IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 18, 2020, accepted May 28, 2020, date of publication June 8, 2020, date of current version June 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000767

Causal Identification Based on Compressive Sensing of Air Pollutants Using Urban Big Data

MINGWEI LI, JINPENG LI, SHUANGNING WAN, HAO CHEN, AND CHAO LIU
School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China

Corresponding author: Mingwei Li (neuqlmw@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 61702088.

ABSTRACT This study addresses the causal identification of air pollutants from surrounding cities affecting Beijing's air quality. A novel compressive sensing causality analysis (CS-Causality) method, which combines Granger causality analysis (GCA) and maximum correntropy criterion (MCC), is presented for efficient identification of the air pollutant causality between Beijing and surrounding cities. Firstly, taking the spatiotemporal correlation into consideration, the original data is mapped into low-dimensional space. Valid information is then obtained based on compressive sensing (CS), which can greatly reduce the dimensions of the data, thus decreasing the amount of data analysis required. Secondly, to analyze the causal relations, GCA, represented by the prediction from one time series to another, is extended to rule out ''Non-Granger'' causes of air pollutants in Beijing originating from its surrounding cities. Thirdly, the greatest impact on Beijing's air quality is confirmed based on MCC. Finally, the accuracy of these results is verified using the transfer entropy.

INDEX TERMS Granger causality analysis, maximum correntropy criterion, data compression, air pollutant.

I. INTRODUCTION

In recent years, air quality has attracted widespread concern due to its rapid deterioration. Air pollution not only directly affects human health, and even seriously threatens human life. Moreover, air pollution is estimated to cause 3.7 million deaths per year and has contributed increasingly to the global burden of disease. This serious phenomenon has caused widely public concern. In particular, the air quality in Beijing and its surrounding cities has received widespread attention from relevant departments and research institutions. According to authoritative reports, the content of PM2.5, PM10, SO2, NO2, CO in the air is an important indicator of the severity of air pollution.

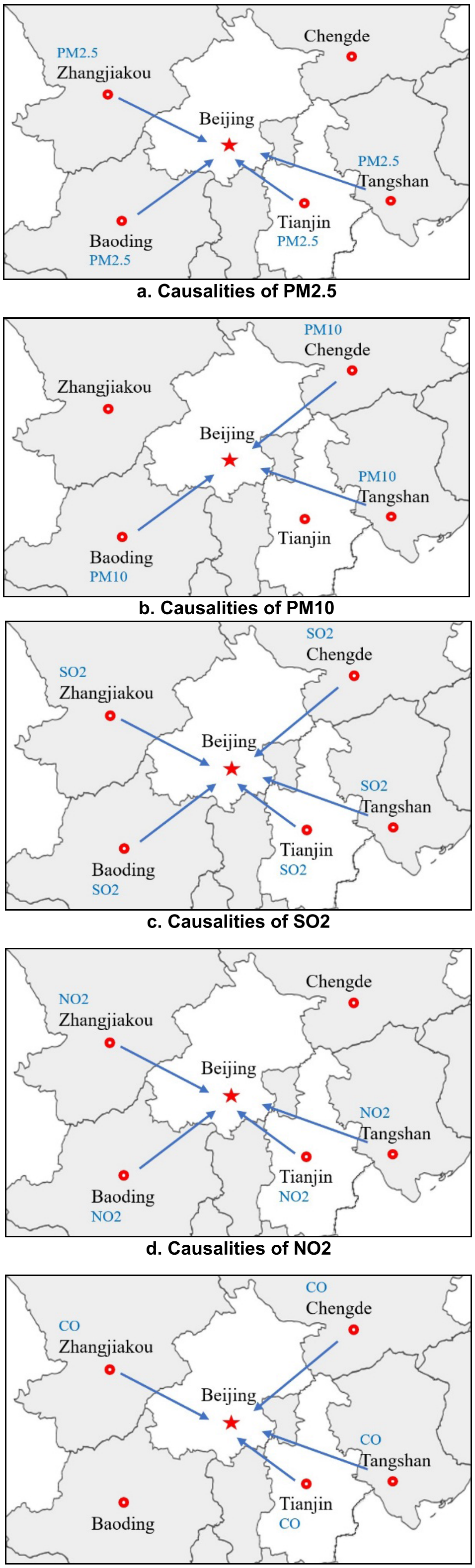
Because of the harm caused by air pollutants to the human body, control of air pollution is critical, especially in terms of reducing the content of pollutants in the air. Now, the primary task is to seek the source of pollutants, to cut off the diffusion of pollutants, so as to fundamentally solve the air pollution problem. Taking the relationship between air quality in Beijing and its surrounding cities as an example, it can be seen from FIGURE 1 that there is a certain correlation between them. However, the huge pollutant data contains a lot of false data and duplicate data, which makes the causality analysis result inaccurate. Therefore the severity of air pollution has not been significantly alleviated.

A. RELATED WORKS

To address air pollution problems, data-driven analysis and causality modeling analysis have become popular tools to predict and analyze the relationship between meteorological data.

1) Data-driven air quality analysis. Several researchers [1]–[6] proposed approaches to forecast and estimate air quality by analyzing and processing the correlations and patterns found in heterogeneous big data. Shang *et al.* [7] estimated gas consumption and pollutant emissions based on GPS data. Zheng *et al.* [8] designed a linear regression-based temporal predictor to model the local air quality factors considering current meteorological data. Simona *et al.* [9] proposed a mobile air pollution monitoring framework coupled together with a data-driven modelling approach for predicting the air quality inside urban areas, at human breathing level. A static monitoring protocol was proposed for comparing the performance of two different mobile sensing units. The proposed experimentation protocols showed not only a significantly higher

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Xiao.



a. Causalities of PM2.5

b. Causalities of PM10

c. Causalities of SO2

d. Causalities of NO2

e. Causalities of CO

FIGURE 1. An illustration of the identification of causalities.

impact of NO2 concentrations, but also poor noise levels registered. The data-driven investigation revealed that data generated by the mobile sensing units when used outdoors can be accurately used to predict future NO2 levels in urban areas.

