, C., Luo, W., Majumder, M. S., Liu, D., Mandl, K. D., Mooring, T. A., & Santillana, M. (2020). The role of environmental factors on transmission rates of the COVID-19 outbreak: An initial assessment in two spatial scales. *Scientific Reports, 10*(1), Article 17002. https://doi.org/10.1038/s41598-020-74089-7

Qi, H., Xiao, S., Shi, R., Ward, M. P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., & Zhang, Z. (2020). COVID-19 transmission in mainland China is associated with temperature and humidity: A time-series analysis. *The Science of the Total Environment, 728*, Article 138778. https://doi.org/10.1016/j.scitotenv.2020.138778

Robinson, P. M. (1988). Root-N-Consistent semiparametric regression. *Econometrica, 56*(4), 931–954. https://doi.org/10.2307/1912705

Sun, C., Chao, L., Li, H., Hu, Z., Zheng, H., & Li, Q. (2022). Modeling and preliminary analysis of the impact of meteorological conditions on the COVID-19 epidemic. *International Journal of Environmental Research and Public Health, 19*(10), 6125. https://doi.org/10.3390/ijerph19106125

Wang, Q., Dong, W., Yang, K., Ren, Z., Huang, D., Zhang, P., & Wang, J. (2021). Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors. *International Journal of Infectious Diseases, 105*, 675–685. https://doi.org/10.1016/j.ijid.2021.03.014

Wang, C. C., Prather, K. A., Sznitman, J., Jimenez, J. L., Lakdawala, S. S., Tufekci, Z., & Marr, L. C. (2021). Airborne transmission of respiratory viruses. *Science, 373*(6558), Article eabd9149. https://doi.org/10.1126/science.abd9149

Weaver, A. K., Head, J. R., Gould, C. F., Carlton, E. J., & Remais, J. V. (2022). Environmental factors influencing COVID-19 incidence and severity. *Annual Review of Public Health, 43*, 271–291. https://doi.org/10.1146/annurev-publhealth-052120-101420

WHO. (2023). *Coronavirus disease (COVID-19) dashboard*. https://covid19.who.int/. (Accessed 4 October 2023).

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *The Science of the Total Environment, 724*, Article 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Xu, R., Rahmandad, H., Gupta, M., DiGennaro, C., Ghaffarzadegan, N., Amini, H., & Jalali, M. S. (2021). Weather, air pollution, and SARS-CoV-2 transmission: A global analysis. *The Lancet Planetary Health, 5*(10), e671–e680. https://doi.org/10.1016/S2542-5196(21)00202-3

Yang, X. D., Li, H. L., & Cao, Y. E. (2021). Influence of meteorological factors on the COVID-19 transmission with season and geographic location. *International Journal of Environmental Research and Public Health, 18*(2), 484. https://doi.org/10.3390/ijerph18020484

Yao, Y., Pan, J., Liu, Z., Meng, X., Wang, W., Kan, H., & Wang, W. (2020). No association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *European Respiratory Journal, 55*(5), Article 2000517. https://doi.org/10.1183/13993003.00517-2020

Yin, C., Zhao, W., & Pereira, P. (2022). Meteorological factors' effects on COVID-19 show seasonality and spatiality in Brazil. *Environmental Research, 208*, Article 112690. https://doi.org/10.1016/j.envres.2022.112690

Zhang, Z., Xue, T., & Jin, X. (2020). Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities. *The Science of the Total Environment, 741*, Article 140244. https://doi.org/10.1016/j.scitotenv.2020.140244

Zheng, R., Xu, Y., Wang, W., Ning, G., & Bi, Y. (2020). Spatial transmission of COVID-19 via public and private transportation in China. *Travel Medicine and Infectious Disease, 34*, Article 101626. https://doi.org/10.1016/j.tmaid.2020.101626

Zhu, Y., Xie, J., Huang, F., & Cao, L. (2020). Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. *The Science of the Total Environment, 727*, Article 138704. https://doi.org/10.1016/j.scitotenv.2020.138704