RESEARCH

The Mann-Kendall-Sneyers test to identify the change points of COVID-19 time series in the United States

Xiang Chen[1,2*†], Hui Wang[3†], Weixuan Lyu[1] and Ran Xu[2,4]

Abstract

Background: One critical variable in the time series analysis is the change point, which is the point where an abrupt change occurs in chronologically ordered observations. Existing parametric models for change point detection, such as the linear regression model and the Bayesian model, require that observations are normally distributed and that the trend line cannot have extreme variability. To overcome the limitations of the parametric model, we apply a nonparametric method, the Mann-Kendall-Sneyers (MKS) test, to change point detection for the state-level COVID-19 case time series data of the United States in the early outbreak of the pandemic.

Methods: The MKS test is implemented for change point detection. The forward sequence and the backward sequence are calculated based on the new weekly cases between March 22, 2020 and January 31, 2021 for each of the 50 states. Points of intersection between the two sequences falling within the 95% confidence intervals are identified as the change points. The results are compared with two other change point detection methods, the pruned exact linear time (PELT) method and the regression-based method. Also, an open-access tool by Microsoft Excel is developed to facilitate the model implementation.

Results: By applying the MKS test to COVID-19 cases in the United States, we have identified that 30 states (60.0%) have at least one change point within the 95% confidence intervals. Of these states, 26 states have one change point, 4 states (i.e., LA, OH, VA, and WA) have two change points, and one state (GA) has three change points. Additionally, most downward changes appear in the Northeastern states (e.g., CT, MA, NJ, NY) at the first development stage (March 23 through May 31, 2020); most upward changes appear in the Western states (e.g., AZ, CA, CO, NM, WA, WY) and the Midwestern states (e.g., IL, IN, MI, MN, OH, WI) at the third development stage (November 19, 2020 through January 31, 2021).

Conclusions: This study is among the first to explore the potential of the MKS test applied for change point detection of COVID-19 cases. The MKS test is characterized by several advantages, including high computational efficiency, easy implementation, the ability to identify the change of direction, and no assumption for data distribution. However, due to its conservative nature in change point detection and moderate agreement with other methods, we recommend using the MKS test primarily for initial pattern identification and data pruning, especially in large data. With modification, the method can be further applied to other health data, such as injuries, disabilities, and mortalities.

†Xiang Chen and Hui Wang contributed equally to this work.

*Correspondence: xiang.chen@uconn.edu

[2] Institute for Collaboration on Health, Intervention, and Policy (InCHIP),
University of Connecticut, Storrs, CT 06269, USA
Full list of author information is available at the end of the article

Background

The Coronavirus Disease 2019 (COVID-19) pandemic has disrupted every aspect of human society. Because of the highly infectious nature of the disease, state governments in the United States (US) have implemented social distancing measures (e.g., closure of non-essential businesses, regional lock-down, and face-covering mandates) to contain the virus spread and flatten the epidemic curve (epi curve) [1]. However, since these state-level measures have differed in the strength and timeline of policy enforcement, it is intractable to rely on a simple rubric to evaluate the policy effectiveness. An alternative step is via analyzing the time series of the COVID-19 cases, which can eventually assist stakeholders with proactive health policymaking, such as determining the optimal timing to relieve social distancing.

One critical variable in the time series analysis is the change point, also called the inflection point, which is the point where a sudden change occurs in chronologically ordered observations. The change point detection has been long employed in statistical theory [2], but its applications to COVID-19 are relatively underexplored. For example, when modeling COVID-19 cases, the majority of studies have defined change points as key dates of policy interventions or social events [1, 3]. Other studies have employed parametric models, such as the linear regression model [4, 5] and the Bayesian model [6, 7] to derive change points. However, most of these parametric models require that the observations are normally distributed and that the trend line cannot have extreme variability. In situations where the observations show large variability over time and the trend line cannot be well fitted, parametric models become less reliable. These situations are not uncommon in fitting the COVID-19 epi curve, as the disease progression has a considerable degree of uncertainties and variability [1].

To overcome the limitations of the parametric model, we have applied a nonparametric model, called the Mann-Kendall-Sneyers (MKS) test, to change point detection in the COVID-19 epi curve. The MKS test, developed from a prototype model by Mann [8], is used to detect the monotonic trends (e.g., upward, downward) and their corresponding change points in time series data. The model has been primarily employed in earth science research to characterize the fluctuation of climatic and environmental variables, such as rainfall, air temperature, and surface runoff [9–11].