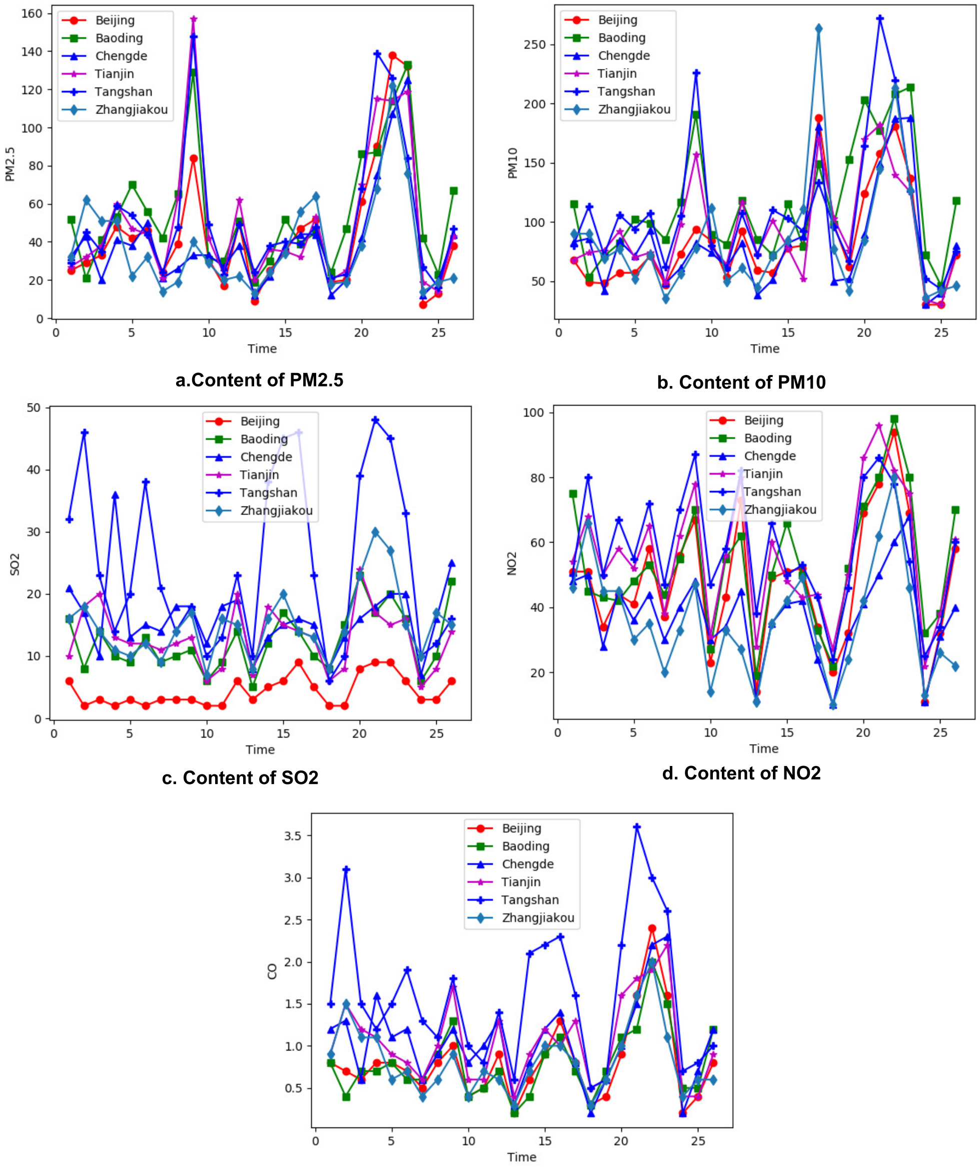
2) Causality modelling for time series. Identifying the source of pollutants has become a critical problem. Recently, due to the availability of air quality big data collected by wireless sensors deployed in different regions, it has become possible to analyze the causalities of pollutants among surrounding cities. To address the problem, Zhu *et al.* [10] proposed a pattern-aided graphical causality analysis approach, based on pattern mining and Bayesian learning, to identify the spatiotemporal causal pathways for air pollutants. Ebert-Uphoff and Deng [11] investigated the causal discovery problem, which set up spatiotemporal problems using a time series and handled temporal and spatial boundaries. Yang *et al.* [12] approximated the Granger causality of the multivariate time series in a unified framework. To predict haze pollution, Peng *et al.* [13] proposed a PS-FCM (Primary Sub-Fuzzy Cognitive Maps) model to reveal causality in the formation of haze. The causality, based on time series data of haze pollution with PS-FCM, were explored and discovered through considering the formation of haze as an evolving process with time. Thus, a multi-dimensional time series data mining method based on the PS-FCM was developed to investigate the formation of haze.

B. MOTIVATION AND MAJOR CONTRIBUTIONS

Although these methods succeeded in improving the performance of data analysis, identifying the source of pollutants from air quality big data is a serious challenge for the following reasons:

1) Large amounts of air quality data can make causality analysis very difficult. Tens of thousands of sensors deployed in different areas sample large amounts of data frequently on a daily basis. Discovering causality in such a large amount of air quality data is very difficult.
2) A large amount of noisy data can lead to inaccuracy in the causality analysis. There are many repetitive or false data points in the original air quality data. These data points are not only useless, but also directly affect the accuracy of the causality analysis.

With such a large amount of data (there is repeatability between the data), an efficient data compression strategy is essential to analyze causality efficiently. Data compression algorithms can be divided into two types: lossless and lossy compression. Lossless compression guarantees the integrity of reconstruction with a cost of relatively poor compression ratios. In lossy compression, compressive sensing (CS) [14], [15] is best known for its performance advantages. The most prominent feature of CS is that redundant temporal-spatial domain information can be reduced for each sensor independently. The method with superior characteristics is very suitable for data compression processing before causal analysis of big data.



a.Content of PM2.5

b. Content of PM10

c. Content of SO2

d. Content of NO2

e. Content of CO

FIGURE 2. Air pollutants in Beijing and its surrounding cities.

Inspired by the robustness and efficiency of the recently proposed pg-Causality, we bring further contributions to the analysis of pollutant causality between Beijing and its surrounding cities based on Granger causality analysis (GCA) and compressive sensing (CS). The spatiotemporal causalities should be reflected by applying the following two aspects:

1) The spatial correlation, which denotes the propagation and diffusion of multiple pollutants in the space.
2) The temporal correlation, which indicates causality is changed at different time lags.

Based on these aspects, we carried out the following innovative research. Firstly, the original air quality data from the different cities (the sparse time series) were respectively mapped into low-dimensional space to obtain the respective low-dimensional vectors. The compression process effectively removed repetitive or false data and reduced computational complexity. Secondly, the causalities between Beijing and its surrounding cities were determined using the Granger causality index. Thirdly, using the maximum correntropy criterion (MCC), the surrounding cities which had the greatest impact on Beijing's air quality were determined.

II. CAUSALITIES OF POLLUTANTS

A. ORIGINAL DATA TEST

The spatiotemporal air quality data in Beijing and its surrounding cities were selected from November 1st to November 26th in 2019. The synchronous correlations in PM2.5, PM10, SO2, CO, and NO2 are shown in FIGURE 2. The horizontal axes show the time range (November 1, 2019 to

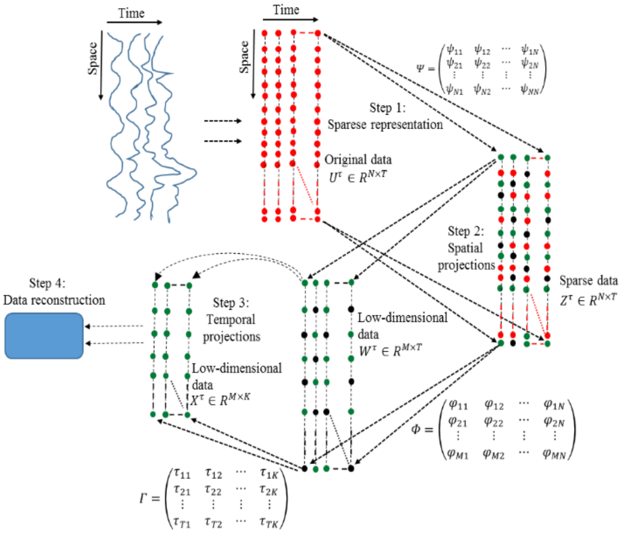

Time
Space
Time
Space
$\Psi = \begin{pmatrix} \psi_{11} & \psi_{12} & \cdots & \psi_{1N} \\ \psi_{21} & \psi_{22} & \cdots & \psi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{N1} & \psi_{N2} & \cdots & \psi_{NN} \end{pmatrix}$
Step 1:
Sparese representation
Original data
$U^\tau \in R^{N\times T}$
Step 2:
Spatial
projections
Sparse data
$Z^\tau \in R^{N\times T}$
Step 4:
Data reconstruction
Step 3:
Temporal
projections
Low-dimensional
data
$X^\tau \in R^{M\times K}$
Low-dimensional
data
$W^\tau \in R^{M\times T}$
$\Phi = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{M1} & \varphi_{M2} & \cdots & \varphi_{MN} \end{pmatrix}$
$\Gamma = \begin{pmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1K} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ \tau_{T1} & \tau_{T2} & \cdots & \tau_{TK} \end{pmatrix}$

FIGURE 3. Overview of CS.

