Smoothing of time-varying graph with the generalized LASSO

WENYU LIU1*, HAO YAN1*, QING WANG2, NAN HU3, NI LIU1, FANGDING XU1, AND JUNBO DUAN1†

1 Key Laboratory of Biomedical Information Engineering of Ministry of Education and Department of Biomedical Engineering, School of Life Science and Technology, Xi’an Jiaotong University, Xi’an 710049, China
2School of Electronic Engineering, Xidian University, Xi’an 710071, China
3First Affiliated Hospital, Xi’an Jiaotong University, Xi’an 710049, China
*Those authors contribute equally.
†Corresponding author: Junbo Duan (junbo.duan@mail.xjtu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 61771381.

ABSTRACT This paper proposes a joint model based on the generalized LASSO to smooth a time-varying graph. The model generalizes the gLASSO from a purely spatial setting to a spatial-temporal one. In the proposed model, the first term measures the fitting error, while the second term incorporates the structural information of graphs and total variations of time sequence, and hence the model can extract both temporal and spatial information. To illustrate the performance of the proposed model, we analyzed the simulated datasets for epidemic diseases and the real datasets for COVID-19 and mortality rate in mainland China. The results show that the proposed model can capture the trends/regions simultaneously in both temporal and spatial domains, being an effective model to analyze the problems that can be modelled as time-varying graphs.

INDEX TERMS Generalized LASSO, penalized regression, spatiotemporal smooth, time-varying graph.

I. INTRODUCTION

The least absolute shrinkage and selection operator (LASSO) is a linear model which has been widely studied and used in many areas [1]. The LASSO also can be used in denoising and smoothing, which is defined by

\[
\beta(\lambda) = \arg \min_{\beta} \left\{ \frac{1}{2} ||y - X\beta||_2^2 + \lambda ||\beta||_1 \right\} \tag{1}
\]

where \(y\) is the data vector, \(X\) is the dictionary matrix, \(\beta\) is the coefficient vector, \(\lambda\) is a tuning parameter which controls the tradeoff between the fitting goodness and sparsity in LASSO solution. \(\lambda\) ranges from 0 to +\(\infty\) [2], and it can be tuned automatically and efficiently by using model selection criteria [3].

To solve the problems appearing in statistics or informatics likes denoising and smoothing, the model was enhanced by adding the constrain of structure instead of pure sparse on the coefficients in linear regression. The generalized LASSO(gLASSO) was proposed by the formulation [4]:

\[
\beta(\lambda) = \arg \min_{\beta} \left\{ \frac{1}{2} ||y - X\beta||_2^2 + \lambda ||D\beta||_1 \right\} \tag{2}
\]

where \(D\) is a specified penalty matrix. \(D\) contains the structure information of \(\beta\), and it’s the famous total variation [5] that assumes \(\beta\) is piecewise smooth. gLASSO solves the two-dimensional problems with the sparse vector \(D\beta\). For overdetermined case of \(D\), generalized LASSO imposes sparseness on \(\beta\) by mapping with \(D\). A gLASSO can be transformed into a LASSO with an equality constraint, which falls into the constrained LASSO (cLASSO) [6], [7].

Due to the flexibility and generality of the gLASSO, it has attracted much attention. The choice of \(D\) depends on the application. Some well-known applications include the fused LASSO [8], trend filtering [9] and generalized fused LASSO [10]. However, in various fields, such as epidemiology, meteorology, ecology, and economics, it is important to consider the temporal and spatial connections. In the above fields, many problems not only have structural information, but also are time-varying, or in other words, the dataset is a set of time series, and these time series also have a structure.
that can be described with a graph. But gLASSO only contains structural information and lacks temporal information.

In the literature, most of the research on the gLASSO focus on one- or two-dimensional problems. Tibshirani et al. [4] proposed an application method of gLASSO, which analyzed the H1N1 flu cases for each (mainland) US state in 2009 and categorize the infection rates into different trend. But the method only analyzed the infection rates in geospatial trend. Hao Wang et al. [11] analyzed the ZIP Code Tabulation Areas (ZCTAs) of residence of the patient as well as the year of diagnosis and discuss the identification of pediatric cancer clusters in Florida between 2000 and 2010 using Poisson model and gLASSO penalty. Although the data are spatiotemporal in nature, Hao Wang et al. aggregate the data on each ZCTA over time and ignore the temporal component.

