# MECHANISM-EMPOWERED MULTIVARIATE TIME SERIES FORECASTING MODEL: APPLICATION TO TUBERCULOSIS PREDICTION

#### Anonymous authors Paper under double-blind review

#### ABSTRACT

Among the current global health challenges, tuberculosis, as a highly contagious chronic disease, remains one of the major public health problems worldwide. Despite significant progress made in the past decades, new challenges, including systematic and effective downscaling, accurate prediction of disease incidence, and implementation of source reduction measures, have added to the difficulty of tuberculosis control. In view of the limitations of the recently proposed EIGHT prediction models in terms of prediction accuracy, this study adopts the Learnable Decomposition and Dual Focus Module Model (Leddam) and then introduces a novel mechanism-supported multivariate spatiotemporal series framework, termed LCHHA-Leddam, to address the challenges in tuberculosis forecasting through an investigation of coal power generation in China. This framework substantially simplifies the complexity of tuberculosis prediction, enhances accurate dimensionality reduction, and improves traceability. It also enhances the explanatory power and accuracy of the Leddam model in the field of tuberculosis prediction. This study provides a fresh perspective for enhancing epidemic forecasting and exploring source reduction measures for industrial activities, demonstrating the feasibility of AI-assisted public health strategies and green production.

006

012 013

014

015

016

017

018

019

020

021

024

025

026

027

#### 1 INTRODUCTION

033 Climate change and environmental pollution have become one of the most important factors affecting public health with the acceleration of urbanization, industrialization, and economic development. Recent research has shown that environmental pollutants lead to approximately nine million premature deaths per year worldwide (Landrigan et al., 2018; Stanaway et al., 2018), and climate change is expected to cause approximately four million additional deaths per year (Zhao et al., 2021). Most 037 studies have explored the relationship between diseases and environmental factor with respect to health risk (Asadgol et al., 2019, Danziger et al., 2022). Fewer studies comprehensively address the concurrent effects of climate change and environmental pollution on disease occurrence and 040 implement source reduction measures. Risk factors for many human diseases and targets for preven-041 tive healthcare strategies are commonly identified based on epidemiological analysis (Chowdhury et 042 al., 2018). However, a variety of factors/processes jointly affect the health outcome or its indicator, 043 suggesting the inability of epidemiological models to simultaneously and systematically consider, 044 and in many cases even impossible to solve hundreds of direct- and indirect environmental factors and processes (Braun et al., 2016).

Life cycle human health assessment (LCHHA) is a systematic and internationally standardized approach for simultaneously and efficiently quantifying the direct and indirect human health footprint of a targeted product, process, or activity from the beginning to the final phase (ISO 14040, 2006; ISO 14044, 2006). In contrast to epidemiological studies, LCHHA quantifies the potential human risk of a regional population via the dose-response curves derived from laboratory animal experiments and extrapolation factors for interspecies differences (Huijbregts et al., 2010). As animal experiments databased LCHHA analysis has high uncertainty because of interspecies differences in physiology and metabolism, the combination of methods to understand better the sources' contribution to potential human health impacts is highly needed. Systematically and effectively implementing source reduction

measures and key factor identification to protect public health from environmental pollutants has
recently become possible. For instance, industrial activity-environmental pollution-carcinogenic
disease relationships calculated by LCHHA and epidemiological cancer incidences can now be further
explored (Chen et al., 2019; Jia et al., 2022). However, relevant research on infectious disease and its
prediction is extremely rare at the global level.

Interestingly, machine learning, as a popular approach for predictive modeling, has been extensively 060 used in disease prediction to achieve early warning and disease intervention because of its capacities 061 on analyzing large amounts of data, handling complex relationships, and learning and adapting over 062 time (Huang et al., 2022; Krittanawong et al., 2020). Multivariate time series are datasets with 063 multiple dimensions, where each dimension represents a separate univariate time series, e.g., climate 064 characteristics (Zhang & Yan, 2023). The prediction of infectious diseases is a multivariate time series prediction as multiple factors/processes collectively influence infectious diseases. Predicting 065 these time series helps to improve decision-making in various application areas such as weather, 066 energy, and finance (Angryk et al., 2020; Demirel et al., 2012; Patton, 2013). Consequently, the 067 incidence of infectious diseases can now be predicted, and the reduction of sources for human health 068 protection can now be explored by integrating LCHHA, climate change, and multivariate time series 069 forecasting models. Tuberculosis (TB) is used as a target disease to achieve the aforementioned goal; it is one of the world's leading infectious disease killers (WHO, 2022) and is involved in the 2030 071 Agenda for the World's Sustainable Development (United Nations, 2015). A quarter of the world's 072 population is infected with Mycobacterium TB (Suárez et al., 2019). In 2021, the number of newly 073 diagnosed TB patients in China was 0.78 million, ranking third among countries and accounting for 074 7.4% of global TB cases (WHO, 2022). In China, the incidence and death of TB rank first within 075 the Class A and Class B notifiable infectious diseases. Two-thirds of the infected individuals are concentrated in the age group (15-70) with solid production capacity. Hence, the issue is regarded as 076 a major public health problem in China. 077

078 TB transmission occurs almost exclusively because of aerosolized particles (Dinkele et al., 2022) 079 and climate conditions such as temperature through the influence on TB bacteria spread and survival, human immune response, and climate patterns (Chong et al., 2022; Chormare and Kumar, 2022). 081 However, the joint effects of climate change and environmental pollution on TB are also rarely assessed. As the world's largest coal consumer and producer, China is confronted with the serious 082 environmental problem caused by coal burning (Ma et al., 2021), such as a global environmental issue 083 of haze over China. More than half of greenhouse gas and aerosolized particles are produced from coal 084 consumption in China (Ma et al., 2021; Shi et al., 2017). The coal power industry in China dominated 085 national coal consumption since 2000 (>40%), although renewable energy has been encouraged by 086 the Chinese government (Chinese Energy Statistic Yearbook, 2001-2021). Approximately  $5.3 \times 10^5$ 087 premature deaths per year were recorded because of air pollution generated by coal burning in China 088 (Yun et al., 2021). This research thereby focuses on the environmental pollutants released by the coal 089 power industry in China and maps and predicts the incidence of TB caused by climate change and 090 environmental pollution. Worth noting, TB incidence caused by climate factors and environmental 091 pollutants is a complex and systematic phenomenon. Direct and indirect pollutants emitted from 092 the coal power industry during its whole life cycle stages for LCHHA analysis generally involve hundreds of contaminants. Therefore, the intervention of low-dimensional feature engineering is 093 highly needed. Creating low-dimensional features is crucial for data scientists to develop models and minimize model variance (Xuan et al., 2019; Rizgar et al., 2020). Domain experts typically undertake this feature engineering task manually, guided by specific domain knowledge. This method 096 is known as mechanism-driven dimensionality reduction (MR) (Lan & Susan, 2019; Ting et al., 2010). Data-driven dimensionality reduction (DR), however, selects feature engineering methods 098 based solely on available data. DR effectively extracts relevant features without relying on domain expertise, serving as a form of automated data cleaning (Neoklis et al., 2017). Despite this, MR 100 remains essential since numerous parameters can significantly influence the performance of DR 101 algorithms and modelling techniques, while MR can scientifically manage the relationships between 102 parameters. This study compares MR and DR in terms of interpretability and accuracy, with LCHHA 103 as a representative of MR. It also introduces a novel mechanism-enabled multivariate spatiotemporal 104 sequence framework that effectively improves the accuracy and interpretability of infectious disease prediction and helps public health managers identify effective management and source reduction 105 strategies. 106



