



# Simultaneous Denoising and Heterogeneity Learning for Time Series Data

Xiwen Jiang<sup>1</sup> · Weining Shen<sup>1</sup>

Received: 4 February 2023 / Revised: 23 July 2023 / Accepted: 28 July 2023  
© The Author(s) 2023

## Abstract

Noisy time series data are often collected in biomedical applications, and it remains an important task to understand the data heterogeneity. We propose an approach that combines the strength of trend filtering and distance-based clustering to simultaneously perform temporal mean denoising and subject-level clustering. We discuss an iterative algorithm that efficiently computes the cluster structure and clusterwise mean trends. Simulation studies confirm the excellent numerical performance of our method. We further consider two data application examples including an U.S. lung cancer mortality study and a suicide rate study.

**Keywords** Cancer mortality · Clustering · *K*-means · Sequential data · Trend filtering

## 1 Introduction

With the advancement of information technology, a significant amount of sequential data has become accessible in the field of biomedical research. This includes data from microarray and RNA-seq in genetic studies and patient health-tracking information in disease studies. These data are often presented as time series, longitudinal, and functional data, with repeated measurements taken over a period of time for each study subject. This type of data has been shown to be highly valuable in gaining insights into the underlying mechanisms of diseases, developing new diagnostic techniques and treatment plans, and ultimately enhancing patient healthcare [1, 2].

---

✉ Weining Shen  
weinings@uci.edu

Xiwen Jiang  
xiwenj@uci.edu

<sup>1</sup> Department of Statistics, University of California, Irvine, 2206 Donald Bren Hall, Irvine, CA 92617, USA

The main focus of this paper is to develop an approach that simultaneously performs denoising and clustering for sequential data. In particular, we focus on time series data and hope our method can be generalized to other data types in a future work. Data denoising and smoothing are a well-recognized challenge due to the existence of complex fluctuations and seasonal variations in time series, including short- and long-term changes. Several methods have been proposed in the literature to address this problem, including total variation denoising [3, 4], Gaussian process filter [5–7], wavelet transform [8, 9], Kalman filter [10, 11], and kernel smoothers [12–14]. We focus on trend filtering [15–17], which is a nonparametric smoothing method that fits a piecewise polynomial model to the data. Compared to other approaches, trend filtering achieves a desirable balance between easy model interpretation and theoretically guaranteed estimation accuracy [18]. Moreover, trend filtering can be implemented efficiently using alternating direction method of multipliers (ADMM) algorithm [19]. Thanks to these properties, trend filtering has been widely used in denoising time series data such as annual GDP [20] and global surface temperature deviation [21].

Despite its success in signal denoising and curve estimation, trend filtering does not directly handle heterogeneity, which is a critical issue to address in biomedical applications. For example, large heterogeneity is known to exist in patients' health information, genomic profile, and treatment effects. There is often a need to identify patient subgroups for improving the accuracy of disease diagnosis and personalized treatment [22]. In epidemiology, it is well known that heterogeneity exists in many diseases (e.g., malaria) across geographical regions and social networks [23, 24]. It is hence our goal in this paper to fill this gap. In particular, we propose a clusterwise trend filtering approach that simultaneously identifies the clustering structure in study subjects where each cluster has a different mean trend over time fitted by different trend filtering models. The result is expected to provide useful insights towards a better understanding of data heterogeneity than what a marginal homogeneous model can offer. For example, in a lung cancer mortality study (more details in Sect. 4.1), the annual mortality rate is monitored over 48 continental states in the US between the year 1969 and 2009. By studying how the mean trend changes over time for different states, we are able to reveal interesting spatial heterogeneity pattern and relate the spatial clusters to environmental factors.

Clustering has been extensively researched in the fields of statistics and machine learning, as indicated in a recent survey [25]. Our proposed method aims to integrate trend filtering with distance-based clustering approaches. We use  $K$ -means [26] as an example and show that our method offers the best of both worlds, inheriting the nice properties of both  $K$ -means and trend filtering in terms of easy implementation and computational efficiency. Through simulation studies, we also show that our method effectively recovers the unknown cluster structure and clusterwise trends. We further demonstrate the utility of our method using two real-world examples. The rest of this paper is organized as follows. Section 2 offers a brief review of trend filtering method and discusses our proposed method. In Sect. 3, we use simulations to compare our method with a few existing approaches. We present two data analysis examples in Sect. 4 and discuss a few future working directions in Sect. 5.

## 2 Methods

In this section, we give a brief review of trend filtering, and then present our method in Sect. 2.2.

### 2.1 Trend Filtering Estimation

Consider a time series with  $T$  time points  $\mathbf{y} = (y_1, \dots, y_T) \in \mathbb{R}^T$ . Trend filtering [15–17] provides a useful way of smoothing the data by considering a piecewise polynomial approximation. In particular, for a given non-negative integer  $q$ , the  $q$ th order trend filtering estimates  $\hat{\beta} \in \mathbb{R}^T$  by solving the following optimization problem:

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^T} \frac{1}{2} \|\mathbf{y} - \beta\|_2^2 + \lambda \|D^{(q+1)}\beta\|_1, \tag{1}$$

where  $\beta$  follows a  $q$ th-order piecewise polynomial,  $\lambda$  is a non-negative tuning parameter to control the trade-off between smoothness of  $\beta$  and approximation error  $\|\mathbf{y} - \beta\|_2$ , and  $D^{(q+1)}$  is the discrete difference operator of order  $q + 1$ . For example, when  $q = 0$ , the fitted values  $\beta = (\beta_1, \dots, \beta_T)$  form a piecewise constant structure, and

$$D^{(q+1)} = D^{(1)} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}_{(T-1) \times T},$$

which means  $\|D^{(1)}\beta\|_1 = \sum_{i=1}^{T-1} |\beta_i - \beta_{i+1}|$ , i.e., (1) yields one-dimensional fused lasso [27].