November 26, 2019), and the vertical axes show the pollutant content(μg/m3).

As can be seen in Fig.2, it is difficult to predict which urban air pollutants have a significant impact on Beijing's air quality based on the original data. Therefore, we will utilize Granger causality with compressive sensing to further explore the impact of pollutants in other cities on Beijing's air quality in the next section.

B. DATA PROCESSING

It is well known that the original collected data will be sparse and repetitive [16], [17]. In order to improve the effectiveness of air quality data, special processing is carried out on the original data to remove invalid data, so as to provide effective information for the data collection and analysis.

Let U $= \{U_1, U_2, U_3, U_4, U_5\}$ be the set of air qualities from surrounding cities, in which U_1, U_2, U_3, U_4, U_5 denote pollutants in Zhangjiakou, Chengde, Tianjin, Baoding, and Tangshan, respectively. The set V represents pollutants in Beijing. Compressive sensing is applied to reduce the amount of required data analysis and processing. The processing consists of four steps, as illustrated in FIGURE 3. Firstly, sparse representation of original data. Generally, the original data are not sparse, so needed to transform into sparse representations on the sparse basis, which can be formulated as (1)-(2).

$$U^\tau = \begin{bmatrix} u_1^\tau & u_2^\tau & u_3^\tau & \cdots & u_N^\tau \end{bmatrix}^T = \Psi Z^\tau \tag{1}$$

$$V = \begin{bmatrix} v_1 & v_2 & v_3 & \cdots & v_N \end{bmatrix}^T = \Psi\Theta \tag{2}$$

where $\Psi \in R^{N\times N}$ is a sparse basis, $Z^\tau \in R^{N\times T}$, $\Theta \in R^{N\times T}$ are sparse coefficients, U^τ, V are time series of original data, where N is the size of the time series. Secondly, spatial projections. In this matrix operation, the original data are projected to an M-dimensional space for eliminating spatial correlations through M times cycle, which can be expressed in (3)-(4). This step achieves data compression in the spatial domain.

$$W^\tau = \Phi U^\tau = \begin{bmatrix} w_1^\tau & w_2^\tau & w_3^\tau & \cdots & w_M^\tau \end{bmatrix}^T \tag{3}$$

$$Q = \Phi V = \begin{bmatrix} q_1 & q_2 & q_3 & \cdots & q_M \end{bmatrix}^T \tag{4}$$

$W^\tau \in R^{M\times T}$, $Q \in R^{M\times T}$ are low-dimensional data in the spatial domain, $\Phi \in R^{M\times N}(M \ll N)$ is the sensing matrix that satisfies the restricted isometry property(RIP). Thirdly, temporal projections. For the sparsity in the time-domain, data compression can further reduce the amount of data for causal analysis. The process can be described in (5)-(6)

$$X^\tau = W^\tau\Gamma = \begin{bmatrix} x_1^\tau & x_2^\tau & x_3^\tau & \cdots & x_M^\tau \end{bmatrix}^T \tag{5}$$

$$Y = Q\Gamma = \begin{bmatrix} y_1^\tau & y_2^\tau & y_3^\tau & \cdots & y_M^\tau \end{bmatrix}^T \tag{6}$$

where $X^\tau \in R^{M\times K}$, $Y \in R^{M\times K}(K < T)$ are low-dimensional data in the spatiotemporal domain, $\Gamma \in R^{T\times K}$ is temporal compressive basis. Finally, data reconstruction. After causal analysis, the original data are recovered from the sparse coefficients using the l_1-norm minimization [17].

Without loss of generality, in the subsequent compression process, the elements of sensing matrix Φ are constructed as Gaussian distribution, sparse basis Ψ is calculated by applying the inverse Discrete Cosine Transform upon the columns of the identity matrix, and elements of temporal compressive basis Γ are designed to be 1.

C. GRANGER CAUSALITY ANALYSIS

To detect and measure which cities have an impact on Beijing's air pollution, causalities between Beijing and its surrounding cities have been discussed. The popular, recent causality models include graphical causality [18], [19], unit-level causality [20], and predictive causality [21], [22]. Granger [23] in predictive causality is applied to test the impact of the surrounding cities on Beijing's air pollution.

$$x_{d,t}^\tau = \sum_{i=1}^{p_1} \alpha_{d,i}^1 x_{t-i}^\tau + \sum_{j=1}^{p_2} \beta_{d,j}^1 y_{d,t-j} + e_{d,1}^\tau(t) \tag{7}$$

$$y_{d,t} = \sum_{i=1}^{p_3} \alpha_{d,i}^2 x_{t-i}^\tau + \sum_{j=1}^{p_4} \beta_{d,j}^2 y_{d,t-j} + e_{d,2}^\tau(t) \tag{8}$$

Here $e_{d,1}^\tau(t)$, $e_{d,2}^\tau(t)$ are the errors at t, $p_i(i = 1, 2, 3, 4)$ is the number of timestamps, and $\alpha_{d,i}$, $\beta_{d,i}$, $(d = 1, 2, \cdots, M)$ are the correspondent weights for the time series X^τ, Y. When the models represented by equations (7) and (8) satisfy the null hypothesis, they degrade to autoregressive (AR) models, which indicates that there is no causality between X^τ and Y. Thus, the AR models for the time series in Granger causality can be expressed as follows:

$$x_{d,t}^\tau = \sum_{i=1}^{p_5} \alpha_{d,i}^3 x_{d,t-i}^\tau + e_{d,3}^\tau(t), \qquad d = 1, 2, \cdots, M \tag{9}$$

$$y_{d,t} = \sum_{i=1}^{p_6} \beta_{d,i}^3 y_{d,t-i} + e_{d,4}(t), \qquad d = 1, 2, \cdots, M \tag{10}$$

where $e^{\tau}_{d,3}(t), e_{d,4}(t)$ are the errors at t; p_5, p_6 are the number of timestamps; and $\alpha^{3}_{d,i}, \beta^{3}_{d,i}$ are the correspondent weights for the time series X^{τ}, Y. In the case of (9) and (10), the time series were caused by their own histories, not by others. In [24], the independents of X^{τ}, Y are defined as follows: $\mathrm{Var}(e^{\tau}_{1}(t)) = \mathrm{Var}(e^{\tau}_{3}(t))$ and $\mathrm{Var}(e^{\tau}_{2}(t)) = \mathrm{Var}(e_4(t))$, where $\mathrm{Var}(e^{\tau}_{i}(t))$, $(i = 1, 2, 3, 4)$ denotes the variance of the error $e^{\tau}_{i}(t)$, $Var(e^{\tau}_{i}(t)) = \sqrt{\sum_{d=1}^{M} Var^2(e^{\tau}_{d,i}(t))}$, $(i = 1, 2, 3, 4)$. Otherwise, we need to distinguish causality between X^{τ} and Y. The Granger causality index from X^{τ} to Y can be defined as follows:

$$F_{X^{\tau}\to Y} = \log \frac{\sum_{t=1}^{T} \mathrm{Var}(e_4(t))}{\sum_{t=1}^{T} \mathrm{Var}(e^{\tau}_{2}(t))} \tag{11}$$