M. Ohishi et al. [12] proposed a method which apply a coordinate descent algorithm to solve the optimization problem for gLASSO and estimated regional effects at 852 areas split Tokyo’s 23 wards using the proposed method. The data involved the changes of rents and environmental conditions of studio apartments in 23 wards in Tokyo over time, but M. Ohishi et al. only analyzed the structural information of the data.

To solve the three-dimensional problems of this category, we proposed a spatiotemporal smoothing model that generalizes the gLASSO from a spatial setting to a spatiotemporal one. It will give an interesting and sparse solution when the data imply some structural properties. The function of \( \lambda \) is similar to the generalized LASSO, which contributes to smoothing on structure. In addition, a new parameter \( \mu \) for \( D \) is introduced in our spatiotemporal smoothing model to control the weight between time and graphical information.

The rest of this paper is organized as follows: in section II, we present the proposed spatiotemporal smoothing model; in section III, we designed a simulation experiment to verify the reliability of the model; in section IV, we illustrate the applications of the proposed model on two specific examples; and we conclude the paper in section V.

II. THE PROPOSED MODEL

Here we introduce the spatiotemporal smoothing model. A data matrix \( Y = \{y_1, y_2, \ldots, y_N\} \in \mathbb{R}^{P \times N} \), each \( y_i \in \mathbb{R}^P \) is a data vector, which stores the responses of \( P \) nodes in a graph. The data of nodes in the graph are time-varying. \( N \) denotes the number of the time spots, which equals to the number of graphs. Consequently, \( Y \) follows a two-dimension structure. To fit our model, we transform the \( Y \) to a cascade structure \( y = \{y_1; y_2; \ldots; y_N\} \in \mathbb{R}^{PN} \).

\( X \) is the dictionary matrix of smoothing with full rank. In our proposed model, we adapt the “signal approximation” case in the generalized LASSO, \( X = I \in \mathbb{R}^{NP \times NP} \).

\( D \) implies the temporal and spatial information of \( Y \). \( D \) consists of \( D_S \), \( D_T \) and \( \mu \). \( D = \begin{bmatrix} D_S & 0 & \cdots & 0 \\ 0 & D_S & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_S \end{bmatrix} \in \mathbb{R}^{(NM+(N-1)P) \times NP} \). 

\( D_S \in \mathbb{R}^{NM \times NP} \) contains the spatial information, and \( D_T \in \mathbb{R}^{(N-1)P \times NP} \) contains temporal information. The spatial information and temporal information have different influence on the data, hence we add an extra parameter \( \mu \) to coordinate their effects.

\( D_S \) represents the structure of graph, which consists of \( D_S \). \( D_S \in \mathbb{R}^{M \times NP} \) represents undirected graph structure. \( P \) columns of \( D_S \) represents the nodes in graph and \( M \) rows of \( D_S \) represent the edges in graph. \( D_S \) has 1 and −1 in a row and it represents the relationship between two nodes, where 1 and −1 are located. The number of \( D_S \) is related to the number of graphs.

\[
D_S = \begin{bmatrix} D_S & 0 & \cdots & 0 \\ 0 & D_S & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_S \end{bmatrix}
\]  

(3)

\( D_S \) has \( N \) matrices \( D_S \) in diagonal line and \( D_S \in \mathbb{R}^{NM \times NP} \). \( D_T \) contains the temporal information and \( D_T \) adopts the idea of total variation [5]. \( D_T \) consists of \( I \), which is \( P \times P \) identity matrix. A pair of \( I \) and \( -I \) represent the connection of the two graphs, which are adjacent in time. The number of time spots depends on the number of \( I \). If there has \( N \) graphs, the number of pairs \( I \) is \( N-1 \). \( D_T \) can be described as follows,

\[
D_T = \begin{bmatrix} I & -I & 0 & \cdots & 0 \\ 0 & I & -I & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I & -I \end{bmatrix}
\]  

(4)

\( D_T \in \mathbb{R}^{(N-1)P \times NP} \) is a block diagonal matrix.

