scientific reports

Check for updates

OPEN

Discovering spatiotemporal patterns of COVID-19 pandemic in South Korea

Sungchan Kim[1], Minseok Kim[1], Sunmi Lee[1]✉ & Young Ju Lee[2]✉

A novel severe acute respiratory syndrome coronavirus 2 emerged in December 2019, and it took only a few months for WHO to declare COVID-19 as a pandemic in March 2020. It is very challenging to discover complex spatial–temporal transmission mechanisms. However, it is crucial to capture essential features of regional-temporal patterns of COVID-19 to implement prompt and effective prevention or mitigation interventions. In this work, we develop a novel framework of compatible window-wise dynamic mode decomposition (CwDMD) for nonlinear infectious disease dynamics. The compatible window is a selected representative subdomain of time series data, in which compatibility between spatial and temporal resolutions is established so that DMD can provide meaningful data analysis. A total of four compatible windows have been selected from COVID-19 time-series data from January 20, 2020, to May 10, 2021, in South Korea. The spatiotemporal patterns of these four windows are then analyzed. Several hot and cold spots were identified, their spatial–temporal relationships, and some hidden regional patterns were discovered. Our analysis reveals that the first wave was contained in the Daegu and Gyeongbuk areas, but it spread rapidly to the whole of South Korea after the second wave. Later on, the spatial distribution is seen to become more homogeneous after the third wave. Our analysis also identifies that some patterns are not related to regional relevance. These findings have then been analyzed and associated with the inter-regional and local characteristics of South Korea. Thus, the present study is expected to provide public health officials helpful insights for future regional-temporal specific mitigation plans.

A novel virus named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as the pathogen for the outbreak of COVID-19 in December 2019[1]. Since then, the COVID-19 pandemic has posed huge challenges to public health officials all around the world. Due to the frequent international flights and human mobility, it took only a few months that COVID-19 spread to more than 200 countries. Currently, many developed countries are in the process of vaccinating their citizens, and some countries hope to soon achieve herd immunity[2]. In fact, the majority of countries with higher proportions of vaccination have shown a significant reduction in the number of COVID-19 cases and deaths from March to June 2021.

Unfortunately, as of July 3, 2021, the confirmed COVID-19 cases have increased worldwide due to the Delta variant[3]. This is one of the new variants of COVID-19 and is a potential threat to the goal of herd immunity. At this point, there are a total of more than 180 million confirmed cases and nearly 4 million deaths in 220 countries[4]. Among others, the US, India, and Brazil are the top three countries of COVID-19 cumulative cases and deaths officially; the US (33,709,176; 605,524), India (30,502,362; 401,050), Brazil (18,687,469; 521,952), respectively. These numbers indicate officially reported cases and may be considerable underestimates due to false negatives[5,6], lack of tracking systems[7], and overloading of healthcare facilities[8]. Therefore, it is urgent to understand the spatial–temporal transmission dynamics of COVID-19 to propose effective interventions to mitigate and reduce further morbidity and mortality. Apparently, COVID-19 has disproportionately affected different regional, social, and economic statuses even in developed countries[9–11]. South Korea shows a significant level of variability in the spatiotemporal patterns of COVID-19 as well. As of March 9, 2020, South Korea had a total of 7382 confirmed cases and the largest outbreak of COVID-19 besides China[12]. This was mainly due to few super-spreading events at the Shincheonji Church in Daegu Province and Daenam health care facility in Gyeongsang Province from February 20 to March 20. As of July 3, 2021, the total confirmed cases and deaths of COVID-19 increased to 159,342 and 2025 in South Korea, respectively. The spatial and temporal heterogeneity of COVID-19 has changed over time.