Recently, some COVID-19 studies have used the Mann-Kendall (MK) test, which is an earlier version of the MKS test, for trend detection [12, 13]. While the MK test is useful in detecting monotonic trends, it cannot detect changes in the trends and the corresponding change points, making it less useful for disease tracking and monitoring in the mid to long term. The MKS test, as a sequential extension of the MK test [14], fills this gap. It can become a valuable tool for long-term disease monitoring and can thus support public health decision-making.

The contributions of the paper are as follows.

- The paper is the first to apply the MKS test to COVID-19 time series analysis.
- The paper identifies six change point patterns for state COVID-19 cases.
- The paper develops an open-access tool for model implementation.

Methods

The nonparametric MKS test [15], oftentimes called the sequential Mann-Kendall-Sneyers test, has been applied to the change point detection for long-term time series data (e.g., hydrological changes, climatic changes). According to the Centers for Disease Control and Prevention (CDC) report, both social distancing and mass gathering can potentially lead to an abrupt change in regional COVID-19 cases, albeit in different directions [16]. Then, we have evaluated the potential of the MKS test for change point detection in short-term time series data, the COVID-19 cases of infection.

In this section, we first articulate the MKS test. Then, we use an example to demonstrate the model implementation.

Method description

The MKS test applied to the COVID-19 time series data can be completed in three major steps.

Step 1: Deriving test statistics (Sk)

We have treated new weekly cases as an independent observation in a 45-week time series data. Under the null hypothesis that the development of new cases remains stable, for each state, we have a time series of the weekly new cases: $X = \{x_1, x_2, x_3 \ldots x_N\}$, where n is the total number of weeks under observation ($N = 45$ in our case

Chen *et al. BMC Medical Research Methodology*   (2022) 22:233

Page 3 of 9

study). m_i ($i = 1, 2, ..., N$) represents the total number of elements x_j preceding x_i ($j < i$) where $x_j < x_i$.

Based on m_i, the test statistic S_k derives the cumulative m_i for each week, as shown in Eq. (1).

$$S_k = \sum_{i=1}^{k} m_i \ (k = 1, 2, 3, \ldots, N) \tag{1}$$

The mean of S_k can be derived by Eq. (2).

$$E(S_k) = k(k-1)/4 \tag{2}$$

The variance of S_k can be derived by Eq. (3).

$$VAR(S_k) = k(k-1)(2k-5)/72 \tag{3}$$

Step 2: Deriving two sequences (Uf and Ub)

Next, we derive two sequences, the forward sequence U_f and the backward sequence U_b, based on the three variables (S_k, $E(S_k)$, and $VAR(S_k)$) in Eqs. (1) through (3). Specifically, the forward sequence U_f of the time series is derived by Equation [4].

$$U_f = (S_k - E(S_k))/\sqrt{VAR(S_k)} \tag{4}$$

Then, we reverse the sequence of the original time series X and term it X_r. An intermediate sequence U_{fr} is derived by applying Eq. (4) to the reversed time series X_r. We reverse the sequence of the values in U_{fr} (i.e., the first value appears the last, and vice versa). We generate the backward sequence U_b by adding a negative sign to the reversed values.

Step 3: Deriving change points

Lastly, we identify the change points of the time series X based on the two generated sequences (U_f and U_b). We first identify the initial set of the change points as the points of intersection between the two sequences. Previous studies show that it is uncertain to recognize all of these change points as abrupt changes, as a change point can be induced by a sudden shift of the mean value over two stable periods [17]. These outlier points could be reevaluated by using additional detection methods, such as the double mass curve [18]. To avoid miscounting the change points while making the proposed method more applicable, we employ a statistical filter—the points of intersection falling beyond the 95% confidence intervals (CIs), which correspond to Z-scores $= \pm 1.96$, are rejected. This filter has been used in relevant MKS studies [19]. It is worth noting that the MKS test can also identify the monotonic trend or the change of direction—if a point of intersection is between the Z-scores of 0 and 1.96, the change is upward; if the point is between the Z-scores of -1.96 and 0, the change is downward.

Model implementation

In this section, we take the state of Virginia as an example to further elaborate on the model implementation. The MKS test can be implemented in Microsoft Excel by calling embedded functions. The datasets and codes are available on GitHub (https://github.com/peter best52/mks).