For  $q \geq 1$ , the operator  $D^{(q+1)} \in \mathbb{R}^{(T-q-1) \times T}$  is defined recursively by  $D^{(q+1)} = D^{(1)} \cdot D^{(q)}$ . For example, when  $q = 1$ ,  $\|D^{(2)}\beta\|_1 = \sum_{i=2}^{T-1} |\beta_{i-1} - 2\beta_i + \beta_{i+1}|$ , which is related to the Hodrick–Prescott filtering [28]. In general,  $\beta$  forms a piecewise linear structure when  $q = 1$  and a piecewise quadratic structure when  $q = 2$ , with

$$D^{(2)} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -2 & 1 \end{pmatrix}, D^{(3)} = \begin{pmatrix} -1 & 3 & -3 & 1 & \dots & 0 \\ 0 & -1 & 3 & -3 & \dots & 0 \\ 0 & 0 & -1 & 3 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -3 & 1 \end{pmatrix}.$$

As shown in Eq. (1), trend filtering estimation is a generalized lasso problem with an identity design matrix  $X = I$  and a specific choice of penalty  $D^{q+1}$ . Thus, it also shares properties of the generalized lasso, e.g., the degrees of freedom for trend filtering estimation are  $df(\hat{\beta}) = \mathbb{E}(\text{number of knots in } \hat{\beta}) + q + 1$  [29]. The number of knots in  $\hat{\beta}$  can be understood as the change points in the time series, which is also the number of non-zero entries in  $D^{(q+1)}\beta$  in the second term of Eq. (1). In addition, because (1) is strictly convex, the trend filtering estimate  $\hat{\beta}$  is the unique minimizer for every  $q \geq 0$ . In summary, trend filtering enjoys several nice properties, including

local adaptivity, computational efficiency, and easy interpretation [18], which makes it an ideal tool for our analysis.

### 2.2 Clusterwise Trend Filtering

Consider a dataset of  $n$  time series,  $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ , where each  $\mathbf{y}_i = (y_{i1}, \dots, y_{iT}) \in \mathbb{R}^T$  is a time series being observed over  $T$  time points. Our goal is to simultaneously perform smoothing for each  $\mathbf{y}_i$  and also cluster these time series. To achieve this goal, we consider a partition of the index set  $\{1, \dots, n\}$ , denoted by  $\mathcal{C} = \{C_1, C_2, \dots, C_K\}$  such that within each cluster, the time series are assumed to follow the same mean structure, which is modeled by a piecewise polynomial sequence obtained from trend filtering. In general, any distance-based clustering methods can be used for inferring the clustering structure  $\mathcal{C}$ . For simplicity, we choose  $K$ -means to demonstrate our idea. We propose to solve the following two optimizations:

$$\begin{aligned} \mathcal{C} &= \operatorname{argmin}_{C_1, \dots, C_K} \sum_{i=1}^n \sum_{k=1}^K \mathbb{1}\{c_i = k\} \|\mathbf{y}_i - \hat{\beta}_k\|_2^2, \\ \hat{\beta}_k &= \operatorname{argmin}_{\beta_k \in \mathbb{R}^T} \frac{1}{2} \|\bar{\mathbf{y}}_k - \beta_k\|_2^2 + \lambda \|D^{(q+1)} \beta_k\|_1, \quad k = 1, \dots, K, \end{aligned} \tag{2}$$

where  $\bar{\mathbf{y}}_k = |C_k|^{-1} \sum_{i=1}^{|C_k|} \mathbf{y}_i$  is the average of time series belonging to cluster  $k$ ,  $K$  is the pre-specified number of clusters, and  $c_i$  is the cluster index for  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ . It can be seen that the first optimization is similar with that of the original  $K$ -means, by treating  $\mathbf{y}_i$  as the input data point and  $\hat{\beta}_k$  as the center for cluster  $k$ , which is obtained by fitting trend filtering to the cluster average to help with interpretation. An alternative approach is to borrow the idea of  $K$ -medoids algorithm, which is proposed as a variant of  $K$ -means to address the influence of outliers [30]. Unlike  $K$ -means,  $K$ -medoids does not use the mean value, but instead finds a data point as the center of the cluster. Because  $K$ -medoids is computationally more expensive than  $K$ -means as it involves computing the distances between all pairs of data points at each iteration [31], we still choose to use  $K$ -means in our numerical implementation.

Optimization in (2) can be conveniently solved by the following procedure:

- (1) Initialization: Set the cluster number  $K$ , and generate an initial partition  $\mathcal{C} = (C_1, \dots, C_K)$  by fitting  $K$ -means to the dataset treating each data point as a  $T$ -dimensional vector.
- (2) Obtain the trend filtering estimator  $\hat{\beta}_k$  for each cluster  $C_k$ ,  $k = 1, \dots, K$ .
- (3) Update partition  $\mathcal{C}$  by assigning each time series to its closest center, i.e., the updated  $\hat{\beta}_k$  from step (2).
- (4) Repeat step (2) and (3) until convergence.

In practice, we stop the algorithm once the partition  $\mathcal{C}$  does not change after a few updates. Our proposed approach inherits the simplicity and computational convenience of  $K$ -means and trend filtering. In particular, both methods can be conveniently implemented in standard software packages such as R, so does our method.

The computational complexity of trend filtering is at most  $O(n^{3/2})$  [15, 17] and the complexity of  $K$ -means is  $O(n^2)$ . Hence our method has an  $O(n^2)$  computational complexity due to the prefixed  $K$ .

At the same time, our method also faces the same challenges as trend filtering and  $K$ -means do. For example, the objective function of  $K$ -means is non-convex, which means that it may converge to a local minimum instead of the global optimum, and the results may be sensitive to the choice of initial values. Therefore, multiple initial values will be used to fully explore the parameter space. Another challenge is the choice of hyperparameters including the cluster number  $K$  and the polynomial order  $q$ . In practice, we choose  $q$  either based on prior knowledge (e.g., shape of the trend) or let  $q$  take values within a range and pick the optimal value that minimizes the total sums of square error between raw clusterwise data average  $\mathbf{y}_k$  and the filtered trend  $\hat{\beta}_k$ . For the choice of cluster number  $K$ , several approaches are available in the literature to determine  $K$  for  $K$ -means, such as the elbow method [26], the Silhouette score [32], and cross-validation [33]. However, there are no universally agreed criteria to determine the optimal value of  $K$ , especially for the large-scale dataset with more overlapping or fuzzy clusters. In our data analysis, we consider a reasonably wide range of values for  $K$  that yields a convenient interpretation depending on the nature of scientific applications and the computational complexity.