The function of \( \mu \) is to weight the temporal information and spatial information. \( \mu \) ranges from 0 to \( +\infty \). If \( \mu \) approaches to zero, the spatial structure has a strong effect, the smoothing in space will get a better solution. If \( \mu \) is large, the temporal information is dominating.

FIGURE 1. A toy example graph with four nodes

To illustrate the formation of \( D \), we start with a toy model, whose graph is showed in Fig. 1. As is shown, \( y_1 \) has 4 nodes and 4 edges, hence \( D_S \) is of size 4 \( \times \) 4. The columns and rows represent the nodes and edges in Fig. 1 respectively, and the \( D_S \) reads

\[
\begin{bmatrix}
D_S & 0 & \cdots & 0 \\
0 & D_S & \ddots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & D_S
d\end{bmatrix}
\]
\[ D_S = \begin{bmatrix}
    1 & 0 & -1 & 0 \\
    1 & 0 & 0 & -1 \\
    0 & 1 & -1 & 0 \\
    0 & 0 & 1 & -1
\end{bmatrix} \]

If the nodes in the graph changes once, \( D_S \) will be of size \( 8 \times 8 \), and \( D_T \) of size \( 4 \times 8 \). \( D_S \) and \( D_T \) read as

\[ D_S = \begin{bmatrix}
    1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
    1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
    0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
    0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
    0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
    0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \]

\[ D_T = \begin{bmatrix}
    1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
    0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\
    0 & 0 & 0 & 1 & 0 & 0 & 0 & -1
\end{bmatrix} \]

Thus \( D \) reads

\[ D = \begin{bmatrix}
    1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
    1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
    0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\
    0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
    0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
    0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
    \mu & 0 & 0 & 0 & -\mu & 0 & 0 & 0 \\
    0 & \mu & 0 & 0 & 0 & -\mu & 0 & 0 \\
    0 & 0 & \mu & 0 & 0 & 0 & -\mu & 0 \\
    0 & 0 & 0 & \mu & 0 & 0 & 0 & -\mu
\end{bmatrix} \]

III. SIMULATION EXPERIMENT

In order to verify the reliability in space and time of the model, we designed the following simulation experiment. Taking the number of infected subjects in an epidemic disease as an example, the experiment simulated five different trends of increasing in the number of infections in one month (30 days) across 31 provinces on the Chinese mainland. As shown in the Fig. 2.

Each trend encompasses several neighbor provinces, which is ordered according to the alphabet: (i) Exponential growth trend encompasses Fujian, Guangdong, Guangxi, Guizhou, Hainan, and Yunnan. (ii) Linear growth trend encompasses Anhui, Chongqing, Hubei, Hunan, Jiangxi, and Zhejiang. (iii) Logarithmic growth trend encompasses Henan, Jiangsu, Shaanxi, Shandong, Shanghai, and Shanxi. (iv) Increased at first and then decreased trend encompasses Beijing, Hebei, Heilongjiang, Jilin, Liaoning, Inner Mongolia, and Tianjin. (v) Sprinkle trend encompasses Gansu, Qinghai, Ningxia, Sichuan, Xinjiang, and Xizang.

Random noises are added to each province to simulate the spatial differences between provinces. As shown in the Fig. 3. In this experiment, the simulated data are time-varying and have structure information. We believe that through the model, we can cluster provinces under the same growth trend and smooth out the noise.
FIGURE 3. Simulation of the increasing trend of the number of infections in mainland China.

Specifically, we assigned the provinces as nodes, which means \( P = 31 \). If two provinces share a border, there is an edge between two corresponding nodes and \( M = 69 \). The dataset spans 30 days and \( N = 30 \). So \( D_S \) is of size \( 2070 \times 930 \). \( D_T \) is of size \( 899 \times 930 \). The data \( Y \) is the vectorized version of \( y_i (i = 1, 2, \ldots, 30) \), which is in chronological order. \( y_i \) contains the infections data of all the nodes in one day. \( D \) is of size \( 2969 \times 930 \).