Figure 1: The prediction and source reduction approach of TB incidence caused by climate change and environmental pollution released by coal power generation.

137 138

139

133

#### 2 METHODOLOGY

#### LCHHA OF COAL POWER GENERATION IN CHINA 2.1

140 The LCHHA of coal power generation in China was estimated based on ISO 14040 series standards 141 (ISO 14040, 2006). Six subcategories (i.e., particulates formation, ozone formation, ozone layer 142 depletion, global warming, carcinogens, and non-carcinogens) are involved for direct and indirect 143 LCHHA (Fig.1). The selection of subcategories in this study is consistent with the commonly 144 used LCHHA models established in Europe (Sala et al., 2012), North America (Bare, 2011), and 145 World (Bulle et al., 2019). Regarding the global scale impact categories of ozone depletion and 146 climate change, the characterization factors published by the World Meteorological Organization (WMO, 2011) and Intergovernmental Panel on Climate Change (IPCC, Arias et al., 2021) are used 147 for quantification. For the rest subcategories, the SDU model, a life cycle impact assessment model 148 developed for China's life cycle assessment analysis (Yang et al., 2019; Ren et al., 2022), is applied. 149 The life cycle inventories of coal power generation at industry level are taken from previously 150 published reference reported by our research group (Li et al., 2024 a, b). The system boundary is 151 set to cradle-to-gate, and 1 kWh coal power generation is used as the functional unit. All inputs 152 and outputs for each life cycle stage of coal power generation are utilized in this study except for 153 infrastructure because of data limitations and their negligible impact (Ecoinvent Centre, 2016). The 154 uses of LCHHA in this study are: 1) to evaluate the human health footprint of coal-fueled power 155 industry initiatives at the national level; 2) to identify the key factors for MR and environmental 156 improvement in China; 3) to provide a training set for predicting the TB incidence caused by direct 157 environmental pollution released by the coal power industry in China; and 4) to enhance the precision and intelligibility of spatio-temporal sequence models. The formula of LCHHA is as follows, 158

159

160 1(

$$\text{LCHHA} = \sum_{i=1}^{n} P_{q_i} \times \alpha_i \tag{1}$$

162 163  $p_{1m}$ 164  $p_{2m}$ 165  $q_{1} \begin{vmatrix} \frac{1}{2} \\ \frac{p_{i1}}{p_{i+1,1}} \\ \vdots \end{vmatrix} + q_{2} \begin{vmatrix} \frac{1}{2} \\ \frac{p_{i2}}{p_{i+1,2}} \\ \vdots \end{vmatrix} + \dots + q_{m}$ 166  $\frac{p_{im}}{p_{i+1,m}}$ 167 168 169 170  $p_{nm}$ 171 172  $p_{11}$  $p_{12}$  $p_{1m}$ 173  $p_{21}$  $p_{22}$  $p_{2m}$ 174  $\frac{p_{i1}}{p_{i+1,1}}, \frac{p_{i2}}{p_{i+1,2}}, \dots,$  $\vdots$ 175  $p_{im}$ 176  $p_{i+1,m}$ 177 178 179  $p_{n1}$  $p_{nm}$ 180 181 = Pq(2)182 183 
$$\begin{split} \alpha_i = \mathbf{X} \mathbf{F}_i \times \mathbf{E} \mathbf{F}_i \times \mathbf{D} \mathbf{F}_i = \left( \frac{\mathbf{B} \mathbf{A} \mathbf{F} \times \mathbf{P} \mathbf{R} \mathbf{O} \mathbf{D} \times \mathbf{P} \mathbf{O} \mathbf{P}}{\mathbf{M} \mathbf{A} \mathbf{S} \mathbf{S}} + \frac{\mathbf{I} \mathbf{N} \mathbf{H} \times \mathbf{P} \mathbf{O} \mathbf{P}}{V_{\mathrm{air}}} \right) \\ \times \frac{0.5}{365 \times \mathbf{L} \mathbf{T} \times \mathbf{B} \mathbf{W} \times \mathbf{E} \mathbf{D}_{50}} \times \mathbf{D} \mathbf{F}_{\mathbf{N} \mathbf{C} / \mathbf{C}, i} \end{split}$$
185 186 187 188 189  $+ \frac{\mathrm{EF}_{\mathrm{PF},i}}{\mathrm{EF}_{\mathrm{PM2.5}}} \times \mathrm{DF}_{\mathrm{PF}}$ 190  $+ \frac{\mathrm{EF}_{\mathrm{OF},i}}{\mathrm{EF}_{\mathrm{NOx}}} \times \mathrm{DF}_{\mathrm{OF}}$ 191 192 193  $+ \text{EESC}_i \times \text{UVB}_i \times \text{EF}_{\text{OD},i} \times \text{DF}_{\text{OD}}$ 194  $+ \frac{\text{GWP}_{\text{GW},i}}{\text{GWP}_{\text{CO}_2}} \times \text{DF}_{\text{GW}}$ 195 (3)196

where  $P_{qi}$  and are the life cycle emission inventory (kg<sub>emission</sub>) and health damage (Daly/kg<sub>emission</sub>) of substance *i*. The meanings of rest symbols are listed in Table 3.