If $\sum_{t=1}^{T} \mathrm{Var}(e_4(t)) > \sum_{t=1}^{T} \mathrm{Var}(e^{\tau}_{2}(t))$, then X^{τ} is the cause of Y. Thus, we can obtain that $F_{X^{\tau}\to Y} > 0$. In particular, if $F_{X^{\tau}\to Y} > 0$, we can conclude that X^{τ} is the cause of Y. Therefore, the sufficient and necessary condition of $F_{X^{\tau}\to Y} > 0$ is that X^{τ} is the cause of Y. Certainly, if $F_{X^{\tau}\to Y} = 0$, there is no causality from X^{τ} to Y. Similarly, the definition of the Granger causality index from Y to X^{τ} is given by:

$$F_{Y\to X^{\tau}} = \log \frac{\sum_{t=1}^{T} \mathrm{Var}(e^{\tau}_{3}(t))}{\sum_{t=1}^{T} \mathrm{Var}(e^{\tau}_{1}(t))} \tag{12}$$

When $F_{X^{\tau}\to Y} > F_{Y\to X^{\tau}}$, the effect of X^{τ} on Y is significantly higher than the effect of Y on X^{τ}. For the theoretical analysis, the original data set of $U_{\tau}(\tau = 1, 2, 3, 4, 5)$ and V were first sampled periodically. The invalid data were then effectively eliminated based on CS and the low-dimensional time series $X^{\tau}(\tau = 1, 2, 3, 4, 5)$ and Y could then be obtained. $F_{X^{\tau}\to Y}$ and $F_{Y\to X^{\tau}}(\tau = 1, 2, 3, 4, 5)$ were calculated and respectively compared using (11)-(12). Thus, it could be determined which X^{τ}(pollutant index of the surrounding cities) was the cause of Y(pollutant index of Beijing).

D. CAUSALITY OBSERVATION

To detect and measure which surrounding cities could be the cause of air pollution in Beijing, a number of assumptions had to be made about the external environment.

1) Uncontrollable factors, such as wind speed, wind direction, temperature, humidity, pressure, and other natural conditions, were not considered. Uncontrollable factors vary from place to place. If they were all taken into account, it would inevitably increase the difficulty of the causality analysis and affect the accuracy of analysis results.

2) The compression error of original data was allowed. The mapping of the original data from high-dimensional space to low-dimensional space is a lossy process, as is the reconstruction of the data from low-dimensional space to high-dimensional space. Typically, the error ratio of the compression process is within 5%.

We utilized the compressed data associated with the five major pollutants to analyze the relationship between the air quality in Beijing and its surrounding cities. From Table 1, the Granger causality conclusions can be drawn as follows:

1) PM2.5 in Baoding, Tianjin, Tangshan, and Zhangjiakou were the cause of PM2.5 in Beijing when the time lag was 2 or 3.
2) PM10 in Baoding, Chengde, and Tangshan were the cause of PM10 in Beijing when the time lag was 2 or 3.
3) SO2 in Baoding, Chengde, Tianjin, Tangshan, and Zhangjiakou were the cause of SO2 in Beijing when the time lag was 2 or 3.
4) CO in Chengde, Tianjin, Tangshan, and Zhangjiakou were the cause of CO in Beijing when the time lag was 3 or 4.
5) NO2 in Baoding, Tianjin, Tangshan, and Zhangjiakou were the cause of NO2 in Beijing when the time lag was 2 or 3.

Regardless of wind direction, wind speed, and other factors, air pollutants arrive in Beijing after 2-3 days of diffusion and have an impact on the air quality in Beijing. From Table 1, the kinds of pollutants in specific cities that have an impact on Beijing's air quality can also be identified, but the degrees of impact of the pollutants from these cities on Beijing's air pollution cannot be determined. In the next section, we will utilize MCC to estimate the degrees of impact of air pollutants around Beijing.

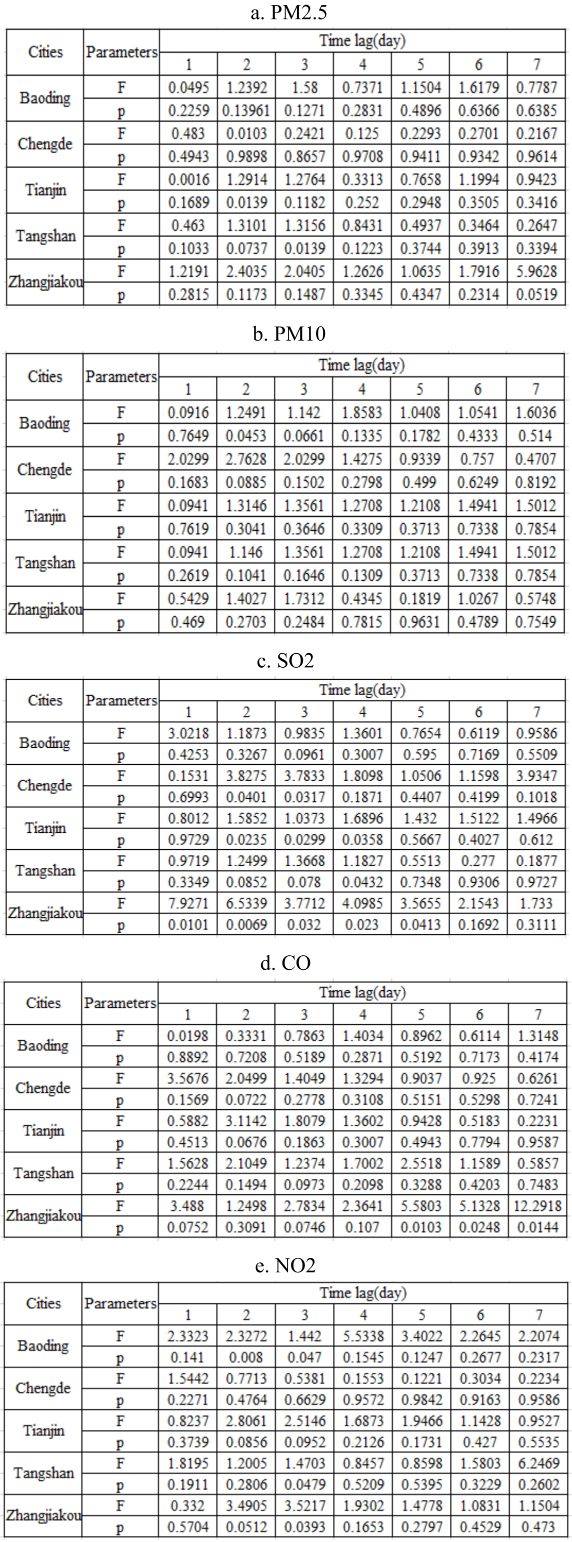
III. GCA BASED ON MCC

There are two traditional measures of similarity between two random variables: minimum error entropy (MEE) and MCC. In this section, MCC is used to determine which surrounding city has the greatest impact on Beijing's air pollution. Firstly, MCC is applied to identify the weights for the time series and guarantees the minimum error. Secondly, the X^{τ} which is the predominant causality of Y is determined through minimum error entropy. If there is causality between X^{τ} and Y, the correntropy is defined using the joint probability density function $f_{X^{\tau}Y}(x^{\tau}_{d,t}, y_{d,t})$ as:

$$\begin{aligned} V\left(X^{\tau}, Y\right) &= E\left(\kappa\left(X^{\tau}, Y\right)\right) \\ &= \frac{1}{MT}\sum_{d=1}^{M}\sum_{t=1}^{K}\iint \kappa(x^{\tau}_{d,t}, y_{d,t}) f_{X^{\tau}Y}(x^{\tau}_{d,t}, y_{d,t})dx^{\tau}_{d,t}dy_{d,t} \end{aligned} \tag{13}$$