Data cleaning

Daily confirmed cumulative COVID-19 case data between March 22, 2020 and January 31, 2021 (in a total of 45 weeks) were obtained from the USAFacts website (https://usafacts.org/data/). Then, we aggregated the data on a weekly basis, generating a 45-week time series for each state representing new weekly cases. Lastly, to demonstrate the method, we extracted the data for Virginia as the time series X.
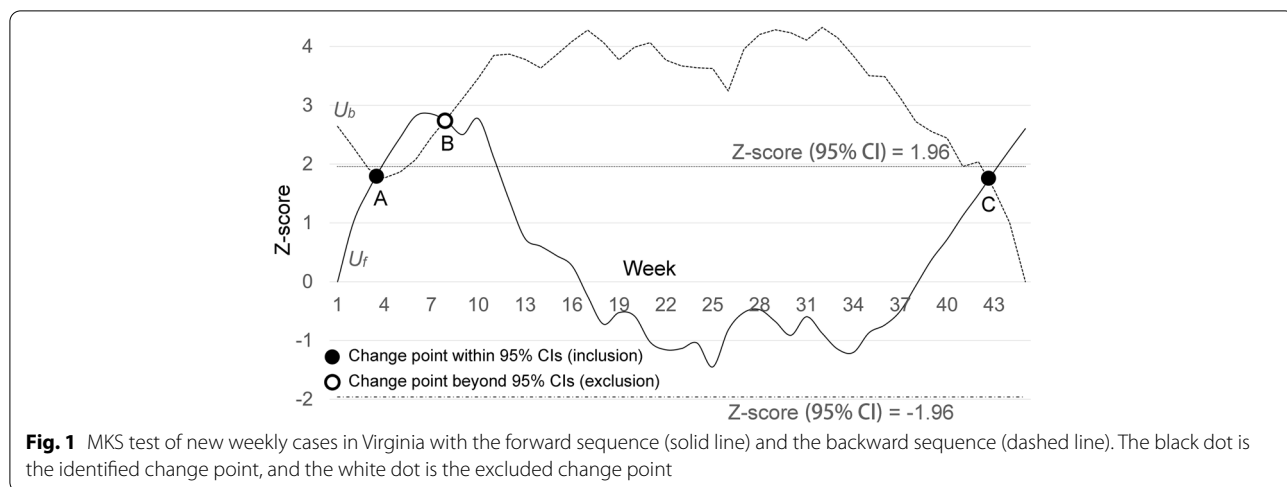
MKS test

For time series X, we derived m_i, the cumulative times that the case value of the current week is larger than that of each preceding week. Following this step, S_k was derived as the cumulative m_i ($i = 1, 2, ..., k$), according to Eq. (1); then, the mean value of S_k or $E(S_k)$ and the variance of S_k or $VAR(S_k)$ were derived by Eqs. (2) and (3), respectively. It is worth noting that, since k is the only independent variable in Eqs. (2) and (3), $E(S_k)$ and $VAR(S_k)$ are the same for all states in this study. Based on Eq. (4), we derived the forward sequence U_f for Virginia (solid line in Fig. 1).

Then, we reversed the time series X and derived X_r. We derived the intermediate sequence U_{fr} by applying Eq. (4) to X_r. Lastly, we derived the backward sequence U_b (dashed line in Fig. 1) by first reversing the sequence of values in U_{fr} and then adding a negative sign to these values.

Change point detection

The forward sequence (U_f) and the backward sequence (U_b) were plotted as the solid line and dashed line, respectively (Fig. 1). The points of intersection between the two sequences became the initial set of the change points. The thresholds of 95% CIs (Z-scores $= \pm$ 1.96) were set as the statistical filter. Only change points within the thresholds were retained. Specifically, in the case of Virginia, three points of intersection were initially detected. Week 4 (Point A in Fig. 1) and Week 43 (Point C in Fig. 1) were identified as the final change points with statistical confidence. Week 8 (Point B in Fig. 1) was excluded (Z-score $= 2.72$), as it fell beyond the thresholds.