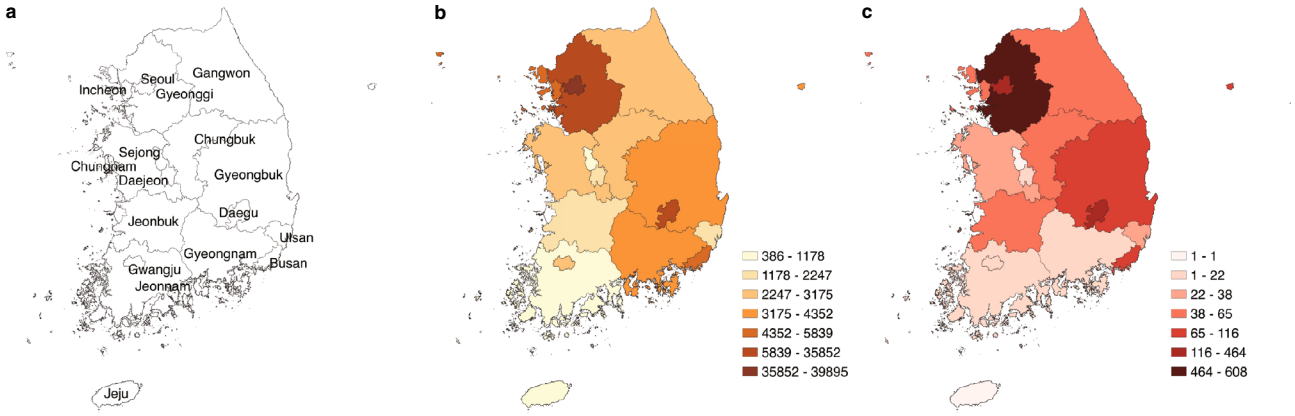
[1]Department of Applied Mathematics, Kyung Hee University, Yongin, Republic of Korea. [2]Department of Mathematics, Texas State University, San Marcos, TX, USA. ✉email: sunmilee@khu.ac.kr; yjlee@txstate.edu

An in-depth understanding of COVID-19 requires the use of mathematical modeling, which has played an essential role to explain complex spatial and temporal transmission dynamics of various infectious diseases. These include recent emerging infectious diseases; novel H1N1 influenza, SARS-CoV-1, Zika, MERS-CoV, and SARS-CoV-2[13]. Recent emerging infectious diseases tend to spread all over the world within a shorter time scale due to dramatic increases in international flights and human mobility[13,14]. There has been much research on spatial–temporal patterns of COVID-19 using various modeling approaches[9,10,15,16]. The spread of COVID-19 during an early stage of the pandemic in South Korea was investigated; 12 significant spatiotemporal clusters were identified and analyzed[17]. They observed that early interventions including 3T (test, trace, treat) were effective so that the cluster size and duration were shortened in time. Castro et al. investigated the spatial and temporal patterns of COVID-19 in Brazil and identified several key factors for failure of region-specific effective interventions[10]. Sartorius et al. employed a Bayesian hierarchical space-time SEIR model to assess the spatiotemporal variability of COVID-19 in England and they examined that mobility and social distancing played a critical role in the spatiotemporal patterns of mobility and mortality[18]. Wang et al. demonstrated the spatiotemporal characteristics and trends of COVID-19 in the United States and the various complex interactions with preventive efforts on COVID-19 were analyzed[19]. Bag et al. explored the spatiotemporal patterns of COVID-19 in India, and further, they examined the interplay between the space-specific patterns and governmental responses[20].

However, it is very challenging to discover spatial–temporal transmission mechanisms by the standard equation-based framework introduced above. In this work, we propose to discover the high complexity of spatial–temporal dynamics for COVID-19 transmission by employing a data-driven approach based on dynamic mode decomposition. The dynamic mode decomposition method (DMD) originated in the fluid dynamics community as a method to decompose complex flows into spatiotemporal coherent structures. DMD is a matrix-free, data-driven method capable of providing an accurate decomposition of a complex system into spatial–temporal coherent structures that may even be able to predict the short-time future state. Since Schmid and Sesterhenn[21] first introduced the DMD algorithm and demonstrated its ability, there have been tremendous works in DMD, and DMD became even more popular and is still in development today. This includes a sparsity-promoting DMD[22], a randomized DMD[23], which scales with the intrinsic rank of the dynamics, a consistent DMD, a new method for computing DMD operator based on a variational framework[24]. DMD has been successfully used for computational epidemiology[25]. Bistrian et al.[26] proposed a framework for reduced-order modeling and forecasting of non-intrusive data with application to epidemiology, using a technique based on randomized DMD combined with ARIMA (AutoRegressive Integrated Moving Average)[27] and this has been used also for modeling of SARS-CoV-2 dynamics obtained from the raw data reported by World Health Organization[28]. Proctor et al.[29] have demonstrated how DMD can aid in the analysis of spatial–temporal disease data. It is shown that DMD is an effective and efficient computational analysis tool for the study of infectious disease taking into account several tests' data such as Google Flu Trends data, pre-vaccination measles in the UK, and paralytic poliomyelitis wild type-1 cases in Nigeria. We note though that in particular, Google Flu Trends data is shown to be overall more influenced by the media clamor than by true epidemiological burden as studied in[30,31].