Here $E(\cdot)$ is the expectation operator, and $\kappa(\cdot)$ denotes a shift-in-variant Mercer kernel. In this study, the Gaussian kernel

TABLE 1. Results of GCA for five pollutants.



a. PM2.5

Cities	Parameters	Time lag(day)						
		1	2	3	4	5	6	7
Baoding	F	0.0495	1.2392	1.58	0.7371	1.1504	1.6179	0.7787
	p	0.2259	0.13961	0.1271	0.2831	0.4896	0.6366	0.6385
Chengde	F	0.483	0.0103	0.2421	0.125	0.2293	0.2701	0.2167
	p	0.4943	0.9898	0.8657	0.9708	0.9411	0.9342	0.9614
Tianjin	F	0.0016	1.2914	1.2764	0.3313	0.7658	1.1994	0.9423
	p	0.1689	0.0139	0.1182	0.252	0.2948	0.3505	0.3416
Tangshan	F	0.463	1.3101	1.3156	0.8431	0.4937	0.3464	0.2647
	p	0.1033	0.0737	0.0139	0.1223	0.3744	0.3913	0.3394
Zhangjiakou	F	1.2191	2.4035	2.0405	1.2626	1.0635	1.7916	5.9628
	p	0.2815	0.1173	0.1487	0.3345	0.4347	0.2314	0.0519

b. PM10

Cities	Parameters	Time lag(day)						
		1	2	3	4	5	6	7
Baoding	F	0.0916	1.2491	1.142	1.8583	1.0408	1.0541	1.6036
	p	0.7649	0.0453	0.0661	0.1335	0.1782	0.4333	0.514
Chengde	F	2.0299	2.7628	2.0299	1.4275	0.9339	0.757	0.4707
	p	0.1683	0.0885	0.1502	0.2798	0.499	0.6249	0.8192
Tianjin	F	0.0941	1.3146	1.3561	1.2708	1.2108	1.4941	1.5012
	p	0.7619	0.3041	0.3646	0.3309	0.3713	0.7338	0.7854
Tangshan	F	0.0941	1.146	1.3561	1.2708	1.2108	1.4941	1.5012
	p	0.2619	0.1041	0.1646	0.1309	0.3713	0.7338	0.7854
Zhangjiakou	F	0.5429	1.4027	1.7312	0.4345	0.1819	1.0267	0.5748
	p	0.469	0.2703	0.2484	0.7815	0.9631	0.4789	0.7549

c. SO2

Cities	Parameters	Time lag(day)						
		1	2	3	4	5	6	7
Baoding	F	3.0218	1.1873	0.9835	1.3601	0.7654	0.6119	0.9586
	p	0.4253	0.3267	0.0961	0.3007	0.595	0.7169	0.5509
Chengde	F	0.1531	3.8275	3.7833	1.8098	1.0506	1.1598	3.9347
	p	0.6993	0.0401	0.0317	0.1871	0.4407	0.4199	0.1018
Tianjin	F	0.8012	1.5852	1.0373	1.6896	1.432	1.5122	1.4966
	p	0.9729	0.0235	0.0299	0.0358	0.5667	0.4027	0.612
Tangshan	F	0.9719	1.2499	1.3668	1.1827	0.5513	0.277	0.1877
	p	0.3349	0.0852	0.078	0.0432	0.7348	0.9306	0.9727
Zhangjiakou	F	7.9271	6.5339	3.7712	4.0985	3.5655	2.1543	1.733
	p	0.0101	0.0069	0.032	0.023	0.0413	0.1692	0.3111

d. CO

Cities	Parameters	Time lag(day)						
		1	2	3	4	5	6	7
Baoding	F	0.0198	0.3331	0.7863	1.4034	0.8962	0.6114	1.3148
	p	0.8892	0.7208	0.5189	0.2871	0.5192	0.7173	0.4174
Chengde	F	3.5676	2.0499	1.4049	1.3294	0.9037	0.925	0.6261
	p	0.1569	0.0722	0.2778	0.3108	0.5151	0.5298	0.7241
Tianjin	F	0.5882	3.1142	1.8079	1.3602	0.9428	0.5183	0.2231
	p	0.4513	0.0676	0.1863	0.3007	0.4943	0.7794	0.9587
Tangshan	F	1.5628	2.1049	1.2374	1.7002	2.5518	1.1589	0.5857
	p	0.2244	0.1494	0.0973	0.2098	0.3288	0.4203	0.7483
Zhangjiakou	F	3.488	1.2498	2.7834	2.3641	5.5803	5.1328	12.2918
	p	0.0752	0.3091	0.0746	0.107	0.0103	0.0248	0.0144

e. NO2

Cities	Parameters	Time lag(day)						
		1	2	3	4	5	6	7
Baoding	F	2.3323	2.3272	1.442	5.5338	3.4022	2.2645	2.2074
	p	0.141	0.008	0.047	0.1545	0.1247	0.2677	0.2317
Chengde	F	1.5442	0.7713	0.5381	0.1553	0.1221	0.3034	0.2234
	p	0.2271	0.4764	0.6629	0.9572	0.9842	0.9163	0.9586
Tianjin	F	0.8237	2.8061	2.5146	1.6873	1.9466	1.1428	0.9527
	p	0.3739	0.0856	0.0952	0.2126	0.1731	0.427	0.5535
Tangshan	F	1.8195	1.2005	1.4703	0.8457	0.8598	1.5803	6.2469
	p	0.1911	0.2806	0.0479	0.5209	0.5395	0.3229	0.2602
Zhangjiakou	F	0.332	3.4905	3.5217	1.9302	1.4778	1.0831	1.1504
	p	0.5704	0.0512	0.0393	0.1653	0.2797	0.4529	0.473

was adopted by:

$$\kappa(x^{\tau}_{d,t}, y_{d,t}) = \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left(\frac{\left(e^{\tau}_{d,2}(t)\right)^2}{2\sigma_d^2}\right) \tag{14}$$

$$\begin{aligned} e^{\tau}_{d,2}(t) &= y_{d,t} - \sum_{i=1}^{p_3} \alpha^2_{d,i} x^{\tau}_{t-i} - \sum_{j=1}^{p_4} \beta^2_{d,j} y_{d,t-j} \\ &= y_{d,t} - W_d^T X_d(t) \end{aligned} \tag{15}$$

where σ_d^2 denotes the variance, which is a constant at temporal space. $W_d = \left(w_1 \; w_2 \; \cdots \; w_m\right)^T$ is the weight, and $X_d(t)$ is defined as:

$$X_d(t) = \left[x^{\tau}_{d,t}, x^{\tau}_{d,t-1}, \cdots, x^{\tau}_{d,t-m+1}\right] \tag{16}$$

The MCC algorithm can be derived through a stochastic gradient as:

$$W_d(i) = W_d(i-1) + \eta_d f(e_d(i)) X_d(i) \tag{17}$$

Here $W_d(i)$ is the weight vector at iteration i, $\eta_d > 0$ denotes the step-size, $e_d(i)$ is the prediction error of $e^{\tau}_{d,2}(i)$ at iteration i, and $f(e_d(i))$ is a function of the error $e_d(i)$. In general, $f(e_d(i))$ can be defined as:

$$f(e_d(i)) = \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left(-\frac{e_d^2(i)}{2\sigma_d^2}\right) e_d(i) \tag{18}$$