To get a satisfying solution, the parameters \( \mu \) and \( \lambda \) must be chosen carefully. After testing different ways to select the parameter, we found the relationship between the two parameters is complicated, and the results of model selection information criteria are not satisfying. As a result, we manually selected \( \mu = 0.8, \lambda = 5 \), and the following results are based on this parameter setting. The detail of parameter selection is discussed in the last section.

To solve the gLASSO problem, the \textit{genlasso} package[13] in R studio was used, and the \textit{ggplot2} package[14] was used to visualize the data. Parameters \textit{maxsteps} and \textit{eps} were set as \( 5000 \) and \( 10^{-6} \) in \textit{genlasso}, and other parameters were left default. The original infections data in the first day is shown in Fig. 4(a), and the smoothed result is shown in Fig. 4(b). Comparing Fig. 4(a) with Fig. 4(b), we can see that after smoothing, the infections are clustered roughly into two regions. The provinces with exponential growth, linear growth, first increased and then decreased, and sprinkles trends are classified as one category, while provinces with logarithmic growth trend are classified as another. This is because the logarithmic growth trend starts at a much faster rate than the other growth trends. Fig. 4(c) and Fig. 4(d) show the original and smoothed infection data on day 28 respectively. We can clearly see that after smoothing, infections in provinces are clustered into 5 groups.
FIGURE 4. The geographical distribution of infections in China in one month (simulated data). (a) and (c) are the original infections data in day 1 and day 28. (b) and (d) are the spatiotemporal smoothing result in day 1 and day 28.

In order to understand the clustering results clearly, we further used `hcluster` package in R to cluster the provinces after smoothing. The principle is to compare each pair of time series. If the two data share similar trend (high correlation), they will be clustered into one category. After clustering, a tree was obtained, with the length of trunk indicating the degree of similarity. We used that package to calculate the infections distance between provinces. This can capture the province clusters which sharing the same trend characteristics. The result is showed in Fig. 5. The clustering result shows that the model successfully distinguished the infections trend in each province and the clustering result is consistent with our hypothesis.

IV. APPLICATION ON TWO SPECIFIC CASES

In the analysis of the real-world datasets, we take the COVID-19 infections outbreak datasets from January 24, 2020, to May 26, 2020, and the mortality rate datasets in China mainland over the last twenty years as two examples.

A. COVID-19 DATASET

The novel coronavirus (COVID-19) caused an epidemic of acute respiratory syndrome in China, which started from late December 2019 in Hubei province. Some researches indicated the virus has the potential for sustained human-to-human transmission [15] [16]. By January 20th, 2020, this epidemic had caused 198 laboratory confirmed infections with 3 death cases. Except Hubei province, Shanghai and Beijing reported 3 COVID-19 cases. On January 23rd, Chinese government adapted unprecedented action of locking down the Wuhan in Hubei to avert a wider epidemic, such as warning citizens not going outdoor, suspending part of public traffic in cities with severe outbreak. When the epidemic outbreak broke out in mainland of China, the government required each one to wear a medical mask in public area. People coming back from other provinces especially high-risk provinces like Hubei need to be quarantined at least for 14 days. These precautions may slow down the epidemic and give government more time to gather the medical resources. Because the virus has been exported to other countries, WHO identified the COVID-19 outbreak as a Public Health Emergency of International Concern (PHEIC) [17].

In this study, we used our model to smooth the COVID-19 data, which will be helpful for the follow-up epidemic research [18]. The analysis of early data of COVID-19 can help us to prevent the fast spread of this epidemic. After spatiotemporal smoothing, we can classify the areas according to the severity of COVID-19, which is conducive in controlling the spread of virus. Furthermore, the results
will provide a plan for the allocation of medical resources, and it also will remind people to avoid going to high-risk areas.

The data we used was retrieved from National Health Commission of the People’s Republic of China (http://www.nhc.gov.cn/), providing the latest epidemiological data of China from January 24, 2020, to May 26, 2020. We found that the infections in Hubei province are much higher than other province. The one of possible reasons is that COVID-19 is a brand-new epidemic. The data we collected started from January 24, 2020, but in fact, China’s first unknown pneumonia patient was found in Wuhan as early as December. This may have led to a large error in the data of Hubei Province. To analyze the transmission of novel coronavirus by our model more informative, the data outside Hubei province were used, and data in Hubei were omit. Since the dataset is much larger than simulated dataset, we used CVX algorithm to reduce the computational resources. After verification testing, we found the error between these two algorithms is exceedingly small on the same simulated data.