#### 2.2 LCHHA-LEDDAM FRAMEWORK

197

198

199 200

201

211

212

202 The Learnable Decomposition and Dual Focus Module (Leddam) was selected for this investigation 203 due to its notable enhancement in predictive precision compared to the recently identified EIGHT 204 predictive models (Yu et al., 2024). Fig.1 presents the prediction and source reduction approach 205 of TB incidence caused by climate change and environmental pollution released from coal power 206 generation in China. The monthly spatiotemporal meteorological elements, the LCHHA of direct 207 66pollutants from coal power generation, and the epidemiological TB incidence in the 2004–2018 period are used for training the Leddam model. Spatiotemporal meteorological elements, including 208 wind speed, rainfall, sunshine duration, temperature, and humidity data are selected as the indicators 209 of climate change because these factors are measurable and easier to track changes over time. 210

$$\text{Percentage}_{i} = \frac{P_{q_{i}} \times \alpha_{i}}{\text{LCHHA}_{\text{All}}} \times 100\%$$
(4)

214 
$$X_{\text{Seasonal}} = X_{\text{Seasonal}}[i_1, i_2, i_3, i_4, i_5]$$
(5)

The analysis of the impact of LCHHA on Leddam begins with a comprehensive examination of the LCHHA effects associated with 34 categories of pollutants. The impact of each pollutant on LCHHA



Figure 2: Flow chart of experiment

is quantified. The overall impact, designated as LCHHA<sub>All</sub>, is determined by the summation of all pollutants' LCHHA impact. Index i represents a distinct pollutant. Subsequently, the percentage contribution of each pollutant to the total is evaluated. These calculations are instrumental in developing an enhanced version of  $X_{Seasonal}$ , where  $X_{Seasonal}$  of the top 5 contributed pollutants are indexed as  $i_1$ ,  $i_2$ ,  $i_3$ ,  $i_4$ , and  $i_5$ .

2.3 DATA SOURCES

Climate data from 699 meteorological stations in China were obtained from the Meteorological Observation Information Centre of China to assess the relationship among various Meteorological elements (CMA, 2021). The spatiotemporal and annual average life cycle inventory of coal power generation was obtained from the CPLCID® (Zhang et al., 2016). For coal power generation in the CPLCID<sup>®</sup>, the data on spatiotemporal coal consumption for coal power generation, import and export coal amount, direct air emission amount of SO<sub>2</sub>, NOx, and particulates, desulfurization gypsum and coal ash utilization rate, hospital beds, medical technicians, GDP per capita, population density, water consumption, and technological transformation were obtained from various reference documents (Annual Statistic Report on Environment in China, 2004-2021; China Electric Power Industry Annual Development Report, 2006-2020; China Statistical Yearbook, 2001-2021; Chinese Energy Statistic Yearbook, 2001-2021). The study by Hong et al. (2019) was also used to derive the data on national average methane and carbon dioxide emitted from low-, high-, outburst-, and open-pit coal mining sites and the life cycle inventory of coal-seam gas utilization. The data on coal quality at the national level were taken from Liu et al. (2015). The data on the primary inventory of coal-based electricity generation were taken from references (Cui et al., 2012; Xu et al., 2015; Zhao et al., 2015), and further combined with spatiotemporal regional statistical data according to the research of Hong et al. (2015). The data collection mentioned above were allowed for the building of a macro-level life cycle inventory of coal power generation in China (Li et al., 2024 a, b). Spatiotemporal TB incidence in China was obtained from China Health Statistical Yearbook (2004-2021) and Public Health Science Data Center (2023), whereas annual TB incidence and burden in China were taken from the Global Burden of Disease Study (IHME, 2020).

## 3 EXPERIMENT

Five downscaling algorithms were chosen based on the fact that dimensionality reduction can be
 categorized into two main types: linear mapping and nonlinear mapping. In this context, linear
 mapping was represented by Principal Component Analysis (PCA), while nonlinear mapping included
 kernel methods such as Kernel Principal Component Analysis (KPCA) and Kernel Fisher Discriminant



Figure 3: Contribution of dominant substances and processes to the annual LCHHA in China: a) substance; b) process

297

Analysis (KFDA), as well as manifold learning techniques like Isometric Mapping (ISOMAP) and 301 Locally Linear Embedding (LLE). The training sets include spatiotemporal annual epidemiological 302 TB incidence, LCHHA values, and meteorological elements at China's provincial level in the 303 2004–2018 period (n=5580). Epidemiological TB data involve spatiotemporal reported incidence 304 rate, incidence cases, and disease burden. Spatiotemporal meteorological elements including wind 305 speed, rainfall, sunshine duration, temperature, and humidity data are selected as the indicators of 306 climate change because these factors are measurable and easier to track changes over time. Spatial 307 average monthly meteorological elements at the province level integrated by daily surface climate data 308 of 699 meteorological stations in China since 2004 are used as original database. Data are randomly split and set into a training, validating and testing set with a ratio of 7:1:2. The mean squared error loss 309 (MSE), mean absolute error (MAE), and coefficient of determination (i.e., R-squared) are used for 310 aforementioned models screening. Fig.2 illustrates the process of our experiment. The LEADDM was 311 initially trained with a database containing all pollutants and meteorological variables. KPCA, PCA, 312 KFDA, LLE, and Isomap were applied to the pollutant data, creating a feature set Z. A new database, 313 combining meteorological data and Z, was then used to retrain the LEADDM. The best algorithm, 314 determined by R<sup>2</sup>, MSE, and MAE, identified the top five pollutant contributors, results shown in 315 Table 1. These, along with meteorological data, formed the RK+M[DM1] database. Similarly, LK+M 316 was created using LCHHA analysis. Both databases were used for further Leddam training.

317 318 319

320 321

322

#### 4 Results

#### . . .

## 4.1 ANALYSIS OF KEY CONTRIBUTORS THROUGH LCHHA

Fig.3 presents the most significant substances and processes contributing to the LCHHA. Mercury, carbon dioxide, nitrogen oxides, sulfur dioxide, and particulates generated by coal power generation



Table 1: Selection of dimensionality reduction method based on the database (P=All pollutants, M=Meteorological)

Figure 4: Each pollutant proportion in KPCA component

site appear to be the main contributors to the overall impact on human health. The potential impact generated by the rest of the substances (e.g., PAHs, lead, arsenic, zinc, and VOCs) and processes (e.g., wastewater and solid waste disposal, limestone, oil, hydrochloric acid, sulphuric acid, sodium hydroxide, coal washing, transportation, and raw materials consumed during coal mining stage) is low. Consequently, the pollutants emitted directly from the coal power generation stage are utilized for subsequent study. The overall LCHHA of coal power generation across China over 30 years has significantly decreased in the LCHHA of nitrogen oxides, sulfur dioxide, and particulates for the last decade, whereas an increasing tendency is observed for the other key substances (Fig.3a). Strict air pollution demands by the Chinese government (GB 13223-2011) and the annual increase in coal power generation can explain the variations.