Substituting (18) into (17), the gradient-based update equation can easily be derived to maximize the correntropy, which is the minimum error entropy. Hence, (17) can be rewritten as:

$$W_d(i) = W_d(i-1) + \frac{1}{\sqrt{2\pi}\sigma_d} \eta_d \exp\left(-\frac{e_d^2(i)}{2\sigma_d^2}\right) e_d(i) X_d(i) \tag{19}$$

In order to facilitate further research, the weight error vector $\tilde{W}_d(i-1)$ at iteration $i-1$ can be expressed as:

$$\tilde{W}_d(i-1) = W_d^0 - W_d(i-1) \tag{20}$$

Here W_d^0 denotes the desired weight vector that was not known beforehand. Clearly, $W_d(i)$ is closest to W_d^0 when the norm of $\tilde{W}_d(i)$ is a minimum. Based on equation (16), the priori error can be represented by:

$$e_a(i) = \tilde{W}_d(i-1)^T X_d(i) \tag{21}$$

Clearly, $e_a(i)$ and $e_d(i)$ have the following relationship:

$$e_d(i) = e_a(i) + \left(W_d(i-1)^T - W_d(i)^T\right) X_d(i) \tag{22}$$

A direct consequence of the energy conservation relation [25] being applied, leads to the equation:

$$\begin{aligned} E\left[\left\|\tilde{W}_d(i)\right\|^2\right] &= E\left[\left\|\tilde{W}_d(i-1)\right\|^2\right] - 2\eta_d E\left[e_a(i) f(e_d(i))\right] \\ &\quad + \eta_d^2 E\left[\|X_d(i)\|^2 f^2(e_d(i))\right] \end{aligned} \tag{23}$$

Based on the analysis,

$$\begin{aligned} W_d^0 &= \underset{W_d^0}{\arg\min}\, E\left[\left\|\tilde{W}_d(i)\right\|^2\right] \\ &= \underset{W_d^0}{\arg\min} \left(E\left[\left\|\tilde{W}_d(i-1)\right\|^2\right] - 2\eta_d E\left[e_a(i) f(e_d(i))\right]\right. \\ &\quad \left. + \eta_d^2 E\left[\|X_d(i)\|^2 f^2(e_d(i))\right]\right) \end{aligned} \tag{24}$$

Through the properties of the function, if W_d^0 satisfies the condition $\eta_d E\left[\|X_d(i)\|^2 f^2(e_d(i))\right] = E\left[e_a(i)f(e_d(i))\right]$, the minimum of equation (23) can be solved for as follows:

$$\min E\left[\left\|\tilde{W}_d(i)\right\|^2\right] = E\left[\left\|\tilde{W}_d(i-1)\right\|^2\right] - \frac{\left(E\left[e_a(i)f(e_d(i))\right]\right)^2}{E\left[\|X_d(i)\|^2 f^2(e_d(i))\right]} \tag{25}$$

We can then easily obtain the optimum of the weight W_d, assuming that $\gamma = \frac{1}{M}\sum_{d=1}^{M} E\left[\left\|\tilde{W}_d(i)\right\|^2\right]$ is the degree of impact of X^τ on Y. This process is repeated to calculate the degree of impact of $X^\tau(\tau = 1, 2, 3, 4, 5)$ on Y. X^{τ_0} has the greatest impact on Y if the expectation of the norm for the weight error $\tilde{W}_d(i)$, $(d = 1, 2, \cdots M)$ is a minimum between X^{τ_0} and Y.

Due to the introduction of CS, large volumes of data can be handled, such as the five major pollutants for Beijing and its surrounding cities in 2019. The compression ratio $\frac{M}{N} = 0.65$ was applied. The degrees of impact of the surrounding cities are shown in Table 2 (Surrounding cities without Granger causality were excluded).

As shown in the TABLE 2, among PM2.5 and SO2, Tangshan has the minimal degree of impact. Therefore Tangshan has the greatest impact on Beijing's air quality. Similarly, Baoding has the greatest impact on Beijing in PM10 and NO2, and Tianjin has the greatest impact on Beijing in CO.

IV. PERFORMANCE ANALYSIS

A. MEAN-SQUARE STABILITY

The mean-square stability is an important index for GCA, which has been extensively studied in the literature [26]. In order to evaluate the stability of the weight error $\tilde{W}_d(i)$, it was assumed that the prediction error $\{e_d(i)\}$ was zero-mean Gaussian distributed, with variance σ_d^2, independent of the $\{X_d(i)\}$.

Now, equations (18) and (19) are used to analyze the mean-square stability of the weight update (19). From (23),

$$\begin{aligned}\lim_{i\to\infty} E\left[\left\|\tilde{W}_d(i)\right\|^2\right] &= \lim_{i\to\infty}\left(E\left[\left\|\tilde{W}_d(i-1)\right\|^2\right] - 2\eta E\left[e_a(i)f(e_d(i))\right]\right.\\ &\quad \left. + \eta_d^2 E\left[\|X_d(i)\|^2 f^2(e_d(i))\right]\right)\end{aligned} \tag{26}$$

Assuming that the weight update (19) is stable, we can obtain:

$$\lim_{i\to\infty} E\left[\left\|\tilde{W}_d(i)\right\|^2\right] = \lim_{i\to\infty} E\left[\left\|\tilde{W}_d(i-1)\right\|^2\right] \tag{27}$$

Steady state in (23) is then:

$$\eta_d E\left[\|X_d(i)\|^2 f^2(e_d(i))\right] \le 2E\left[e_a(i)f(e_d(i))\right] \tag{28}$$

TABLE 2. Degrees of impact.

Pollutants	Baoding	Chengde	Tianjin	Tangshan	Zhangjiakou
PM2.5	0.0321		0.0383	0.0214	0.0332
PM10	0.0387	0.0415		0.0523	
SO2	0.0343	0.0413	0.0216	0.0127	0.0279
CO		0.0142	0.0128	0.0232	0.0155
NO2	0.0173		0.0192	0.0202	0.0307

Considering the assumption and equation (22), we can derive:

$$\begin{aligned}&\lim_{i\to\infty} E\left[e_a(i)f(e_d(i))\right]\\ &= \lim_{i\to\infty}\left[E\left(e_d(i)f(e_d(i))\right) - E\left(\left(W_d(i-1)^T - W_d(i)^T\right)\right.\right.\\ &\quad \left.\left.\times\ \left(X_d(i)f(e_d(i))\right)\right)\right]\\ &= \lim_{i\to\infty}\int_{-\infty}^{+\infty} e_d^2(i)\exp\left(-\frac{e_d^2(i)}{\sigma_d^2}\right)de_d(i)\\ &= \frac{1}{4\sqrt{\pi}\sigma_d}\end{aligned} \tag{29}$$

In this way,

$$\begin{aligned}&\lim_{i\to\infty} E\left[\|X_d(i)\|^2 f^2(e_d(i))\right]\\ &= \lim_{i\to\infty} E\left(\|X_d(i)\|^2\right)E\left(f^2(e_d(i))\right)\\ &= \frac{1}{2\sqrt{3}\pi\sigma_d^2}E\left(\|X_d(i)\|^2\right)\end{aligned} \tag{30}$$

By substituting (25) and (26) into (24), we have:

$$\eta_d \le \frac{\sqrt{3\pi}\sigma_d}{E\left(\|X_d(i)\|^2\right)} \tag{31}$$

Note that if step-size η_d satisfies condition (31), $E\left[\left\|\tilde{W}(i)\right\|^2\right]$ will be decreasing, and hence stable.