Because we have neglected the data of Hubei Province, our model was modified accordingly. The node number \( P \) was changed from 31 to 30. There are six provinces geographically adjacent to Hubei province, hence the edge number \( M \) was reduced to 63. The dataset spans 122 days and \( N = 122 \). \( D_S \) is of size \( 7686 \times 3660 \), \( D_T \) is \( 3630 \times 3660 \) and \( D \) is \( 11316 \times 3660 \). Because of the large amount of data, the running speed of automatic selection program is too slow and needs to be redesigned. After empirical testing, the \( \lambda \) was manually set to 40 and \( \mu \) to 0.8. The results show that the parameter combination can meet our expectations.

As shown in Fig. 6, we selected the original data and smoothed data of national epidemic data on February 2, 2020. Compared with the original data in Fig. 6(a), the smoothed data in Fig. 6(b) have obvious blocking phenomenon in the map, which are affected by geographical and time factors. We also selected the data from April 24, the 60th day after the outbreak, which is showed in Fig. 6(c). As can be seen from the Fig. 6(b), the map is divided into five levels according to the severity of the epidemic. First of all, there is no doubt that the severity of the epidemic in Hubei Province can be divided into the first level. Next, in the Fig. 6(a) and Fig. 6(b), the severity of the epidemic in Guangdong and Zhejiang is second only to Hubei so they can be divided into second level.

**FIGURE 6.** The geographical distribution of COVID-19 in China. (a) is the original infections data on Feb 2nd, (b) is the spatiotemporal smoothing result on Feb 2nd, (c) is the spatiotemporal smoothing result on April 24th.
Then, Anhui, Henan, Hunan, and Jiangxi provinces are divided into the third level after smoothing. The severity of the epidemic in Anhui and Jiangxi, neighboring Hubei Province, has been raised. What’s more, they are raised a level because their neighboring provinces like Zhejiang and Guangdong also are severe. After that, Chongqing, Fujian, Jiangsu, Shandong, and Shanghai are classified into forth level. Although Fujian’s situation is similar with southwest areas in original image, it’s adjacent with Guangdong and Zhejiang. So, Fujian is classified as the fourth level considering the influence of time trend and neighboring provinces. Chongqing is a little higher than these three provinces, although it’s near with Hubei province, geographically, it is inconvenient to transport because of the mountain range. According to the Fig. 6(a), the epidemic situation in Shandong, Jiangsu, and Shanghai are similar, slightly lower than that in other provinces in Southeast China, but higher than that in Northwest and Northeast China. In the other side, considering the time trend, Chongqing has pulled down to the fourth level. Finally, the other provinces in mainland are divided into fifth level. From the original data image, Sichuan and Tianjin have higher epidemic severity than other provinces in the fifth grade. However, due to the influence of geographical factors and time factors, these three provinces have been pulled down to the fifth level after smoothing the data.

Compared Fig. 6(a) with Fig. 6(b), the data smoothed by our model in some provinces has changed. These changes are obvious in the Fig. 6(b). According to the severity of the epidemic, the data of some provinces, such as Fujian, have been raised to a higher level, while those of some provinces, such as Sichuan, have been lowered. We think these changes are understandable. Some might be caused by geography reason and poor transport may have hindered the spread of COVID-19. When smoothing the data, the model not only considers the geographical factors among the provinces, but also the data changing during 122 days. This is why Anhui and Jiangxi province have similar geographical structures, original data and neighboring provinces’ situation, but they are divided into different levels. Compared Fig. 6(a) with Fig. 6(c), we can see with the time goes by, the classification of regions varies according to the outbreak data. Therefore, the data classification on a day will be affected.

Then, we used hcluster method to cluster the provinces. Hubei Province is still omitted in the classification. The result is showed in Fig. 7. After the province is classified by hcluster method, we draw the line charts for each category. The results are shown in Fig. 8.