Our method can be easily generalized to integrate with other distance-based clustering methods. For example, one may consider hierarchical clustering: start by trend filtering every times series to form separate clusters, then calculate the pairwise distance to merge the two closest clusters, and repeat this process till a proper number of clusters is obtained. Other distance-based clustering can also be adjusted based on the smooth version of individual time series.

### 3 Simulation

#### 3.1 Setting

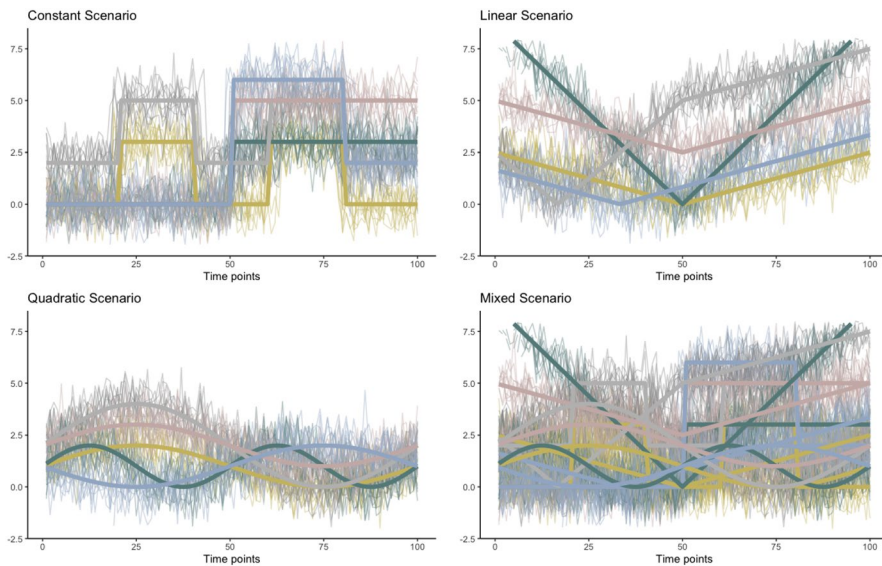
We conduct simulation studies to evaluate the empirical performance of our proposed approach. We generate data with a mean structure following a piecewise polynomial model under four settings, including a constant scenario, a linear scenario, a quadratic scenario, and a mixed scenario. For example, when the order is 0, time series from different clusters all present a piecewise constant trend, which contains several unknown phases (varying over clusters) and takes a constant value under each phase. Under the mixed scenario, time series in different clusters follow a piecewise polynomial with different orders. More specifically, under the first three scenarios, there are five different types of mean trends which correspond to five clusters. Under the mixed scenario, there are a total of 15 types of mean trends. For instance, for the constant scenario, the number of phases, the length of each phase, and the signal values may vary cluster by cluster. Under all scenarios, each cluster includes 10 time series observed over  $T = 100$  time points. We then add Gaussian noise to the generated mean trends with a standard deviation taking values from 0.4

to 1.8. A demonstration is given in Fig. 1, where the colored solid lines are the generated mean trends for each cluster. For example, there are 5 clusters and 50 time series in each of the first three panels under the piecewise constant, linear, and quadratic scenarios; and 15 clusters and 150 time series under the mixed scenario.

We compare our proposed method with two alternative approaches:  $K$ -medoids [34] and functional  $K$ -means clustering [35, 36]. Functional  $K$ -means provides a useful way to identify common patterns and trends among different groups of functional data. For implementation, all numerical experiments are conducted in R on a compute server (256 GB RAM, with 8 AMD Opteron 6276 processors, operating at 2.3 GHz, with 60 processing cores). The average running time is 9.3 (SD = 1.3) minutes for one simulated dataset analysis. Our method can be implemented based on *genlasso* package [37] for trend filtering step. The  $K$ -medoids is implemented using *clust* package. In the simulation, we assume the order of polynomials  $q$  is known for the first three scenarios. For the mixed scenario, we assume  $q$  is unknown. To determine its value, we let  $q$  take values between 0 and 3, and pick the one that minimizes the fitted square error.

### 3.2 Results

We conduct the simulation for 1000 replications and summarize the percentage of times when the true cluster structures are correctly identified in Table 1. In addition, we calculate the Rand index (RI), which is a metric that can be used to evaluate the performance of a clustering algorithm taking values between 0 and 1 (higher is better) [38]. The RIs for our method and two competing approaches



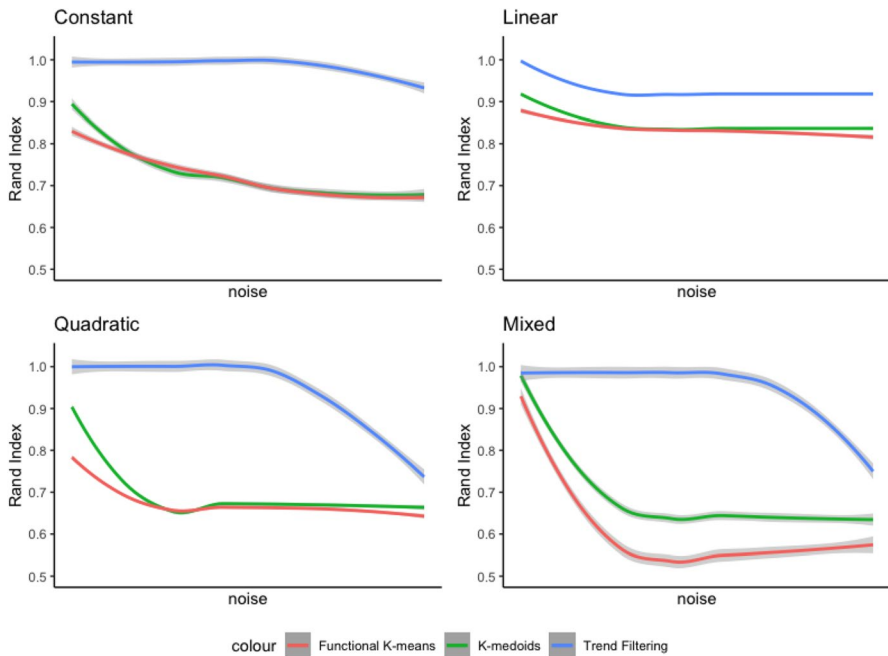
**Fig. 1** Four simulation data generation scenarios (noise SD = 1.2) (Color figure online)