B. RESULT VALIDATION

Transfer entropy (TE), in terms of stability and accuracy, is a very useful tool in quantifying directional causality for both linear and nonlinear relationships. To verify the accuracy of the results based on MCC, TE was applied to explore the air pollutant causality between the surrounding cities and Beijing. $TE_{X^\tau\to Y}$ is defined from conditional entropies as

$$TE_{X^\tau\to Y} = \frac{1}{M}\sum_{d=1}^{M} \mathrm{P}(y_{d,t+u}\,|y_{d,t}) - \mathrm{P}(y_{d,t+u}\,|x_{d,t}, y_{d,t}) \tag{32}$$

where $y_{d,t} = \left(y_{d,t}\ y_{d,t-1} \cdots y_{d,t-m+1}\right)$ is an m-dimensional vector, and $x_{d,t} = \left(x_{d,t}\ x_{d,t-1} \cdots x_{d,t-m+1}\right)$

is an m-dimensional vector, $\mathrm{P}(y_{d,t+u}\,|y_{d,t})$ is the entropy of the conditional process Y in its past, and can be calculated as:

$$\mathrm{P}(y_{d,t+u}\,|y_{d,t}) = -\sum_{y_{d,t+u}} \mathrm{P}\left(y_{d,t+u}, y_{d,t}, x_{d,t}\right) \cdot \log \mathrm{P}(y_{d,t+u}\,|y_{d,t}) \tag{33}$$

Similarly,

$$\mathrm{P}(y_{d,t+u}\,|x_{d,t}, y_{d,t}) = -\sum_{y_{d,t+u}} \mathrm{P}\left(y_{d,t+u}, y_{d,t}, x_{d,t}\right) \times \log \mathrm{P}(y_{d,t+u}\,|x_{d,t}, y_{d,t}) \tag{34}$$

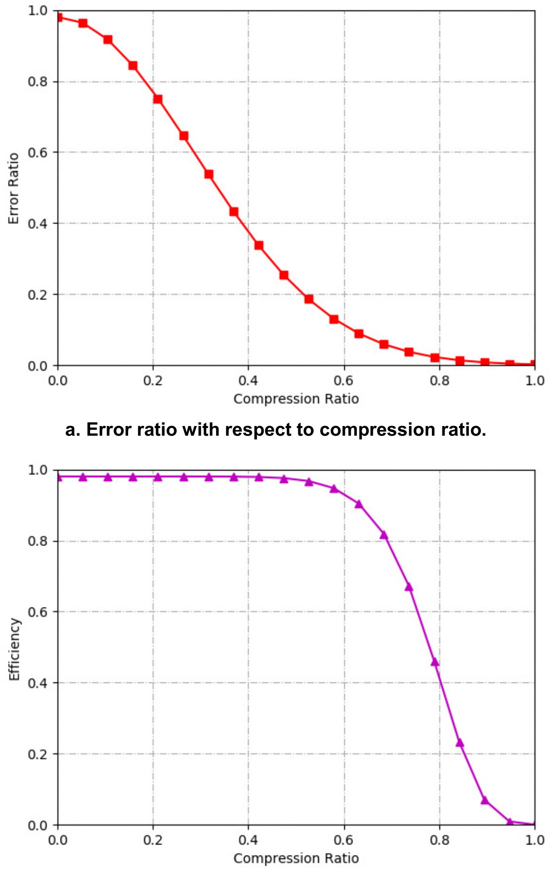
According to the properties of conditional probabilities, equations (33) and (34) are then substituted into equation (32),

$$\begin{aligned} TE_{X^\tau \to Y} &= \frac{1}{M}\sum_{d=1}^{M}\sum_{y_{d,t+u}} \mathrm{P}\left(y_{d,t+u}, y_{d,t}, x_{d,t}\right) \times \log \frac{\mathrm{P}(y_{d,t+u}\,|x_{d,t}, y_{d,t})}{\mathrm{P}(y_{d,t+u}\,|y_{d,t})} \\ &= \frac{1}{M}\sum_{d=1}^{M}\sum_{y_{d,t+u}} \mathrm{P}\left(y_{d,t+u}, y_{d,t}, x_{d,t}\right) \times \log \frac{\mathrm{P}\left(y_{d,t+u}, y_{d,t}, x_{d,t}\right)\mathrm{P}(y_{d,t})}{\mathrm{P}(y_{d,t+u}, y_{d,t})\mathrm{P}\left(y_{d,t}, x_{d,t}\right)} \end{aligned} \tag{35}$$

According to equation (35), the impact of X^τ on Y can be calculated. $TE_{X^\tau \to Y}(\tau = 1, 2, 3, 4, 5)$ is then calculated, enabling the determination of which X^τ has the greatest impact on Y by the maximum $TE_{X^\tau \to Y}$. By calculation, Tangshan has the greatest impact on Beijing's air quality in terms of PM2.5 and SO2, Baoding has the greatest impact on Beijing in terms of PM10 and NO2, and Tianjin has the greatest impact on Beijing in terms of CO. This result is consistent with the result of MCC (TABLE 2).

V. EXPERIMENTS

To verify the effectiveness of our algorithm for CS and GCA, experiments were conducted using Python 3.7, and data was derived from the Online Monitoring and Analysis Platform for Air Quality in China [27]. According to the comparative analysis of a large number of data, it is found that the data sampled every 4 hours are concentrated in distribution. For the convenience of data compression, the elements in temporal compressive basis are all 1, that is, the average value of 6 groups data in every day is taken as the causal analysis data. Therefore, only the spatial compression ratio is considered. When the data is compressed in the temporal space, the average value of the 6 data sampled in a day is acquired, so it is equivalent to adopting only one data every day. Therefore, the step size $\eta_d = 1$. Since our main goal is causal analysis rather than data compression; we only use traditional sensing matrix Φ and sparse basis Ψ to achieve preliminary data processing.



a. Error ratio with respect to compression ratio.

b. Efficiency with respect to compression ratio.

FIGURE 4. Tradeoff between error ratio and efficiency.

To facilitate the experiment, $\sigma_d^2 = 1$. For implementing the analysis, the following steps had to be completed:

1) Original data from the open official website was sampled.
2) The original data was compressed based on CS.
3) The causality of pollutants between Beijing and its surrounding cities was analyzed based on GCA.
4) The pollutants in each city which had the greatest impact on Beijing's air quality were determined based on MCC.

To achieve these processes, the related problems had to be addressed from two perspectives.

1) Because the data set to be processed was large, coupled with the fact that CS is a lossy compression process, the data compression accuracy varied with the compression ratio.
2) The efficiency of data processing and analysis was greatly affected by the compression ratio, causing a kind of tradeoff between error ratio, efficiency, and compression ratio.

From Fig.4a, it can be seen that the error ratio decreased as the compression ratio increased. When the compression ratio increased to 65%, the error ratio dropped to almost 5%. Similarly, the efficiency decreased as the compression ratio increased as shown in Fig.4b. When the compression ratio increased to 65%, the efficiency was still 90%. As the compression ratio continued to increase, the efficiency dropped sharply. Therefore, a tradeoff point was reached when the compression ratio reached 65%.