# 4.2 ANALYSIS OF KEY CONTRIBUTORS THROUGH ML DIMENSIONALITY REDUCTION METHODS

Table 1 presents the performance of various dimensional reduction techniques applied to the pollutant meteorology dataset. The results indicate that Kernel PCA (P+M+KPCA) achieved the highest Test R<sup>2</sup> of 0.767, a lowest Test MAE of 0.266 and not the highest Test MSE of 0.264, demonstrating strong capability in capturing different pollutants relationships. KPCA is selected to do the analysis of key contributors. Fig.4 demonstrates the proportion of different pollutants in the KPCA component. Acenaphthene, Acenaphthylene, Anthracene, Benzo(a)pyrene, and Benzo(k)fluoranthene are the top five contributors to KPCA component, which is different from the analysis through LCHHA. Table 4 and Fig.8-11 in the appendix show the proportion of the various contaminants present in the alternative downscaling techniques. It is notable that no two of the five downscaling techniques vield an identical ranking of the first five key contributors, which suggests that the data-driven method lack sufficient interpretability. This phenomenon also demonstrates the limitations of existing dimensionality reduction methods in the field of AI for science. 

#### 4.3 COMPARISON BETWEEN MR AND DR

The results in Table 2 highlight the comparative performance of the different mechanisms in Leddam. The model that incorporates all pollutants along with meteorological data (P+M) exhibits the lowest

Table 2: Comparison between MR and DR (P=All pollutants, M=Meteorological, LK=Key factor
 selected by LCHHA, RK=Key factor selected by KPCA)

	P+M	KPCA+M	LK+M	RK+M
Test- $R^2$	-50.624	0.767	0.882	0.794
Test-MSE	0.083	0.264	0.096	0.146
Test-MAE	0.150	0.143	0.185	0.253

test  $R^2$  of -50.624, suggesting the need to select key pollutants. In contrast, the model based on 388 meteorology and LCHHA selection on pollutants (LK+M), has the highest test  $R^2$  of 0.882. It also 389 has the lowest Test MSE of 0.096 and lower Test MAE of 0.185 compared with KPCA selection 390 (RK+M) and KPCA reduction (KPCA+M). Notably, TB incidence caused by climate factors and 391 environmental pollutants is a complex and systematic phenomenon. LCHHA quantifies all the inputs 392 and outputs of coal power generation during its life cycle stages. Every single input has its life cycle and interacts as a part of the coal power generation system. Therefore, LCHHA analysis can 394 significantly reduce the complexity of the integrated assessment model applied in this study. The performance of LK+M underscores the effectiveness of the mechanism in enhancing Leddam model accuracy. Compared to data-driven methods, the higher accuracy of key factor locking via mechanism 396 models highlights deficiencies of data-driven methods in exploring feature interrelationships. It also 397 reveals that the data-driven downscaling algorithms still face multiple limitations, such as challenges 398 in handling high-dimensional sparse data, the the effectiveness of feature extraction, and the inherent 399 complexity of scientific problems. These limitations not only affect the depth and breadth of data 400 analysis but may also restrict the process of scientific discovery. 401

402 403

380 381 382

#### 4.4 SOURCE REDUCTION

404 405

#### 406 Fig. 3 shows that carbon, mercury, and particulate matter (e.g., nitrogen dioxide, sulfur dioxide. particulates) emitted from coal burning for power generation are the main contributors of the direct 407 LCHHA, which is the key contributor to TB prediction (Fig.1 and Table 2). The uptrend of life cycle 408 carbon emissions from China's coal power industry and actual annual average temperature (Appendix 409 A.1) are consistent with the IPCC opinion that global warming is mainly caused by greenhouse gas 410 emissions (Arias et al., 2021). As the annual average temperature increase in China correlates with 411 the annual humidity, rainfall, and sunshine duration changes (Fig. 5b), source reduction measures on 412 TB become possible through pollutants control, specifically on mercury, carbon dioxide, nitrogen 413 oxides, sulfur dioxide, and particulate emissions control. China continues to be the largest carbon 414 (IEA, 2022) and mercury (Hu et al., 2018) emitter in the world because of its high gross national 415 coal consumption. However, significant environmental control achievements have been attained in 416 recent years (Fig. 3 a). Approximately 2.2 Mt of NOx, 0.8 Mt of SO<sub>2</sub> and 643.1 t of atmospheric Hg 417 were emitted by the entire coal power generation chain in 2021. For the same year, in the entire coal power generation chain, approximately 40.1% of NOx, 69.9% of SO<sub>2</sub>, and 98.0% of Hg were emitted 418 from the coal power generation stage; the direct emissions of the aforementioned key substances 419 by coal power generation sites in China in 2013 and the extensive haze episodes in China were 420 nearly 89.8%, 96.6%, and 97.8%. The results indicate significant environmental control achievements 421 in national $SO_2$ , PM, and NOx control after setting a strict emission standard for thermal power 422 plants (GB 13223-2011) and increasing the efficiency of coal consumption year by year (Chinese 423 Energy Statistic Yearbook, 2001-2021). A similar trend of carbon control through coal consumption 424 efficiency improvement can also be seen in Appendix Fig.12. However, mercury control is weak, 425 and the current mercury mitigation policy in China's coal power industry remains unsystematic and 426 ineffective. The USEPA (1977) reported that 60% of mercury could be removed using coal-washing 427 technology, while the United Nations Environment Programme reported it to be 78% (UNEPA, 428 2002). Thus, the proper selection of coal-washing technology can significantly reduce the key factors and overall environmental burden. In 2020, the coal-washing rate in China was approximately 429 74.1%, lower than that in the United States (90%) (Ghosh, 2013). Clearly, increasing the national 430 coal-washing rate and the efficiency of coal consumption is another key factor for protecting human 431 health.