**Table 1** Simulation results: percentage of correct cluster structure identification (and associated maximum standard errors) for our method (Trend Filtering), functional *K*-means (Functional), and *K*-medoids, based on 1000 replications

Noise (max SE)	Constant			Linear		
	Trend filtering	Functional	<i>K</i> -medoids	Trend filtering	Functional	<i>K</i> -medoids
	(0.054)	(0.067)	(0.087)	(0.029)	(0.030)	(0.031)
0.4	0.9922	0.8410	0.9170	0.9997	0.8850	0.9184
0.6	0.9940	0.7604	0.7553	0.9303	0.8512	0.8635
0.8	0.9953	0.7548	0.7476	0.9184	0.8365	0.8367
1.0	0.9956	0.7201	0.7153	0.9184	0.8356	0.8367
1.2	0.9939	0.6906	0.6913	0.9184	0.8333	0.8367
1.4	0.9839	0.6742	0.6798	0.9184	0.8295	0.8367
1.6	0.9638	0.6678	0.6770	0.9184	0.8247	0.8367
1.8	0.9245	0.6629	0.6761	0.9184	0.8199	0.8367
Noise (max SE)	Quadratic			Mixed		
	Trend filtering	Functional	<i>K</i> -medoids	Trend filtering	Functional	<i>K</i> -medoids
	(0.116)	(0.048)	(0.084)	(0.107)	(0.147)	(0.127)
0.4	1.0000	0.7880	0.9176	0.9843	0.9252	0.9772
0.6	1.0000	0.6682	0.6735	0.9847	0.7081	0.7772
0.8	1.0000	0.6650	0.6733	0.9860	0.5364	0.6470
1.0	0.9983	0.6629	0.6727	0.9848	0.5402	0.6429
1.2	0.9855	0.6607	0.6708	0.9751	0.5447	0.6392
1.4	0.9208	0.6571	0.6684	0.9531	0.5487	0.6354
1.6	0.8174	0.6512	0.6661	0.8751	0.5532	0.6329
1.8	0.7351	0.6406	0.6636	0.7323	0.5565	0.6308

are summarized in Fig. 2, where blue lines are for our method, and green and red lines are for functional *K*-means and *K*-medoids.

We find that our method achieves the highest accuracy in all scenarios and under different noise levels. The advantage over the competing methods becomes more significant as the noise level increases (e.g., noise  $\geq .8$ ), which indicates that our method works quite well especially for more fluctuating curves. As the order of polynomial becomes larger, the clustering accuracy deteriorates for all methods as expected. Also it is worthy mentioning that for the most difficult case, mixed scenario, our method manages to achieve a high RI value of above 90% even when the noise level is fairly large ( $\geq 1.4$ ). For the piecewise constant case, our method manages to maintain a RI above 90% over all noise levels while the other two methods have an RI dropping down to around 65%. All these observations confirm the excellent performance of our method.



**Fig. 2** Simulation accuracy: Rand index and associated 95% confidence bands for three methods under different data generating scenarios and noise levels (Color figure online)

## 4 Real-Data Examples

### 4.1 Lung Cancer Mortality Rate

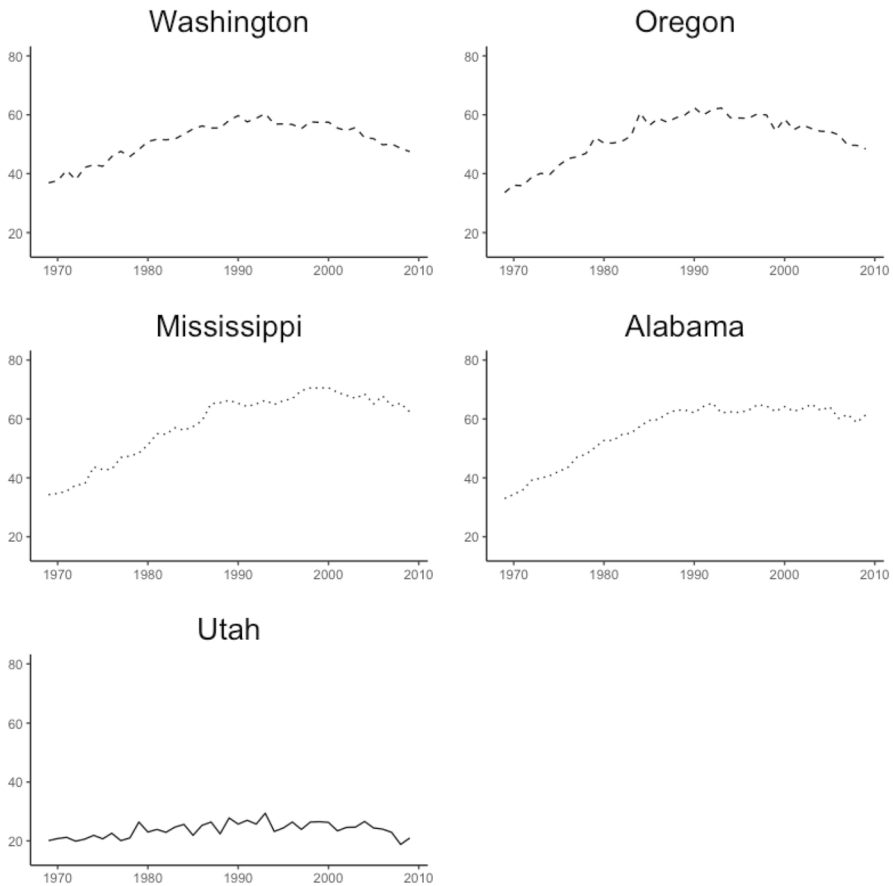
Cancer is a leading cause of death in the United States. Among all types of cancer, the bronchial and lung cancers are associated with the highest number of deaths. In 2019, it is estimated that 0.6 million people died of cancer in the United States, with 0.14 million due to lung cancer. Past studies supported by National Institutes of Health have suggested the existence of geographical pattern in bronchial and lung cancers, e.g., the highest incidence was found in the south (76.0 per 100,000) and the lowest incidence was in the west (58.8 per 100,000) [39]. In addition, many studies have discovered a temporal change pattern in lung cancer mortality [40–44].