According to the severity of the epidemic, we can still divide the provinces into five grades. The first level still is Hubei province. But interestingly, For the second level, the result of clustering is a little different from above single day’s analysis. Compared with the classification on February 2nd, the classification based on time analysis is closer to that on April 24th which is showed in Fig. 6(c). The second level includes Anhui, Guangdong, Henan, Hunan, Jiangxi, and Zhejiang province, which is showed in Fig. 8(a). One of the reasons is the Feb 2nd data are early on the COVID-19 outbreak. The provinces which adjacent with Hubei except Chongqing and Shaanxi are all divided into second level. From Fig. 8(a) we can discover that the second level’s affected people all near to a thousand.

The third level contains Beijing, Chongqing, Fujian, Heilongjiang, Jiangsu, Shandong, and Shanghai province. From Fig. 8(b), Beijing, and Heilongjiang within 12 days of a national outbreak are relatively lower than other third level’s provinces. Although the number of people infected in Chongqing and Fujian increased rapidly in the early stage, it was controlled at about 500 people later. As can be seen from the level 3 classification, after the overall time analysis, the provinces in the fourth level of the smoothing classification on February 2nd were upgraded by one level. This may be caused by the increasing number of infected people in Northwest and Southwest China.

In Fig. 8(c), no more than 360 people had been infected on the fourth level. In the fourth level, only more than 300 people were infected in Sichuan. Guangxi, Guizhou, Hainan, Hebei, Shaanxi, Shanxi, Tianjin, and Yunnan province are all in the fourth level. These provinces grew faster in early time than fifth level.

Gansu, Jilin, Liaoning, Neimongo, Ningxia, Qinghai, Xinjiang, and Xizang province are divided into fifth level. Because other provinces have more infections than them, the
most of them have less infections than the smoothing result. In Fig. 8(d), we can know the fifth level has the slowly increasing trend.

The time trend analysis clearly shows the trend of infections in each province. In the previous section, we used \texttt{hcluster} to classify and analyze the total smoothing results and get a good result. But observing the line charts, we can see that there are two different trends. Most provinces are increasing rapidly in early time and reaching a high level around 25 days after January 23rd, and then the number of infections begins to stabilize or slowly increase. Some provinces like Guangdong, Heilongjiang, Shanxi, Shanghai, Tianjin, these provinces showed an increasing trend after 60 days of the outbreak.

March 24, China Railway Wuhan Bureau Group Company issued a notice to resume the arrival and departure business in Hubei Province except for 17 railway passenger stations in Wuhan City from 0:00 on March 25 [19]. So we guess one of the reasons for this large-scale outbreak is the movement of people living with asymptomatic infections. Compared with the growth trend of the first outbreak, this time is relatively slow and the growth rate is not huge. 76 days after the closure of the city, the government lifted the ban on Wuhan, Hubei Province [20], which did not have a rebound effect on the epidemic. Around the beginning of May, the epidemic situation in the whole country was basically under control. On the 35th day or so, the data of some provinces suddenly decreased and increased after smoothing. On the 35th day or so, showed in Fig 8(b) and Fig 8(c), the data of some provinces suddenly decreased or increased and then stabilized again after smoothing. We analyze that this may be due to the geographic factors during smoothing. When we had less data, the analysis in time trend had a lot of bumps after smoothing. As our data volume increases, this protrusion or depression decreases considerably. On the other hand, the results may be over-fitting, which may be because the selection of parameters is not optimal enough.

B. MORTALITY RATE IN CHINA

The mortality is related to the variation of urbanization, geographic dispersion, disease, medical level and so on [21]. By using the proposed model to integrate time trend and geographical distribution information, regional imbalance can be discovered, and which can help administrators to give targeted policies for different provinces. Here we use our model to smooth the mortality of provinces in mainland China.

Because of China’s Hukou (household registration) system, the mortality data can be collected from China statistical yearbook and National Bureau of statistics of China [22]. Similar to the previous one, the node number $P$ and the edge number $M$ are 31 and 69 respectively. The dataset spans 20 years and $N = 20$. So $D_S$ is of size $1380 \times 620$. $D_T$ is of size $589 \times 620$. And $D$ is of size $1969 \times 620$. The parameters $\lambda$ and $\mu$ are manually selected as 0.8 and 0.2.