Fig. 1 MKS test of new weekly cases in Virginia with the forward sequence (solid line) and the backward sequence (dashed line). The black dot is the identified change point, and the white dot is the excluded change point

Since both Point A and Point C were between Z-scores of 0 and 1.96, these changes were upward.
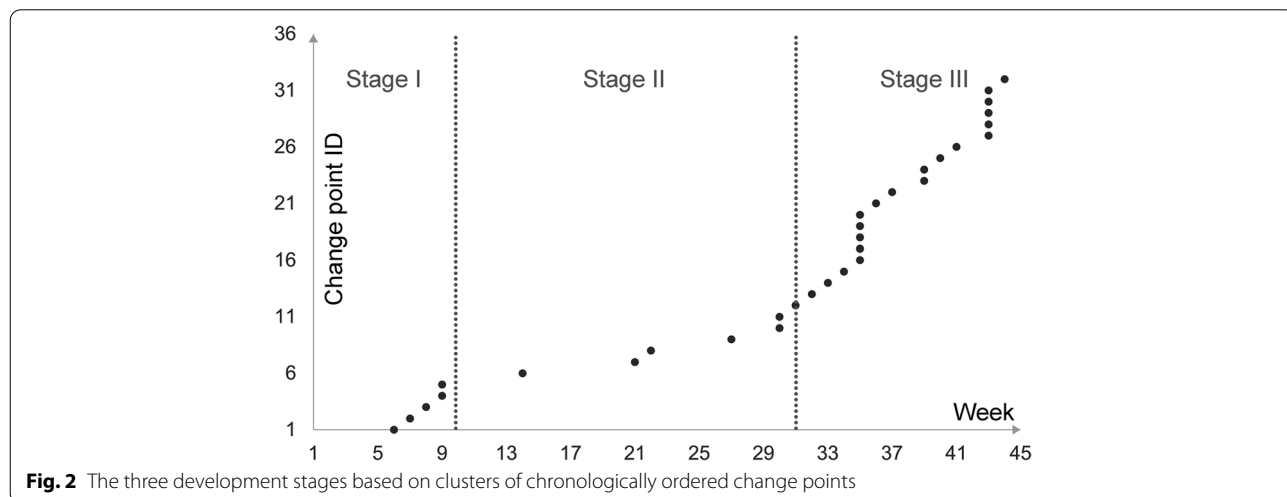
Results

By applying the MKS test to weekly new COVID-19 cases in 50 states, we identified that 30 states (60.0%) have at least one change point within the 95% CIs. For the unqualified states, most of them have no change points within the 95% CIs but have at least one change point beyond the 95% CIs. Only the state of Vermont has no change points either within the 95% CIs or beyond, meaning that there is no abrupt case decrease or increase during the entire study period.

To characterize the temporal distribution of these change points, we further divided the study period into three disease development stages, namely, Weeks 1–10 (March 23 through May 31, 2020), Weeks 11–30 (June 1 through November 19, 2020), and Weeks 31–45 (November 19, 2020 through January 31, 2021). These three stages were determined by the three clusters of chronologically ordered change points, as shown in Fig. 2. Based on the three development stages, we then mapped out the emergence of the change point for each state, as shown in Fig. 3.