We analyze the lung cancer mortality rate data collected by the American Cancer Society, which covers the annual age-adjusted death rate due to lung cancer in 48 states in the US (excluding Alaska and Hawaii) from 1969 to 2009. In other words, the data consist of 48 time series being observed over 41 years. As shown in Fig. 4, the temporal trend for most states has a parabolic trajectory. The mortality rate continuously increases for the first two decades until the peak around 1990 followed by a decrease over the next two decades, with some states beginning to stabilize at the same level and others experiencing significant declines in mortality. Our goal is to explore the heterogeneity in the state-level mortality rate curves. This is important because the resulting clusters provide insights into which factors may influence the

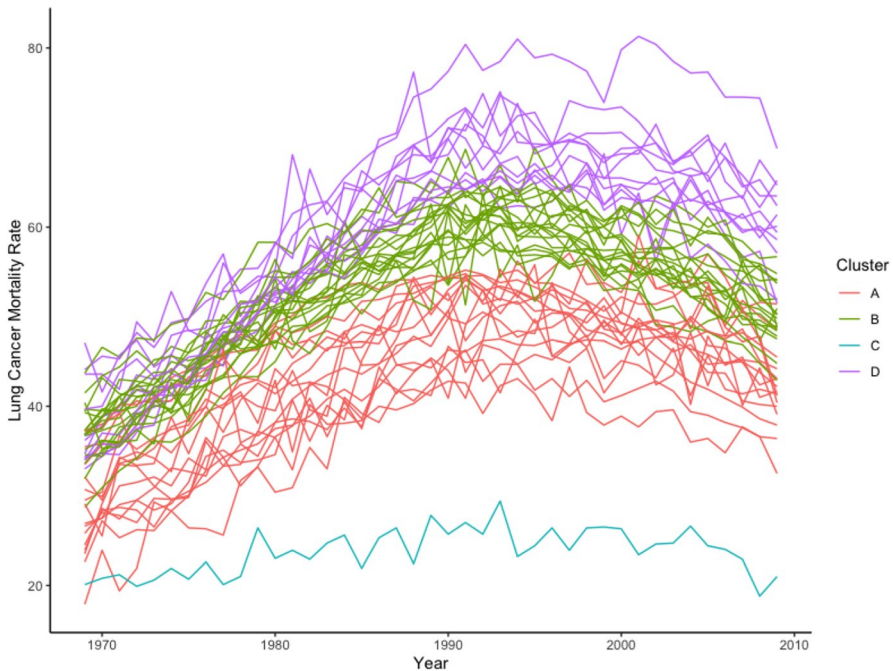


mortality rates. For example, a spatial pattern can be seen in the mortality rate map. Some neighboring states, such as Washington and Oregon, have similar changes in their mortality rates, with a sharp decline after reaching the peak of the parabola. In comparison, some neighboring states in the southeast, such as Mississippi and Alabama, have mortality rates stabilized after a previous upward trend. In addition, Utah's pattern is distinctly different from its nearby states, as shown in Figs. 3 and 4.

We apply our proposed method and set  $q = 2$  to capture the parabolic trajectory in the curves. To determine the cluster number  $K$ , we fit the model by letting  $K$  take values from 3 to 8, and consider several criteria, including the elbow method, the silhouette coefficient [32], the Calinski–Harabasz index [45], and the Gap statistics [46]. We choose  $K = 4$  since it is preferred by most of the criteria being considered. The clustering result is shown in Fig. 5, where four clusters are marked by four colors. There seems a quite obvious geographic pattern in the result. For example, cluster A consists of spatially contiguous states



**Fig. 3** Age-adjusted mortality rates of lung cancer from 1969 to 2009 for five selected states



**Fig. 4** Spaghetti plot of lung cancer mortality rates for 48 continent states in the U.S. (Color figure online)

located in the Rocky Mountains and the Mid-Atlantic region, and cluster D is mainly located in the middle-east and southern part of US except Nevada and Maine. Utah forms its own cluster due to its low mortality rate compared to the rest of U.S. Our cluster result is also presented in Fig. 4, where it is obvious that cluster A has the lowest mortality rate (excluding Utah), while cluster D has the highest mortality rate and the fastest growth during the year 1970–1990. Cluster D also has a higher variation compared to the other clusters.