VI. CONCLUSION

In this paper, we proposed a novel method of causal identification based on compressive sensing of air pollutants using urban big data. The compression model was developed to compress spatiotemporal correlation data representing the amount of air pollutants in Beijing and its surrounding cites. By extending the existing Granger causality theory, the spatiotemporal Granger causality analysis model was adopted for algorithm implementation to identify which cities had an impact on the air quality in Beijing. Specifically, degrees of impact were determined by applying the MCC among the surrounding cities. To verify the effectiveness of the algorithm, the mean-square stability of GCA was confirmed and the accuracy of the degrees of impact of the MCC was obtained based on transfer entropy. By conducting real meteorological big sensing data experiments, it was demonstrated that our proposed algorithm significantly improved data causality analysis performance within the allowed error ratio.

In fact, we have discussed the linear causality of air pollutants between Beijing and surrounding cities. Next, we will continue to study the nonlinear causality between them, which is of great significance for the study of air pollution.

REFERENCES

[1] C. Wu, W. Hu, M. Zhou, S. Li, and Y. Jia, "Data-driven regionalization for analyzing the spatiotemporal characteristics of air quality in China," *Atmos. Environ.*, vol. 203, no. 15, pp. 172–182, Apr. 2019.

[2] G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air quality prediction: Big data and machine learning approaches," *Int. J. Environ. Sci. Develop.*, vol. 9, no. 1, pp. 8–16, Jan. 2018.

[3] K. Ahuja and N. N. Jani, "Air quality prediction data-model formulation for urban areas," in *Proc. Int. Conf. Comput. Netw. Commun. Technol.*, Singapore, Sep. 2018, pp. 111–118.

[4] Y. Zhang, Y. Wang, M. Gao, Q. Ma, J. Zhao, R. Zhang, Q. Wang, and L. Huang, "A predictive data feature exploration-based air quality prediction approach," *IEEE Access*, vol. 7, pp. 30732–30743, 2019.

[5] X. Li and Z. Zhang, "Research and analysis for real-time streaming big data based on controllable clustering and edge computing algorithm," *IEEE Access*, vol. 7, pp. 171621–171632, 2019.

[6] Y. Zheng, C. Mascolo, and C. T. Silva, "Guest editorial: Urban computing," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 124–125, Jun. 2017.

[7] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, Aug. 2014, pp. 1027–1036.

[8] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Sydney, NSW, Australia, Aug. 2015, pp. 2267–2276.

[9] A. S. Mihăiţă, L. Dupont, O. Chery, M. Camargo, and C. Cai, "Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling," *J. Cleaner Prod.*, vol. 221, pp. 398–418, Jun. 2019.

[10] J. Y. Zhu, C. Sun, and V. O. K. Li, "An extended spatio-temporal Granger causality model for air quality estimation with heterogeneous urban big data," *IEEE Trans. Big Data*, vol. 3, no. 3, pp. 307–319, Sep. 2017.

[11] I. Ebert-Uphoff and Y. Deng, "Causal discovery from spatio-temporal data with applications to climate science," in *Proc. 13th Int. Conf. Mach. Learn. Appl.*, Detroit, MI USA, Dec. 2014, pp. 606–613.

[12] D. Yang, H. Chen, Y. Song, and Z. Gong, "Granger causality for multivariate time series classification," in *Proc. IEEE Int. Conf. Big Knowl. (ICBK)*, Hefei, China, Aug. 2017, pp. 103–110.

[13] Z. Peng, W. Liu, and S. An, "Haze pollution causality mining and prediction based on multi-dimensional time series with PS-FCM," *Inf. Sci.*, vol. 523, pp. 307–317, Jun. 2020.

[14] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[15] A. Payani, A. Abdi, X. Tian, F. Fekri, and M. Mohandes, "Advances in seismic data compression via learning from data: Compression for seismic data acquisition," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 51–61, Mar. 2018.

[16] Y. Zheng, W. Wang, and M. Fan, "Learning a limited text space for cross-media retrieval," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, in Lecture Notes in Computer Science, vol. 10424, Jul. 2017, pp. 292–303.

[17] P. Lindstrom and M. Isenburg, "Fast and efficient compression of floating-point data," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 1245–1250, Sep. 2006.

[18] A. Shojaie and G. Michailidis, "Discovering graphical Granger causality using the truncating lasso penalty," *Bioinformatics*, vol. 26, no. 18, pp. 517–523, Sep. 2010.

[19] G. Yang, L. Wang, and X. Wang, "Network reconstruction based on grouped sparse nonlinear graphical Granger causality," in *Proc. 35th Chin. Control Conf. (CCC)*, Chengdu, China, Jul. 2016, pp. 2229–2234.

[20] M. D'Aló, S. Falorsi, and F. Solari, "Space-time unit-level EBLUP for large data sets," *J. Off. Statist.*, vol. 33, no. 1, pp. 61–77, Mar. 2017.

[21] L. Luo, W. Liu, I. Koprinska, and F. Chen, "Discovering causal structures from time series data via enhanced Granger causality," in *Proc. Australas. Joint Conf. Artif. Intell.*, Canberra, ACT, Australia, Nov. 2015, pp. 365–378.

[22] J. Y. Zhu, C. Zhang, H. Zhang, S. Zhi, V. O. K. Li, J. Han, and Y. Zheng, "Pg-causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data," *IEEE Trans. Big Data*, vol. 4, no. 4, pp. 571–585, Dec. 2018.

[23] C. Aviles-Cruz, E. Rodriguez-Martinez, J. Villegas-Cortez, and A. Ferreyra-Ramirez, "Granger-causality: An efficient single user movement recognition using a smartphone accelerometer sensor," *Pattern Recognit. Lett.*, vol. 125, pp. 576–583, Jul. 2019.

[24] B. Chen, X. Wang, Y. Li, and J. C. Principe, "Maximum correntropy criterion with variable center," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1212–1216, Aug. 2019.

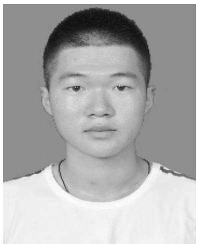
[25] N. Takahashi and I. Yamada, "Steady-state mean-square performance analysis of a relaxed set-membership NLMS algorithm by the energy conservation argument," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3361–3372, Sep. 2009.

[26] B. Chen, L. Xing, B. Xu, H. Zhao, N. Zheng, and J. Príncipe, "Kernel risk-sensitive loss: Definition, properties and application to robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2888–2900, Jun. 2017.

[27] *Statistics of Air Quality in Beijing*. Accessed: Nov. 26, 2019. [Online]. Available: https://www.aqistudy.cn/Period

MINGWEI LI received the M.S. degree in operations research and control theory and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2008 and 2014, respectively. Since 2017, she has been an Assistant Professor with Northeastern University at Qinhuangdao. Her research interests include data analysis, big data processing, and compressive sensing.

JINPENG LI received the B.S. degree in information and computing sciences from Shanxi Agricultural University, China, in 2018. He is currently pursuing the master's degree with Northeastern University, China. His research interests include applied mathematics, artificial intelligence, machine learning, data mining, and programming languages.



SHUANGNING WAN is currently pursuing the bachelor's degree in learning applied statistics with Northeastern University at Qinhuangdao (NEUQ). His current research interests include data mining, data handling, and programming languages.

HAO CHEN is currently pursuing the bachelor's degree in information and computing science with Northeastern University at Qinhuangdao (NEUQ). His current research interests include machine learning, web development, and programming languages.

CHAO LIU received the M.S. degree in fundamental mathematics and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2006 and 2009, respectively. Since 2011, he has been an Assistant Professor with Northeastern University at Qinhuangdao. His research interests include modeling and dynamical analysis of the stochastic biological and infectious disease systems.