Figure 4 shows the change points detected by the MKS test for the 30 states with at least one change point within the 95% CIs. Among these states, we identified that a single change point exists for 25 states, two change points exist for 4 states (i.e., LA, OH, VA, and WA), and three change points exist for one state (i.e., GA). Then, we further derived 6 change patterns based on the emergence and direction of the change point at the three stages, as shown in Table 1.
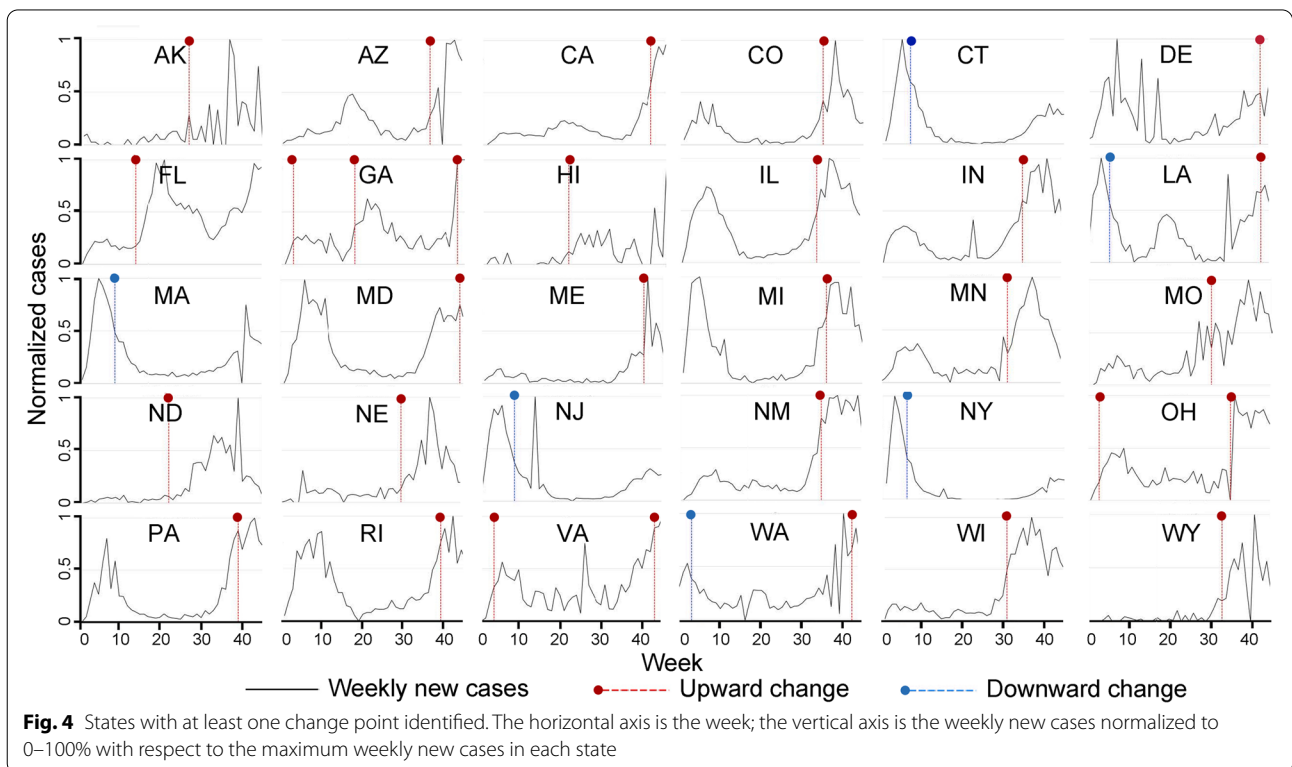

Fig. 2 The three development stages based on clusters of chronologically ordered change points

Fig. 3 The emergence of the change point for each state **a** at the first stage (Weeks 1–10), **b** at the second stage (Weeks 11–30), and **c** at the third stage (Weeks 31–45). The map is created by the authors

Fig. 4 States with at least one change point identified. The horizontal axis is the week; the vertical axis is the weekly new cases normalized to 0–100% with respect to the maximum weekly new cases in each state

Chen *et al. BMC Medical Research Methodology*   (2022) 22:233

Page 6 of 9

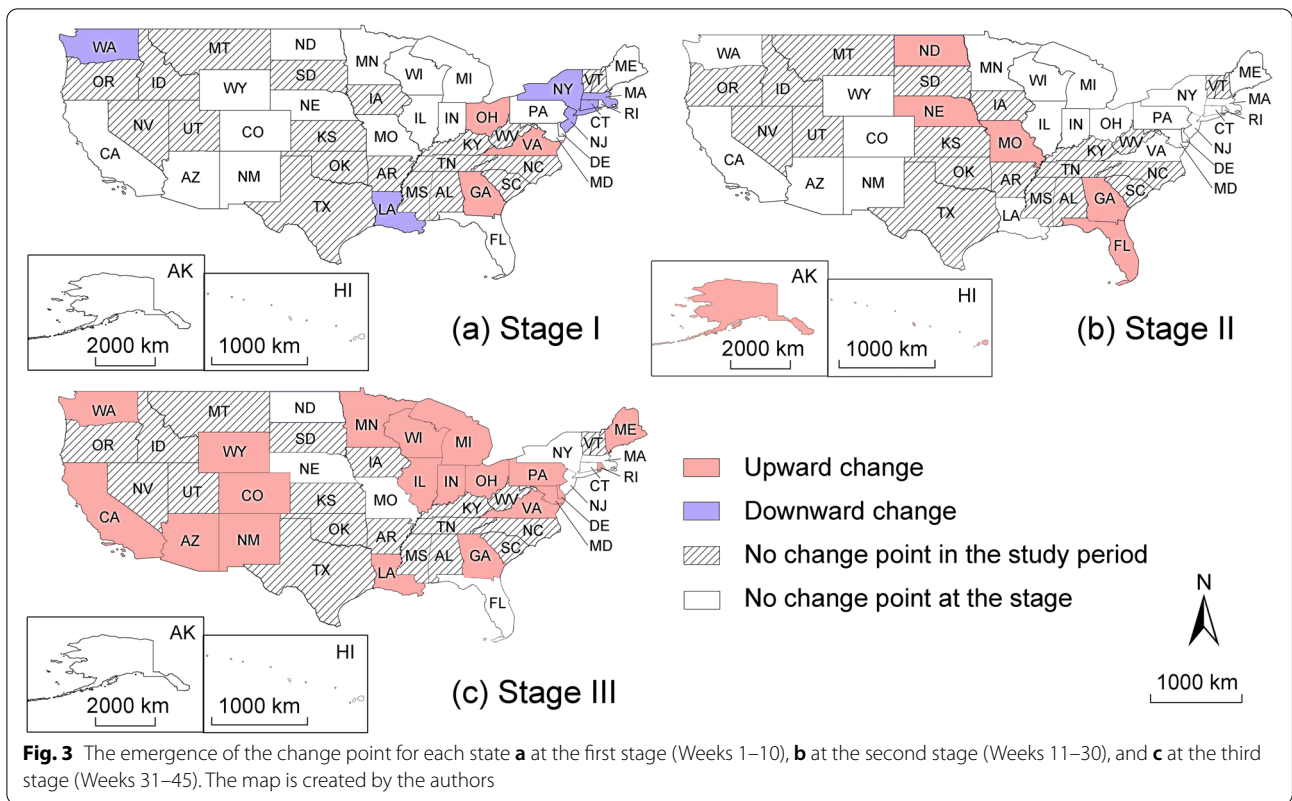**Table 1** Summary of change patterns based on the emergence and direction of change points at three stages