We are also able to relate our cluster result with two main risk factors of lung cancer, including smoking, which is the number one risk factor [47–49], and air pollution, which also contributes to lung cancer [50–52]. As a reference, we present the state-level plots for both risk factors in Fig. 6a and b. It is clear that Utah has the lowest adult smoking rate in the country at 9%. While most states in cluster D have a higher smoking rate [darker blue color in panel (a)], e.g., Arkansas at 22% and Kentucky at 23%. Similar findings can be obtained in Fig. 6b. For example, cluster A and C in general have a lower air pollution rate compared to the other regions, which matches with the fact that these two clusters have a lower mortality rate. Meanwhile, cluster D has the highest air pollution index and, hence, the highest mortality rate. These findings highlight the

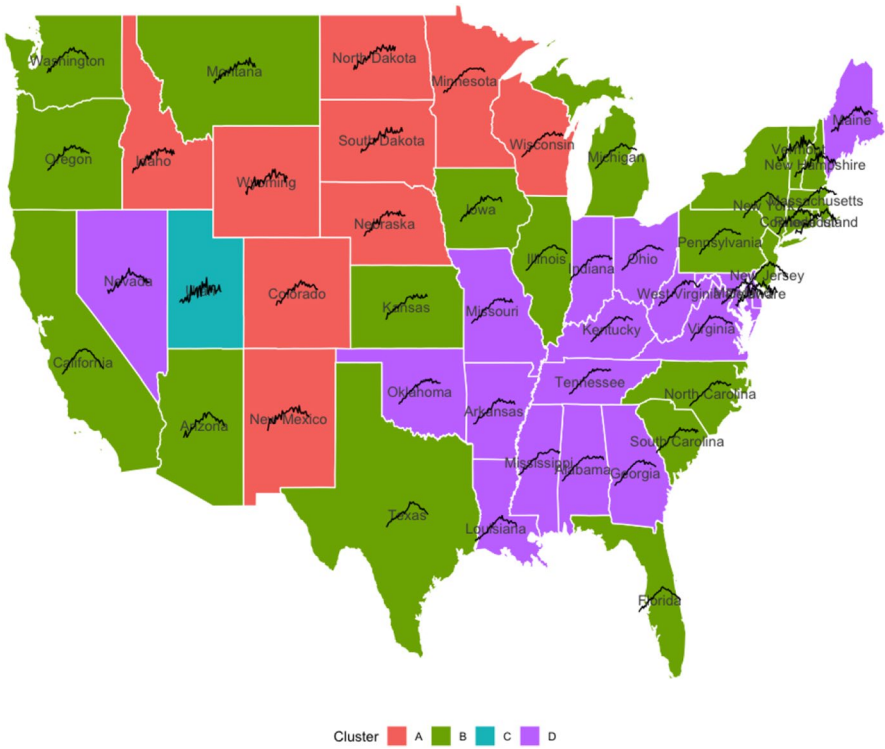


Fig. 5 U.S. map with four clusters obtained by our proposed method (Color figure online)

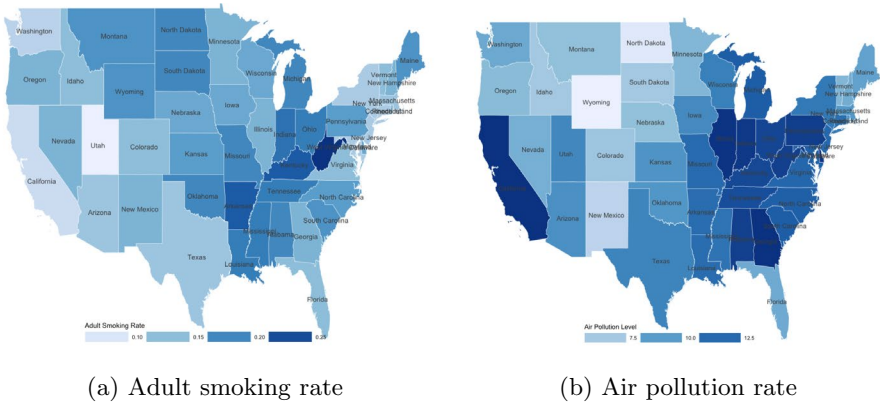


Fig. 6 U.S. map based on two leading risk factors associated with the lung cancer (Color figure online)

utility of our method in discovering meaningful clustering and temporal patterns in mortality rate curves.

### 4.2 Suicide Rate Study

Next we consider a suicide mortality study. According to World Health Organization (WHO), more than 0.7 million people die due to suicide every year. This number has kept increasing since COVID-19 [53–55]. Many factors contribute to the risk of suicide, including mental illness, stigma, financial reasons, alcohol, and drug misuse [56]. In recent years, researchers have also discovered temporal and spatial patterns in the suicide rates [57–59]. To verify the effectiveness of our method, we study a 30-year-long data on suicide mortality in the U.S. The data are available as a CDC Wide-ranging Online Data for Epidemiologic Research (WONDER) dataset. It provides the annual suicide mortality rates for all 48 contiguous states in the continental United States (excluding Alaska and Hawaii) from 1990 to 2019. As shown in Fig. 7a, the suicide rate exhibits a ‘V’ shape for most states, i.e., there are two phases over the observed 30-year period. During the first phase (first 10–15 years), the suicide mortality rate keeps going down. This trend is especially obvious for states such as California, Nevada, Illinois, and New York. The next 15–20 years is the second phase for a strong rebound, where the mortality rate in many states has far exceeded the initial 1990 level by the end of 2010.

We apply our method and choose  $q = 2$ , i.e., a piecewise quadratic trend. The cluster number  $K$  is decided to be 3 according to a combination of elbow method, the silhouette coefficient, and Gap statistic. The clustering results are provided in Fig. 7a and the clusterwise average curves are given in Fig. 7b. The results exhibit a clear geographical pattern despite we did not include any spatial information in our analysis. For example, cluster II (green) consists of 11 contiguous states in the middle west part of U.S; the cluster III (blue) has 29 states where the majority are states in east and middle east except Washington; and cluster I is the smallest cluster that contains California, Illinois, New Jersey, and Massachusetts. As shown in