# 432 5 CONCLUSION

433 434

This study employs a novel LCHHA-Leddam model framework to predict the incidence of TB in 435 China, taking into account spatiotemporal climate change and environmental pollution caused by the 436 country's coal power industry. The LCHHA model can identify the connections between industrial 437 activities, direct and indirect environmental pollution, associated risks, and their impact on human 438 health, thereby facilitating the identification of source control measures. The results demonstrate 439 that the LCHHA-Leddam model outperforms the traditional Leddam model due to its superior 440 accuracy in identifying key factors through mechanism models. This suggests that a combination of mechanistic and data-driven models can significantly enhance the scientific accuracy of TB disease 441 predictions. Given that the LCHHA of coal power generation in China is primarily characterized 442 by direct emissions of atmospheric mercury, nitrogen oxides, sulfur dioxide, and particulate matter 443 during the coal combustion stage for power generation, it becomes feasible to implement source 444 reduction measures for TB disease control. These measures include controlling the aforementioned 445 pollutants, improving coal consumption efficiency, and increasing the national coal washing rate. 446

In summary, this study's novelty lies not only in providing scientific data for early TB warning but 447 also in its potential to facilitate effective management and source reduction. To improve the accuracy 448 of predicted outcomes and the efficacy of public health protection measures, it is necessary to consider 449 other factors that may influence epidemic diseases, such as high spatial-temporal resolution, genetics, 450 diet, lifestyle regularity, and population distribution characteristics. Furthermore, incorporating 451 interpretable machine learning algorithms, such as SHAP (SHapley Additive exPlanations), into our 452 framework could enhance interpretability and provide a more comprehensive understanding of the 453 factors influencing epidemic predictions. This holistic approach will better equip us to address the 454 environmental and health challenges of the future.

455 456 457

458

459

#### 6 REPRODUCIBILITY STATEMENT

To foster reproducibility, we will make our code available in supplementary materials and will make it public online after acceptance. We give details on our experimental protocol in Appendix

#### Reference

Angryk, R.A., Martens, P.C., Aydin, B., Kempton, D.J., Mahajan, S.S., Basodi, S., Ahmadzadeh, A.,
Cai, X., Boubrahimi, S.F., Hamdi, S.M., Schuh, M.A., Georgoulis, M.K., 2020. Multivariate time series dataset for space weather data analytics. Scientific Data.

Annual Statistic Report on Environment in China, 2004-2020. Ministry of Ecology and Environment of the People's Republic of China. https://www.mee.gov.cn/hjzl/sthjzk/zghjzkgb/.

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M.,
Plattner, G.-K., Rogelj, J., 2021. Climate Change 2021: The Physical Science Basis. Contribution
of Working Group14 I to the Fourth Assessment Report of the Intergovernmental Panel on Climate
Change; Technical Summary. https://www.ipcc.ch/report/fourth-assessment-report-working-group-i.

Asadgol, Z., Mohammadi, H., Kermani, M., Badirzadeh, A., Gholami, M., 2019. The effect of
climate change on cholera disease: The road ahead using artificial neural network. PloS one 14 (11),
e0224813. https://doi.org/10.1371/journal.pone.0224813.

Bare, J., 2011. TRACI 2.0: the tool for the reduction and assessment of chemical and other environmental impacts 2.0. Clean Techn Environ Policy 13 (5), 687–696. https://doi.org/10.1007/s10098010-0338-9.

Braun, J.M., Gennings, C., Hauser, R., Webster, T.F., 2016. What Can Epidemiological Studies Tell

Bulle, C., Margni, M., Patouillard, L., Boulay, A.-M., Bourgault, G., Bruille, V. de, Cao, V., Hauschild,

- 483 M., Henderson, A., Humbert, S., Kashef-Haghighi, S., Kounina, A., Laurent, A., Levasseur, A.,
- Liard, G., Rosenbaum, R.K., Roy, P.-O., Shaked, S., Fantke, P., Jolliet, O., 2019. IMPACT World+: a
- 485 globally regionalized life cycle impact assessment method. Int J Life Cycle Assess 24 (9), 1653–1674. https://doi.org/10.1007/s11367-019-01583-0.

Chen, B., Wang, M., Duan, M., Ma, X., Hong, J., Xie, F., Zhang, R., Li, X., 2019. In search of key: Protecting human health and the ecosystem from water pollution in China. Journal of Cleaner Production 228, 101–111. https://doi.org/10.1016/j.jclepro.2019.04.228.

China Electric Power Industry Annual Development Report 2006-2020, 2007. China electricity council. http://www.cec.org.cn.

- China Health Statistical Yearbook, 2004-2021. National Health Commission of the People's Republicof China. Peking Union Medical College Press.
- China Statistical Yearbook, 2001-2021. National Bureau of Statistics of the People's Republic of
   China. China Statistics Press. http://www.stats.gov.cn/tjsj./ndsj/.
- 496 Energy 2001-2021. Chinese Statistic Yearbook. National Bureau of 497 Statistics of the People's Republic of China. China Statistics Press. 498 https://data.cnki.net/v3/trade/Yearbook/Single/N2022060061?zcode=Z025. 499
- 500 Chong, K.C., Yeoh, E.K., Leung, C.C., Lau, S.Y.F., Lam, H.C.Y., Goggins, W.B., Zhao, S., Ran, J.,
- Chormare, R., Kumar, M.A., 2022. Environmental health and risk assessment metrics with special mention to biotransfer, bioaccumulation and biomagnification of environmental pollutants. Chemosphere 302, 134836. https://doi.org/10.1016/j.chemosphere.2022.134836.
- Chowdhury, R., Ramond, A., O'Keeffe, L.M., Shahzad, S., Kunutsor, S.K., Muka, T., Gregson, J.,
  Willeit, P., Warnakula, S., Khan, H., Chowdhury, S., Gobin, R., Franco, O.H., Di Angelantonio, E.,
  2018. Environmental toxic metal contaminants and risk of cardiovascular disease: systematic review and meta-analysis. BMJ (Clinical research ed.) 362, k3310. https://doi.org/10.1136/bmj.k3310.
- 508
   509
   509
   509
   509
   510
   510
   510
   510
   511
   512
   513
   514
   514
   514
   515
   516
   517
   518
   518
   518
   519
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
   510
- Cui, X., Hong, J., Gao, M., 2012. Environmental impact assessment of three coal-based electricity
   generation scenarios in China. Energy 45 (1), 952–959. https://doi.org/10.1016/j.energy.2012.06.063.
- Danziger, J., Dodge, L.E., Hu, H., Mukamal, K.J., 2022. Susceptibility to Environmental Heavy Metal Toxicity among Americans with Kidney Disease. Kidney360 3 (7), 1191–1196. https://doi.org/10.34067/KID.0006782021.
- Demirel, O.F., Zaim, S., Caliskan, A., Ozuyar, P., 2012. Forecasting natural gas consumption in Istanbul using neural networks and multivariate time series methods. Turkish Journal of Electrical Engineering and Computer Sciences.
- Dinkele, R., Gessner, S., McKerry, A., Leonard, B., Leukes, J., Seldon, R., Warner, D.F., Wood, R.,
  2022. Aerosolization of MycobacteriumTBby Tidal Breathing. Am. J. Respir. Crit. Care Med. 206
  (2), 206–216. https://doi.org/10.1164/rccm.202110-2378OC.
- Ecoinvent Centre, 2016. Swiss Centre for life cycle inventories, ecoinvent v3.3 reports, 2016.
   https://ecoinvent.org/the-ecoinvent-database.
- 525 Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., 526 Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, 527 J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, 528 M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, 529 S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, 530 Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, 531 P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N., 532 Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pan, 533 N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., 534 Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, 535 A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., 536 Tubiello, F., van der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, 537 A., Wright, R., Yuan, W., Yue, C., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 538 2022, Earth Syst. Sci. Data, 14, 4811-4900, https://doi.org/10.5194/essd-14-4811-2022, 2022.
- 539