The results are shown in Fig. 9. The original mortality data of 1999 is showed in Fig. 9(a), and the smoothed result is shown in Fig. 9(b). Comparing Fig. 9(a) with Fig. 9(b), we
can find that after smoothing, the mortality rate is clustered into two regions. There is a gap between west and east region in the year 1999. The west region has a higher mortality rate than that of east. The highest mortality rate is in the northwest border area. The central part of the east has a higher mortality rate than that in the coastal areas. According to the comparison of 1999 (Fig. 9(b)) and 2017 (Fig. 9(c)), it is shown that the imbalance between east and west regions declined as time goes by, and some provinces like Gansu and Xinjiang even lower than others, and the mortality rate in mainland provinces tends to converge to a similar level.

After that, we used *hcluster* method to cluster the smoothed results. The result is shown in Fig. 10.

![Cluster Dendrogram](image)

**FIGURE 10.** The trend cluster result of mortality rate of provinces in China after spatial-temporal smoothing.

According to the clustering result, the time trend of provinces can be clustered into four patterns. As shown in Fig. 11. In Fig. 11(a), the mortality rate declines at beginning and then keeps steady, then continues to decline and finally rebounds. Ningxia and Xinjiang province belong to this pattern. The second pattern is showed in Fig. 11(b), the mortality rate tends to be stable after decreasing. A group of provinces suit this pattern like Anhui and Hebei. We can discover in Fig. 11(c), provinces like Guangdong and Hainan have the third pattern, which first declines and then tends to rise to its original level. The last pattern is in Fig. 11(d), provinces Xinjiang and Xizang are typical examples, and the general trend is downward. The mortality rate in the western provinces of China decreased significantly around 2000. We believe this result is consistent with the policy of western development at that time [23].

![Graph](image)

**FIGURE 9.** The geographical distribution of mortality rate in China over the past twenty years. (a) is the original mortality data in 1999, (b) and (c) are the spatiotemporal smoothing result in 1999 and 2017.
After smoothing, we were able to categorize the data into different trend, e.g., low-risk and high-risk areas. Note that the proposed model is not limited to three-dimensional data, such as the example of the epidemic or mortality data in this paper; it can be applied to other problems that can be modelled as time-varying graphs.

Compared with other research on gLASSO, our proposed model further expands the application scope of gLASSO. Previous researches tend to focus on one-dimensional or two-dimensional problems, but ignore the impact of temporal information on the problems. In our proposed model, we generalized the gLASSO from a purely spatial setting to a spatial-temporal one through a spatial-temporal constraint, which makes the gLASSO more widely used and solves the problem of gLASSO that lacks temporal information.

In the current model, edges in a graph share the same weight $\mu$, i.e., non-weighted graph. For a weighted graph, the extension of the proposed model is straightforward by adopting $\mu$’s in $D$ to edge weights correspondingly.

The selection of parameters is important and is still an open question. Akaike information criterion (AIC) [24] and Bayesian information criterion (BIC) [25] are widely used in weighting the complexity of model and the goodness of fitting data. We tried to use them as an automatic way to tune parameters, but the results are quite different from the parameters we selected manually, and the method is very computationally intensive. In further research, we tried to simplify the problem of parameter selection. To be specific, we took $\mu \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, indicating the ratio of the strength of the temporal penalty over that of the spatial penalty is in the range of 20% and 100% and tried to select the $\lambda$ based on the AIC or BIC. In general, $\mu$ increases with the number of time spots. In this way, we found that the selected parameters are not very different from those we selected manually. However, the parameters selected by this method are still not precise enough. We still hope to develop some automatic way to select parameters in this model in the future.

REFERENCES

[1] R. Tibshirani, “Regression shrinkage and selection via the lasso,” Journal of the Royal Statistical Society: Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996.

[2] B. E. Trevor, T. Hastie, L. Johnstone, and R. Tibshirani, “Least angle regression,” Annals of Statistics, vol. 32, pp. 407–499, 2002.