No.	Pattern	State
1	+++	GA
2	+/+	OH, VA
3	−/+	LA, WA
4	−//	CT, MA, NJ, NY
5	/+/	AK, FL, HI, MO, ND, NE
6	//+	AZ, CA, CO, DE, IL, IN, MD, ME, MI, MN, NM, PA, RI, WI, WY

+ upward change point, − downward change point, / no change point

Discussion

Two epidemiologic patterns can be identified in Table 1. First, the downward changes at the first stage (Pattern 4) appear only in Northeastern states (e.g., CT, MA, NJ, NY), as confirmed in Fig. 3a. This pattern can be explained by the immediate state policy actions on social distancing in this region during the early outbreak. After COVID-19 was declared a national emergency by the presidential proclamation on March 1, 2020 [20], most Northeastern states enforced social distancing regulations in late March and early April, including the closure of non-essential businesses and schools [21]. These policies largely restricted face-to-face interactions, slowed the virus diffusion, and eventually, suppressed the epi curves. Second, the upward changes at the third stage appear mostly in the Western states (e.g., AZ, CA, CO, NM, WA, WY) and the Midwestern states (e.g., IL, IN, MI, MN, OH, WI), as shown in Fig. 3c. This result is consistent with the observation that most Western and Midwestern states experienced an abrupt case surge in the late summer and fall [22]. The rising trend could be linked to their less restrictive reopening policies, especially reopening indoor dining without a statewide face-covering mandate [23].

To further validate the MKS test, we compared it with two other change point detection methods, the pruned exact linear time (PELT) method and the regression-based method (Table 2), both of which are commonly used for detecting multiple change points in time series data. Specifically, the PELT method searches for change points by minimizing a cost function over possible numbers and locations of change points, and it implements an efficient pruning to increase the computational efficiency [24, 25]. The regression-based method analyzes the time series using a regression model with multiple segments, where the coefficients shift from one stable regression relationship to another. It implements a dynamic programming approach to find segments that can minimize the residual sum of squares [26, 27]. We implemented the PELT method using the 'changepoint' package in R [25] and the regression-based method using the 'strucchange' package in R [28].

The validation tested if the MKS-identified change points can be confirmed by the two other methods. A confirmation is accepted if an MKS-identified change point is validated by another method within a two-week window. The comparison results are shown in Table 2. Based on the 36 MKS-identified change points, the MKS-test reaches 41.7% agreement (15/36) with the PELT method and 47.2% agreement (17/36) with the regression-based method. It is also worth mentioning that the other two methods identified at least one change point for every state, even when there is no obvious change of direction. The comparison results signify that the MKS test is a relatively conservative method for change point detection, as it can only detect abrupt changes and can thus avoid false-positive results.