- GB 13223-2011. Emission standard of air pollutants for thermal power plants. Ministry of ecology and environment of the People's Republic of China (in Chinese).
- Ghosh, S.R., 2013. Global Coal Beneficiation Scenario and Economics of Using Washed Coal. Indian
   Central Mine Planning & Design Institute Limited.
- Hong, J., Yu, Z., Fu, X., Hong, J., 2019. Life cycle environmental and economic assessment of coal
  seam gas-based electricity generation. Int J Life Cycle Assess. 24, 1828-1839.
- Hong, J., Zhang, F., Xu, C., Xu, X., Li, X., 2015. Evaluation of life cycle inventory at macro level:
  a case study of mechanical coke production in China. Int J Life Cycle Assess 20 (6), 751–764.
  https://doi.org/10.1007/s11367-015-0876-9.
- Hu, Y., Cheng, H., Tao, S., 2018. The growing importance of waste-to-energy (WTE) incineration in China's anthropogenic mercury emissions: Emission inventories and reduction strategies. Renewable and Sustainable Energy Reviews 97, 119–137. https://doi.org/10.1016/j.rser.2018.08.026.
- Huang, X., Wu, L., Ye, Y., 2019. A review on dimensionality reduction techniques. International Journal of Pattern Recognition and Artificial Intelligence 33, 10 (2019), 1950017.
- Huang, Y., Ren, Y., Yang, H., Ding, Y., Liu, Y., Yang, Y., Mao, A., Yang, T., Wang, Y., Xiao, F., He,
  Q., Zhang, Y., 2022. Using a machine learning-based risk prediction model to analyze the coronary
  artery calcification score and predict coronary heart disease and risk assessment. Comput. Biol. Med.
  151, 106297. https://doi.org/10.1016/j.compbiomed.2022.106297.
- Huijbregts, M., Hauschild, M., Jolliet, O., Margni, M., McKone, T., Rosenbaum, R.K., van de Meent,
  D., 2010. USEtoxTM User Manual Version 1.01. Substance database of the UNEP/SETAC model
  for the comparative assessment of chemicals released to air, water and soil and their toxic effects on
  the human population and ecosystems, USEtoxTM. http://www.usetox.org
- IEA (International Energy Agency), 2022. Global Energy Review: CO2 Emissions in 2021. https://www.iea.org/reports/global-energy-review-co2-emissions-in-2021-2.
- IHME, 2020. Global Burden of Disease Study 2019 Results. Institute for Health Metrics and
   Evaluation (IHME) the united states. Available from https://vizhub.healthdata.org/gbd-results/.
- ISO 14040, 2006. Environmental Management-Life Cycle Assessment-Principles and Framework,
   International Standard.
- ISO 14044, 2006. Environmental Management-Life Cycle Assessment-Requirements and Guidelines, International Standard.
- Jia, Y., Zhang, T., Zhai, Y., Bai, Y., Ren, K., Shen, X., Cheng, Z., Zhou, X., Hong, J.,
  2022. Exploring the potential health and ecological damage of lead–zinc production activities
  in China: A life cycle assessment perspective. Journal of Cleaner Production 381, 135218.
  https://doi.org/10.1016/j.jclepro.2022.135218.
- Krittanawong, C., Virk, H.U.H., Bangalore, S., Wang, Z., Johnson, K.W., Pinotti, R., Zhang, H.,
  Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J.L., Tang, W.H.W., 2020. Machine
  learning prediction in cardiovascular diseases: a meta-analysis. Scientific reports 10 (1), 16057.
  https://doi.org/10.1038/s41598-020-72685-1.
- Landrigan, P.J., Fuller, R., Hu, H., Caravanos, J., Cropper, M.L., Hanrahan, D., Sandilya, K.,
  Chiles, T.C., Kumar, P., Suk, W.A., 2018. Pollution and Global Health An Agenda for Prevention.
  Environmental health perspectives 126 (8), 84501. https://doi.org/10.1289/EHP3141.
- Li, D., Chang, J., Hong, J. 2024. Toward a comprehensive life-cycle carcinogenic impact assessment: A statistical regression approach based on cancer burden. Science of The Total Environment, 921, 170851.
- Li, D., Qin, J., Hong, J., 2024. Toward a comprehensive life cycle aquatic ecotoxicity assessment via
   machine learning: Application to coal power generation in China. Journal of Cleaner Production, 445, 141373.
- Liu, Z., Guan, D., Wei, W., Davis, S.J., Ciais, P., Bai, J., Peng, S., Zhang, Q., Hubacek, K.,
  Marland, G., Andres, R.J., Crawford-Brown, D., Lin, J., Zhao, H., Hong, C., Boden, T.A., Feng,

K., Peters, G.P., Xi, F., Liu, J., Li, Y., Zhao, Y., Zeng, N., He, K., 2015. Reduced carbon emission estimates from fossil fuel combustion and cement production in China. Nature 524 (7565), 335–338. https://doi.org/10.1038/nature14677.

Ma, X., Zhang, T., Ji, C., Zhai, Y., Shen, X., Hong, J., 2021. Threats to human health and ecosystem:
Looking for air-pollution related damage since 1990. Renewable and Sustainable Energy Reviews 145, 111146. https://doi.org/10.1016/j.rser.2021.111146.

Mohammad, K.N., Chan, R.W.Y., Lai, C.K.C., Chan, P.K.S., Leung, C.S.Y., Chen, V.X.Y., Wang,
Y., Wei, Y., 2022. Independent effect of weather, air pollutants, and seasonal influenza on risk
ofTBhospitalization: An analysis of 22-year hospital admission data. The Science of the total
environment 837, 155711. https://doi.org/10.1016/j.scitotenv.2022.155711.

N., Abbastabar, H., Abd-Allah, F., 2018. Global, regional, and national comparative risk assessment
of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195
countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study
2017. Lancet 392 (10159), 1923–1994.

NASA, 2022. World of Change: Global Temperatures. https://earthobservatory.nasa.gov/world ofchange/global-temperatures.