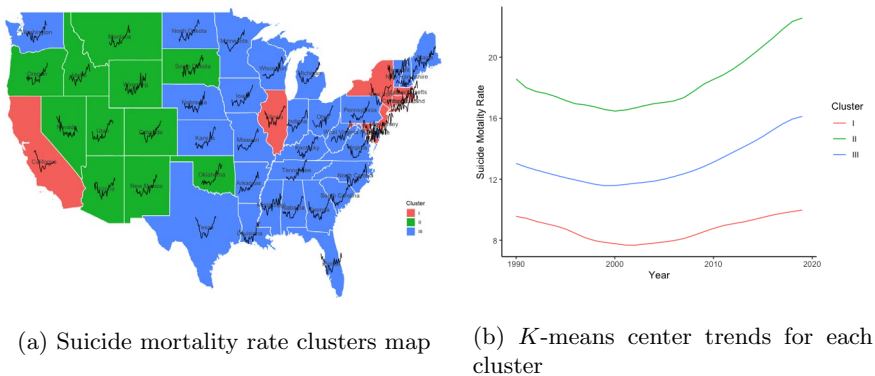


Fig. 7 Suicide mortality rate clusters (Color figure online)

Fig. 7b, the suicide mortality rate is the highest in cluster II, followed by cluster III and I. One possible explanation is that a high suicide rate is often associated with a low economic status. For example, WHO reports that 77% of global suicides occur in low- and middle-income countries. This is reflected in our results, e.g., cluster I, despite having the least number of states, has the best economic and welfare development and hence the lowest suicide rate.

## 5 Discussion

In this paper, we propose a new time series clustering method that performs smoothing over temporal direction and learns heterogeneity at subject level. Our method builds on the idea of  $K$ -means clustering and trend filtering, and can be extended to integrate with other distance-based clustering methods. Numerical results have confirmed the utility of our method in terms of cluster structure recovery and time series denoising. Our data analysis results suggest that the cluster results can be useful to provide guidance on the inclusion of covaraites (e.g., spatial, environmental, and economic factors) in a future analysis such as regression.

Several future work directions remain open for this topic. First, it will be of interest to generalize our method to analyze longitudinal and functional data where the observations are collected at non-equally spaced time points. Classical trend filtering cannot perform smoothing over irregular time intervals. Instead, one may consider other smoothing methods such as wavelet or kernel approaches. Second, it will be of interest to develop Bayesian methods that could take account for the uncertainty associated with cluster number and polynomial order estimation by using Gaussian process and its generalizations [60, 61]. In addition, studying theoretical properties such as convergence analysis of the algorithm and risk analysis of the curve estimation in this context is another important direction. Finally, developing a spatial clustering method that accounts for the spatial dependence may help improve the performance of our method in our data examples.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Qi Z, Liu D, Fu H, Liu Y (2020) Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *J Am Stat Assoc* 115(530):678–691

2. Mo W, Qi Z, Liu Y (2021) Learning optimal distributionally robust individualized treatment rules. *J Am Stat Assoc* 116(534):659–674
3. Vogel CR, Oman ME (1996) Iterative methods for total variation denoising. *SIAM J Sci Comput* 17(1):227–238
4. Condat L (2013) A direct algorithm for 1-D total variation denoising. *IEEE Signal Process Lett* 20(11):1054–1057
5. Ko J, Fox D (2009) GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models. *Auton Robot* 27(1):75–90
6. Quinonero-Candela J, Rasmussen CE (2005) A unifying view of sparse approximate Gaussian process regression. *J Mach Learn Res* 6:1939–1959
7. Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S (2013) Gaussian processes for time-series modelling. *Philos Trans R Soc A Math Phys Eng Sci* 371(1984):20110550
8. Pan Q, Zhang L, Dai G, Zhang H (1999) Two denoising methods by wavelet transform. *IEEE Trans Signal Process* 47(12):3401–3406
9. Fligge M, Solanki S (1997) Noise reduction in astronomical spectra using wavelet packets. *Astron Astrophys Suppl Ser* 124(3):579–587
10. Kalman R (1960) A new approach to linear filtering and prediction problems. *ASME J Basic Eng* 82:35–45
11. Meinhold RJ, Singpurwalla ND (1983) Understanding the Kalman filter. *Am Stat* 37(2):123–127
12. Wand MP, Jones MC (1994) Kernel smoothing. CRC Press, Boca Raton
13. Hall P, Huang L-S (2001) Nonparametric kernel regression subject to monotonicity constraints. *Ann Stat* 29(3):624–647
14. Kong D, Bondell H, Shen W (2018) Outlier detection and robust estimation in nonparametric regression. In: International conference on artificial intelligence and statistics. PMLR, pp 208–216
15. Kim S-J, Koh K, Boyd S, Gorinevsky D (2009)  $\ell_1$  trend filtering. *SIAM Rev* 51(2):339–360
16. Steidl G, Didas S, Neumann J (2006) Splines in higher order TV regularization. *Int J Comput Vis* 70(3):241–255
17. Tibshirani RJ (2014) Adaptive piecewise polynomial estimation via trend filtering. *Ann Stat* 42(1):285–323
18. Wang Y-X, Sharpnack J, Smola A, Tibshirani R (2015) Trend filtering on graphs. In: Artificial intelligence and statistics. PMLR, pp 1042–1050
19. Ramdas A, Tibshirani RJ (2016) Fast and flexible ADMM algorithms for trend filtering. *J Comput Graph Stat* 25(3):839–858
20. Yamada H, Jin L (2013) Japan’s output gap estimation and  $\ell_1$  trend filtering. *Empir Econ* 45(1):81–88
21. Roualdes EA (2015) Bayesian trend filtering. arXiv Preprint. <http://arxiv.org/abs/1505.07710>
22. Gao X, Shen W, Ning J, Feng Z, Hu J (2022) Addressing patient heterogeneity in disease predictive model development. *Biometrics* 78(3):1045–1055
23. Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Sabot O, Rodriguez MH, Abeyasinghe RR, Ghebreyesus TA et al (2010) Shrinking the malaria map: progress and prospects. *The Lancet* 376(9752):1566–1578
24. Yin F, Shen W, Butts CT (2022) Finite mixtures of ERGMS for modeling ensembles of networks. *Bayesian Anal* 17(4):1153–1191
25. Rai P, Singh S (2010) A survey of clustering techniques. *Int J Comput Appl* 7(12):1–5
26. MacQueen J (1967) Classification and analysis of multivariate observations. In: 5th Berkeley symposium on mathematical statistics and probability. pp 281–297
27. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005) Sparsity and smoothness via the fused lasso. *J R Stat Soc Ser B (Stat Methodol)* 67(1):91–108
28. Hodrick RJ, Prescott EC (1997) Postwar us business cycles: an empirical investigation. *J Money Credit Bank*. <https://doi.org/10.2307/2953682>
29. Tibshirani RJ, Taylor J (2011) The solution path of the generalized lasso. *Ann Stat* 39(3):1335–1371
30. Park H-S, Jun C-H (2009) A simple and fast algorithm for k-medoids clustering. *Expert Syst Appl* 36(2):3336–3341
31. Arora P, Varshney S et al (2016) Analysis of k-means and k-medoids algorithm for big data. *Procedia Comput Sci* 78:507–512
32. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65
33. Smyth P (1996) Clustering using Monte Carlo cross-validation. In: KDD, vol 1. pp 26–133

34. Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis. Wiley, Hoboken
35. García MLL, García-Ródenas R, Gómez AG (2015) K-means algorithms for functional data. *Neurocomputing* 151:231–245
36. Meng Y, Liang J, Cao F, He Y (2018) A new distance with derivative information for functional k-means clustering algorithm. *Inf Sci* 463:166–185
37. Tibshirani RJ, Arnold TB (2020) Introduction to the genlasso package. Citeseer
38. Rand WM (1971) Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 66(336):846–850
39. Schabath MB, Cress WD, Muñoz-Antonia T (2016) Racial and ethnic differences in the epidemiology and genomics of lung cancer. *Cancer Control* 23(4):338–346
40. Gadgeel SM, Severson RK, Kau Y, Graff J, Weiss LK, Kalemkerian GP (2001) Impact of race in lung cancer: analysis of temporal trends from a surveillance, epidemiology, and end results database. *Chest* 120(1):55–63
41. Zhang Y, Luo G, Etxeberria J, Hao Y (2021) Global patterns and trends in lung cancer incidence: a population-based study. *J Thorac Oncol* 16(6):933–944
42. Yang X, Man J, Chen H, Zhang T, Yin X, He Q, Lu M (2021) Temporal trends of the lung cancer mortality attributable to smoking from 1990 to 2017: a global, regional and national analysis. *Lung Cancer* 152:49–57
43. Wang N, Mengersen K, Kimlin M, Zhou M, Tong S, Fang L, Wang B, Hu W (2018) Lung cancer and particulate pollution: a critical review of spatial and temporal analysis evidence. *Environ Res* 164:585–596
44. Schabath MB, Thompson ZJ, Gray JE (2014) Temporal trends in demographics and overall survival of non-small-cell lung cancer patients at Moffitt Cancer Center from 1986 to 2008. *Cancer Control* 21(1):51–56
45. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat Theory Methods* 3(1):1–27
46. Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Ser B (Stat Methodol)* 63(2):411–423
47. Trichopoulos D, Kalandidi A, Sparros L, Macmahon B (1981) Lung cancer and passive smoking. *Int J Cancer* 27(1):1–4
48. Correa P, Fontham E, Pickle LW, Lin Y, Haenszel W (1983) Passive smoking and lung cancer. *The Lancet* 322(8350):595–597
49. Sun S, Schiller JH, Gazdar AF (2007) Lung cancer in never smokers—a different disease. *Nat Rev Cancer* 7(10):778–790
50. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, Hoffmann B, Fischer P, Nieuwenhuijsen MJ, Brunekreef B et al (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European study of cohorts for air pollution effects (escape). *Lancet Oncol* 14(9):813–822
51. Cohen AJ (2000) Outdoor air pollution and lung cancer. *Environ Health Perspect* 108(suppl 4):743–750
52. Fajersztajn L, Veras M, Barrozo LV, Saldiva P (2013) Air pollution: a potentially modifiable risk factor for lung cancer. *Nat Rev Cancer* 13(9):674–678
53. Tandon R (2021) COVID-19 and suicide: just the facts. Key learnings and guidance for action. *Asian J Psychiatry* 60:102695
54. Mamun MA (2021) Suicide and suicidal behaviors in the context of COVID-19 pandemic in Bangladesh: a systematic review. *Psychol Res Behav Manag* 14:695
55. John A, Pirkis J, Gunnell D, Appleby L, Morrissey J (2020) Trends in suicide during the covid-19 pandemic. *BMJ* 371. <https://doi.org/10.1136/bmj.m4352>
56. Pirkis J, John A, Shin S, DelPozo-Banos M, Arya V, Analuisa-Aguilar P, Appleby L, Arensman E, Bantjes J, Baran A et al (2021) Suicide trends in the early months of the COVID-19 pandemic: an interrupted time-series analysis of preliminary data from 21 countries. *The Lancet Psychiatry* 8(7):579–588
57. Sy KTL, Shaman J, Kandula S, Pei S, Gould M, Keyes KM (2019) Spatiotemporal clustering of suicides in the us from 1999 to 2016: a spatial epidemiological approach. *Soc Psychiatry Psychiatr Epidemiol* 54(12):1471–1482
58. Gould MS, Wallenstein S, Kleinman M (1990) Time-space clustering of teenage suicide. *Am J Epidemiol* 131(1):71–78

59. Hempstead K (2006) The geography of self-injury: spatial patterns in attempted and completed suicide. *Soc Sci Med* 62(12):3186–3196
60. Jones A, Townes FW, Li D, Engelhardt BE (2023) Alignment of spatial genomics data using deep Gaussian processes. *Nat Methods*. <https://doi.org/10.1038/s41592-023-01972-2>
61. Li D, Jones A, Banerjee S, Engelhardt BE (2021) Multi-group Gaussian processes. arXiv Preprint. <http://arxiv.org/abs/2110.08411>