Conclusions

To sum up, the MKS test has several advantages in change point detection. First and foremost, it is characterized by high computational efficiency and easy implementation. Users can easily implement this method in Microsoft Excel without any prior statistical knowledge or modeling skills. Second, the method can detect the change of direction, whereas some other methods (e.g., PELT) can only identify the existence of a change without specifying the direction. Third, since the MKS test is a nonparametric model, it can be applied to time series data where the distribution is not normal or has extreme variability. However, due to its conservative nature and moderate agreement with the other slower but more sensitive methods, we recommend using the MKS test primarily for initial pattern identification and data pruning, especially in large data. For example, to identify the change points in a long sequence of COVID-19 infection data, we can first use the MKS test to narrow down the time window where changes are likely to occur, and then use a second method (which has a higher computational cost but is more sensitive) to reconfirm the change pattern. In addition, as the conservativeness of the MKS test can be easily modified by adjusting the width of the statistical filter, future studies should examine how the quality of the results derived from the MKS test may vary as a function of the statistical filter.

This pilot study is the first to implement the MKS test for COVID-19 studies. An open-access tool is developed to facilitate the model implementation. With further validation and modification, the method can be applied to other health data, such as injuries, disabilities, and

Chen *et al. BMC Medical Research Methodology*   (2022) 22:233

Page 7 of 9

Table 2 Summary of the identified change points (CP) by the three methods

State	MKS			PELT		Regression-based			
	CP#1	CP#2	CP#3	CP#1	CP#2	CP#1	CP#2	CP#3	CP#4
AK	27			36		36			
AL				41		39			
AR				36		18	25	37	
AZ	37			40		38			
CA	41			40		39			
CO	36			34		10	34		
CT	8			9		10	35		
DE	43			14	38	14	38		
FL	15			15		16	22	29	39
GA	3	18	43	42		17	24	39	
HI	21			44		22			
IA				34		34			
ID				31		17	33	39	
IL	34			33		12	33		
IN	35			34		34			
KS				36		27	36		
KY				36		30	38		
LA	6	44		7	34	7	34		
MA	9			11		11	37		
MD	44			12	36	6	12	35	
ME	40			36		36			
MI	36			7	33	7	33		
MN	32			33		33	39		
MO	30			26		26	35		
MS				42		15	37		
MT				30		18	30		
NC				40		36			
ND	23			27		27	33	39	
NE	30			33		33	39		
NH				40		14	36		
NJ	9			14		14			
NM	35			34		34			
NV				36		17	24	36	
NY	6			7		7			
OH	3	36		35		12	35		
OK				36		17	31	39	
OR				35		35			
PA	39			36		11	36		
RI	39			13	36	13	36		
SC				40		16	23	39	
SD				31		26	32	38	
TN				36		17	31	37	
TX				39		17	23	33	39
UT				33		33			
VA	4	43		37		37			
VT				36		6	36		
WA	4	43		40		7	38		
WI	31			30		30	39		

Table 2 (continued)

State	MKS			PELT		Regression-based			
	CP#1	CP#2	CP#3	CP#1	CP#2	CP#1	CP#2	CP#3	CP#4
WV				36		31	37		
WY	33			34		34			

mortalities. By identifying key time points where chronologically ordered observations have a drastic change, the method can eventually contribute to revealing the etiology of these health outcomes and supporting public health decision-making.

Abbreviations
CDC: Centers for Disease Control and Prevention; CI: Confidence interval; COVID-19: Coronavirus Disease 2019; CP: Change point; MKS: Mann-Kendall-Sneyers; PELT: Pruned exact linear time; US: United States.

Authors' contributions
XC contributed towards conceptualization, writing the initial draft, and revising the draft. HW contributed towards conceptualization, methodology, and writing the initial draft. WL contributed towards visualizing the results and revising the draft. RX contributed towards methodology and revising the draft. All authors read and approved the manuscript.

Availability of data and materials
The data and codes for the study can be accessed on Github [https://github.com/peterbest52/mks].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Geography, University of Connecticut, Storrs, CT 06269, USA. [2]Institute for Collaboration on Health, Intervention, and Policy (InCHIP), University of Connecticut, Storrs, CT 06269, USA. [3]Department of Geosciences, Mississippi State University, Mississippi State, MS 39762, USA. [4]Department of Allied Health Sciences, University of Connecticut, Storrs, CT 06269, USA.