Nguyen, L.H., Holmes, S., 2019. Ten quick tips for effective dimensionality reduction. PLoS computational biology 15, 6 (2019), e1006907. https://doi.org/10.1371/journal.pcbi.1006907

Patton, A., 2013. Copula methods for forecasting multivariate time series. Handbook of Economic
 Forecasting.

Polyzotis, N., Roy, R., Whang.S.E., Zinkevich, M., 2017. Data Management Challenges in Production
Machine Learning. In Proceedings of the 2017 ACM International Conference on Management of
Data (Chicago, Illinois, USA) (SIGMOD '17). Association for Computing Machinery, New York,
NY, USA, 1723–1726. https://doi.org/10.1145/3035918.3054782

Public Health Science Data Center, 2023. Chinese center for disease control and prevention.
 https://www.phsciencedata.cn/Share/loginShare.jsp?id=1d0a4e38-90f8-11eb-a123-7cd30ac209ea#

Ren, K., Zhang, T., Bai, Y., Zhai, Y., Jia, Y., Zhou, X., Cheng, Z., Hong, J., 2022. Environmental and economical assessment of high-value utilization routes for coke oven gas in China. Journal of Cleaner Production 353, 131668. https://doi.org/10.1016/j.jclepro.2022.131668.

Sala, S., Wolf, M.-A., Pant, R., 2012. Characterisation factors of the ILCD Recommended Life Cycle
 Impact Assessment methods. Database and supporting information.

Shi W., Lin C., Chen W., Hong J., Chang J., Dong Y., Zhang Y., 2017. Environmental effect of current desulfurization technology on fly dust emission in China, Renewable & Sustainable Energy Reviews 72, 1-9.

Stanaway, J.D., Afshin, A., Gakidou, E., Lim, S.S., Abate, D., Abate, K.H., Abbafati, C., Abbasi,

Suárez, I., Fünger, S.M., Kröger, S., Rademacher, J., Fätkenheuer, G., Rybniker, J., 2019. The
Diagnosis and Treatment of Tuberculosis. Deutsches Arzteblatt International 116 (43), 729–735.
https://doi.org/10.3238/arztebl.2019.0729.

<sup>636</sup> UNEPA (U.S. Environmental Protection Agency), 2002. Global mercury assessment.

<sup>637</sup>United Nations, 2015. A/RES/70/1 - Transforming our world: the 2030 Agenda for Sustainable
 <sup>639</sup>Development. https://sdgs.un.org/2030agenda.

Us about the Impact of Chemical Mixtures on Human Health? Environmental health perspectives124 (1), A6-9. https://doi.org/10.1289/ehp.1510569.

USEPA (U.S. Environmental Protection Agency), 1977. Mercury study report to congress.
 https://www.fda.gov/OHRMS/DOCKETS/ac/02/briefing/3872\_Studies%2015.pdf.

644<br/>645<br/>646WHO (World health organization), 2022.GlobalTBreports 2010-2021.646https://www.who.int/teams/global-tuberculosis-programme/tb-reports.

WMO (2011) Scientific assessment of ozone depletion: 2010, Global Ozone Research and Monitoring
 Project-report no.52. World Meteorological Organization, Geneva.

Xu, C., Hong, J., Ren, Y., Wang, Q., Yuan, X., 2015. Approaches for controlling air pollutants and their environmental impacts generated from coal-based electricity generation in China. Environmental science and pollution research international 22 (16), 12384–12395. https://doi.org/10.1007/s11356015-4539-y.

- Yang, D., Gao, S., Hong, J., Ye, L., Ma, X., Qi, C., Li, X., 2019. Life cycle assessment of rare earths
  recovery from waste fluorescent powders A case study in China. Waste management (New York,
  N.Y.) 99, 60–70. https://doi.org/10.1016/j.wasman.2019.08.038.
- Yu, G., Zou, J., Hu, X., Aviles-Rivero, A. I., Qin, J., Wang, S., 2024. Revitalizing Multivariate
  Time Series Forecasting: Learnable Decomposition with Inter-Series Dependencies and Intra-Series
  Variations Modeling. In Forty-first International Conference on Machine Learning
- Yu,T., Simoff, S., Jan, T., 2010. VQSVM: A case study for incorporating prior domain knowledge into inductive machine learning. Neurocomputing 73, 13 (2010), 2614–2623. https://doi.org/10.1016/j.neucom.2010.05.007 Pattern Recognition in Bioinformatics Advances in Neural Control.
- Yun X., Meng W., Xu H., Zhang W., Yu X., Shen H., Chen Y., Shen G., Ma J., Li B., Cheng H., Hu J.,
  Tao S., 2021. Coal is dirty, but where it is burned especially matters. Environ. Sci. Technol. 55, 11,
  7316–7326. https://pubs.acs.org/doi/full/10.1021/acs.est.1c01148.
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J. 2020. A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. Journal of Applied Science and Technology Trends 1, 2 (2020), 56–70.
- Zhang, Y., and Yan, J., 2023. Crossformer: Transformer utilizing cross-dimension dependency for
   multivariate time series forecasting. In The eleventh international conference on learning representa tions.
- Zhang, Y., Sun, M., Hong, J., Han, X., He, J., Shi, W., Li, X., 2016. Environmental footprint of aluminum production in China. Journal of Cleaner Production 133, 1242–1251. https://doi.org/10.1016/j.jclepro.2016.04.137.
- Zhao, J., Hong, J., Song, Z., Wang, Q., Zhao, X., Ma, C., 2015. Cost combined life cycle assessment of lignite-based electricity generation. Fuel 159, 666–674. https://doi.org/10.1016/j.fuel.2015.07.026.
- Zhao, Q., Guo, Y., Ye, T., Gasparrini, A., Tong, S., Overcenco, A., Urban, A., Schneider, A., Entezari, 681 A., Vicedo-Cabrera, A.M., Zanobetti, A., Analitis, A., Zeka, A., Tobias, A., Nunes, B., Alahmad, 682 B., Armstrong, B., Forsberg, B., Pan, S.-C., Íñiguez, C., Ameling, C., La Cruz Valencia, C. de, Aström, 683 C., Houthuijs, D., van Dung, D., Royé, D., Indermitte, E., Lavigne, E., Mayvaneh, F., Acquaotta, F., 684 de'Donato, F., Di Ruscio, F., Sera, F., Carrasco-Escobar, G., Kan, H., Orru, H., Kim, H., Holobaca, 685 I.-H., Kyselý, J., Madureira, J., Schwartz, J., Jaakkola, J.J.K., Katsouyanni, K., Hurtado Diaz, M., 686 Ragettli, M.S., Hashizume, M., Pascal, M., Sousa Zanotti Stagliorio Coélho, M. de, Valdés Ortega, 687 N., Ryti, N., Scovronick, N., Michelozzi, P., Matus Correa, P., Goodman, P., Nascimento Saldiva, P.H., 688 Abrutzky, R., Osorio, S., Rao, S., Fratianni, S., Dang, T.N., Colistro, V., Huber, V., Lee, W., Seposo, 689 X., Honda, Y., Guo, Y.L., Bell, M.L., Li, S., 2021. Global, regional, and national burden of mortality 690 associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. The Lancet. Planetary health 5 (7), e415-e425. https://doi.org/10.1016/S2542-5196(21)00081-4. 691
- 692
- 693 694
- 695
- 696
- 697
- 698
- 699 700
- 700