[3] P. Stoica and Y. Selen, “Model-order selection: a review of information criterion rules,” IEEE Signal Processing Magazine, vol. 19, pp. 21–28, July 2004.

[4] R. J. Tibshirani and J. Taylor, “The solution path of the generalized lasso,” Ann. Statist., vol. 39, pp. 1335–1371, 06 2011.

[5] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Physica D: Nonlinear Phenomena, vol. 60, pp. 259–268, 11 1992.

[6] James G M , Paulson C , Rusmevichientong P. “The Constrained Lasso,” 2013. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.306.8795 &rep=rep1&type=pdf.

[7] J. Duan, J. Idier, Y. Wang, and M. Wan, “A joint least squares and least absolute deviation model,” IEEE Signal Processing Letters, vol. 26, pp. 543–547, April 2019.
[8] Tibshirani R, Saunders M, Rosset S, et al. “Sparsity and smoothness via the fused lasso.” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67: 91-108. [Online]. Available: https://doi.org/10.1111/j.1467-9868.2005.00490.x.

[9] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, “11 trend filtering,” SIAM Review, vol. 51, no. 2, pp. 339–360, 2009.

[10] Xin B., Kawahara Y., Wang Y., et al. “Efficient generalized fused lasso and its application to the diagnosis of Alzheimer’s disease,” AAI Press, 2014. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/Pape r8261

[11] Wang H, Rodriguez A. “Identifying pediatric cancer clusters in Florida using log-linear models and generalized lasso penalties,” Statistics and Public Policy, 2014, 11(1): 86-96.

[12] Ohishi M, Fukui K, Okamura K, et al. “Coordinate optimization for generalized fused Lasso,” Communications in Statistics-Theory and Methods, 2021: 1-19.

[13] R. T. Taylor B. Arnold, “Introduction to the genlasso package,” [Online]. Available: https://pbil.univlyon1.fr/CRAN/web/packages/genlasso/vignettes/article.pdf.

[14] H. Wickham, “ggplot2,” Wiley Interdisciplinary Reviews: Computational Statistics, vol. 3, no. 2, pp. 180–185, 2011.

[15] Riu J, Althaus C L. “Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-ncov), December 2019 to January 2020,” Eurasurveillance, 2020, 25(4): 2000058.

[16] Li Q, Guan X, Wu P, et al. “Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia,” New England Journal of Medicine, vol. 382, 01, 2020.

[17] Centers for disease control and prevention, “Cdc first travel-related case of 2019 novel coronavirus detected in united states,” [Online]. Available: http://www.wuhan.gov.cn/zwgk/tzgg/20200404/20200414999326.shtml, 1 2020.

[18] L. Zhong, L. Mu, J. Li, J. Wang, Z. Yin, and D. Liu, “Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model,” IEEE Access, vol. 8, pp. 51761–51769, 2020.

[19] C. R. W. B. G. Company, “From the 28th, you can take the train to Wuhan,” [Online]. Available: http://www.wuhan.gov.cn/sy/whyw/202004/20200414998897.shtml.

[20] W. M. P. Government, “Hubei novel coronavirus infection prevention and control command announcement,” [Online]. Available: http://www.wuhan.gov.cn/zwgk/tzgg/20200404/20200414999326.shtml.

[21] Y. Chen, A. Ebenstein, M. Greenstone, and H. Li, “Evidence on the impact of sustained exposure to air pollution on life expectancy from china’s huai river policy,” Proceedings of the National Academy of Sciences, vol. 110, no. 32, pp. 12936–12941, 2013.

[22] N. B. of Statistics of China, “China statistical yearbook,” [Online]. Available: http://www.stats.gov.cn/tjjs/jsystj/.

[23] Yetian H X W B Z. “Carry out the strategy of western region development and promote the opening up within china,” Reform of Economic System, 2000: 05.

[24] H. Akaike, “A new look at the statistical model identification,” IEEE Transactions on Automatic Control, vol. 19, pp. 716–723, December 1974.

[25] K. P. Burnham and D. R. Anderson, “Multimodel inference: Understanding aic and bic in model selection,” Sociological Methods & Research, vol. 33, no. 2, pp. 261–304, 2004.