References
1. Chen X, Zhang A, Wang H, Gallaher A, Zhu X. Compliance and containment in social distancing: mathematical modeling of COVID-19 across townships. Int J Geogr Inf Sci. 2021;35(3):446–65.
2. Chen J, Gupta AK. On change point detection and estimation. Commun Stat Simul Comput. 2001;30(3):665–97.
3. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020;369(6500).
4. Vokó Z, Pitter JG. The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis. GeroScience. 2020;42(4):1075–82.
5. Zhang S, Xu Z, Peng H. Change Point Modeling of Covid-19 Data in: the United States; 2020.
6. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V. Research article summary: Inferring COVID-19 spreading rates and potential change points for case number forecasts. medRxiv. 2020. https://doi.org/10.1101/2020.04.02.20050922.
7. Mbuvha R, Marwala T. Bayesian inference of COVID-19 spreading rates in South Africa. PLoS One. 2020;15(8):e0237126.
8. Mann HB. Nonparametric tests against trend. Econometrica. 1945;13(3):245–59.
9. Wang J, Kwan M-P. An analytical framework for integrating the spatiotemporal dynamics of environmental context and individual mobility in exposure assessment: A study on the relationship between food environment exposures and body weight. Intern J Environ Res Public Health. 2018;15(9):2022.
10. Rahman MA, Yunsheng L, Sultana N. Analysis and prediction of rainfall trends over Bangladesh using Mann–Kendall, Spearman's rho tests and ARIMA model. Meteorol Atmospher Phys. 2017;129(4):409–24.
11. Dawood M. Spatio-statistical analysis of temperature fluctuation using Mann–Kendall and Sen's slope approach. Clim Dyn. 2017;48(3–4):783–97.
12. Ison D. Statistical procedures for evaluating trends in coronavirus disease-19 cases in the United States. Int J Health Sci. 2020;14(5):23.
13. Shaharudin SM, Ismail S, Samsudin MS, Azid A, Tan ML, Basri MAA. Prediction of epidemic trends in COVID-19 with mann-kendall and recurrent forecasting-singular spectrum analysis. Sains Malays. 2021;50(4):1131–42.
14. Fenta AA, Yasuda H, Shimizu K, Haregeweyn N. Response of streamflow to climate variability and changes in human activities in the semiarid highlands of northern Ethiopia. Reg Environ Chang. 2017;17(4):1229–40.
15. Sneyers R. On the statistical analysis of series of observations. Technical Note No. 143, World Meteorological Organization, Geneva, Switzerland. 1990.
16. Moreland A, Herlihy C, Tynan MA, Sunshine G, McCord RF, Hilton C, et al. Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement—United States, March 1–May 31, 2020. Morb Mortal Wkly Rep. 2020;69(35):1198.
17. Fu C, Wang Q. The Definition and Detection of the Abrupt Climatic Change. Chin J Atmos Sci. 1992;04:482–93.
18. Searcy JK, Hardison CH. Double mass curves. Geological Survey Water Supply Paper 1541-B, U.S. Geological Survey, Washington, D.C. 1960.
19. Some'e BS, Ezani A, Tabari H. Spatiotemporal trends and change point of precipitation in Iran. Atmos Res. 2012;113:1–12.
20. House W. Proclamation on declaring a national emergency concerning the Novel Coronavirus Disease (COVID-19) outbreak 2020 [Available from: https://www.whitehouse.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novelcoronavirus-disease-covid-19-outbreak/
21. Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J. Pandemic politics: Timing state-level social distancing responses to COVID-19. J Health Politic Policy Law. 2021;46(2):211–33.
22. Clark JK, McChesney R, Munroe DK, Irwin EG. Spatial characteristics of exurban settlement pattern in the United States. Landsc Urban Plan. 2009;90(3–4):178–88.

Chen *et al. BMC Medical Research Methodology*   (2022) 22:233

Page 9 of 9

23. Kaufman BG, Whitaker R, Mahendraratnam N, Smith VA, McClellan MB. Comparing associations of state reopening strategies with COVID-19 burden. J Gen Intern Med. 2020;35(12):3627–34.
24. Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc. 2012;107(500):1590–8.
25. Killick R, Eckley I. changepoint: An R package for changepoint analysis. J Stat Softw. 2014;58(3):1–19.
26. Bai J, Perron P. Estimating and testing linear models with multiple structural changes. Econometrica. 1998;66(1):47–78.
27. Zeileis A, Leisch F, Hornik K, Kleiber C. strucchange: An R package for testing for structural change in linear regression models. J Stat Softw. 2002;7:1–38.
28. Zeileis A, Leisch F, Hornik K, Kleiber C, Hansen B, Merkle EC, Zeileis MA. Package 'strucchange'. J Stat Softw. 2015.

Publisher's Note