## A APPENDIX

702

703 704 705

Table 3: LCHHA	formula ex	planation
----------------	------------	-----------

706	Parameter	Meaning	
707	$XF_i$	Exposure factor of substance i	
708	$\mathrm{EF}_i$	Effect factor of substance i	
709	NC	Human health damage (DLAY) caused by carcinogenicity	
710	С	Human health damage (DLAY) caused by non-carcinogenicity	
711	PF	Human health damage (DLAY) caused by particulate formation	
712	OF	Human health damage (DLAY) caused by ozone formation	
713	OD	Human health damage (DLAY) caused by ozone depletion	
714	GW	Human health damage (DLAY) caused by global warming	
715	BAF	Bioconcentration factor	
716	PROD	Intake per unit time	
710	MASS	Total mass of pollutants in the exposure area	
717	INH	Average respiratory rate per person	
/18	Vair	Total air volume in the region	
719	LT	Average life expectancy of the population	
720	BW	Average body weight of the population	
721	$ED_{50}$	Median effective dose: the dose that elicits the desired effect in 50% of subjects	
722	$\mathrm{EF}_{PF,i}$	Health impact per kilogram of inhaled substance i	
723	$\mathrm{EF}_{PM2.5}$	Health impact per kilogram of inhaled PM2.5	
724	$\mathrm{DF}_{PF}$	Human health damage caused by per unit intake of PM2.5	
725	$\mathrm{EF}_{OF,i}$	Health impact per kilogram of inhaled substance i	
726	$\mathrm{EF}_{NOx}$	Health impact per kilogram of inhaled NOx	
727	$\mathrm{DF}_{NOx}$	Human health damage caused by per unit intake of NOx	
728	UVB	Increase in radiation in the area	
720	$EF_{OD}$	Additional disease incidence rate due to increased UV radiation in the area	
720	$DF_{OD}$	Human health damage caused by disease incidence due to UV radiation	
730	$\mathrm{GWP}_{GW,i}$	Global warming potential of greenhouse gas i as reported by the latest IPCC	
/31	$\mathrm{GWP}_{CO_2}$	Global warming potential of CO2 as reported by the latest IPCC	
732	AGWP	Absolute global warming potential in $yr \cdot W \cdot m^{-2} \cdot kg^{-1}$	
733	TF	Temperature factor in $C \cdot m^2 \cdot W^{-1}$	
734	RR	Relative health risks from rising temperatures	
735	$\mathrm{DF}_{GW,i}$	Damage factor for health damage caused by emissions of greenhouse gas i,	
736		calculated through AGWP, TF, RR, and DALY $yr^{-1} \circ C^{-1}$	

737 738

739

#### A.1 CLIMATE CHANGE IN CHINA

740 Carbon dioxide, the main driver of climate change, has Carbon dioxide, the main driver of climate 741 change, has attracted global attention. In China, carbon dioxide emitted from the whole country 742 (Friedlingstein et al., 2022) and its life cycle emissions from the coal power industry show a significant 743 uptrend during the last seven decades. The life cycle carbon emissions of the coal power industry 744 currently account for half of the national carbon emission in China. Meanwhile, the annual average 745 temperature in China shows a sharp rise (Fig.5a), with a warming rate of 0.25 °C/10 years, which is higher than the global average level for the same period  $(0.15^{\circ}C \text{ to } 0.20^{\circ}C/10 \text{ years; NASA})$ 746 2022). Spatiotemporal average humidity and rainfall show an uptrend with increasing average 747 temperature and a downtrend with increasing sunshine duration, whereas annual average wind speed 748 with temperature below -2.5 °C presents a downtrend (Fig.5b). For spatiotemporal wind speed 749 beyond -2.5 °C, no significant changes are observed with increasing annual average temperature. The 750 average rainfall distribution in China shows a decreasing tendency from the southeast coast to the 751 northwest inland (Fig.5c). A similar tendency can be observed in the average humidity distribution, 752 as the humidity distribution is closely related to the precipitation distribution. By contrast, the annual 753 average sunshine duration shows an increasing tendency from the southeast coast to the northwest 754 inland. The average temperature distribution is high in the south and low in the north, whereas a 755 contrasting phenomenon is observed in the average wind speed distribution.



Figure 5: Spatiotemporal meteorological element changes: a) annual average variation since 1951; b)
 relationship among meteorological elements since 1951 (color represents different region); c) spatial variation in 2019 (white color means no data).







Figure 7: Secular variation of China's coal supply, coal production, and coal consumption in coal



Figure 10: Each pollutant proportion in KFDA component



#### Table 4: Top 5 Key Contributors for 5 Different DR Methods

978					
979	PCA	LLE	KPCA	KFDA	ISOMAP
980	Feature	Feature	Feature	Feature	Feature
981	NMVOC	Benzo(g,h,i)perylene	Anthracene	Benzo(ghi)perylene	Chromium
982	Anthracene	Cadmium	Benzo(k)fluoranthene	Benzo(b)fluoranthene	Pyrene
983	Benzo(k)fluoranthene	Carbon dioxide	Acenaphthylene	Benzo(a)anthracene	Particulates
984	Acenaphthylene	Arsenic	Acenaphthene	Benzo(a)pyrene	Zinc
985	Acenaphthene	Methane	Benzo(a)pyrene	Acenaphthene	Carbon monoxide

#### Table 5: Parameter values

Parameter	Value	
seq_len	96	
pred_len	1	
d_model	256	
n_layers	3	
learning_rate	1.00E-04	
kernel_size	25	
features	MS	
target	OT	
freq	m	
label_len	48	
enc_in	11	
dec_in	11	
c_out	11	
pe_type	no	
dropout	0	
revin	TRUE	
num_workers	0	
itr	1	
train_epochs	100	
batch_size	32	
patience	6	
Îradj	constant	
use_amp	TRUE	
use_gpu	TRUE	
gpu	0	