In this paper, we propose a compatible window-wise dynamic mode decomposition (CwDMD). The notable difference of our work from other available works is that we tackle COVID-19 time series data in a way that the data sets are made to be consistent in the sense of Tu et al.[32]. Basically, the compatible window is a selected the data set that can be modeled by a linear operator, thereby making DMD analysis meaningful. Further, we show that the consistency is equivalent to the linearity and demonstrate that DMD produces misleading data interpretation for inconsistent or nonlinear data in general. This indicates that the direct and reliable DMD analysis of large time-series data such as COVID-19 data is not feasible. We develop a strategy to choose an adequate set of representative subdomains called windows in which an appropriate balance or compatibility between spatial and temporal resolutions is built. The total size-times duration of all the windows serving a given system depends only on local situations that can arise in the full time-series data. We then apply DMD to each window that results in robust and reliable data analysis. It is easy to see that if the data is linear, DMD analysis will be adequate while it is not for nonlinear data. Oftentimes such an inadequacy has been justified through the Koopman mode analysis in the framework of Hankel DMD. However, it is well-known that Hankel DMD is proven to work only for ergodic data[33,34]. These frameworks, therefore, can not be applied in general, for highly nonlinear data. Such data includes internal solitary wave as discussed in[35] as well as COVID-19 data analyzed in the present paper, which are not necessarily ergodic. It is notable that a recent work by Zhang et al.[35] is closely relevant to our method. However, their work is not based on compatible windows, i.e., the choice of windows is constructed without respecting the consistency. Phase studies are not investigated either unlike the proposed study in this paper. Furthermore, we make significant and novel progress from the consistency assumption that the data fitting for any given window can be achieved accurately only by finding the coordinate of any single data within the window in terms of DMD modes. This allows us to achieve a significant computational reduction. The identified coordinate is then used as a certain scale for the selection of important DMD modes.

Our new method is used to investigate the spatiotemporal patterns of COVID-19 in South Korea from January 20, 2020 to May 10, 2021. A total of four compatible windows have been selected from the given COVID-19 time series data. The spatiotemporal patterns of these four windows are then analyzed by a few important DMD modes selected based on our new criterion. Several hot and cold spots were identified, their spatial–temporal relationships, and some hidden regional patterns were discovered. Our analysis reveals that the first wave was contained in the Daegu and Gyeongbuk area, but it spread rapidly to the whole of South Korea after the second wave. Later on, the spatial distribution is seen to become more homogeneous after the third wave. These findings have then been associated with the inter-regional and local characteristics of South Korea. We expect that the present study can provide public health officials helpful insights for future regional-temporal specific mitigation plans.



d

	Population (× 1,000)	Area (km²)	Density (per km²)	Cumulative cases					Cumulative deaths
				Total	*1st wave*	*2nd wave*	*3rd wave*	*after the 3rd wave*	
Total	51781	100350.7	516.0	124679	10298	10077	50827	48408	1873
Seoul	9602	605.2	15865.8	39895	629	3990	18183	15468	464
Busan	3344	770.2	4341.7	5286	137	320	2179	2501	122
Daegu	2419	883.5	2738.0	9448	6846	201	1138	1123	221
Incheon	3951	1065.3	3708.8	5839	92	588	2773	1994	60
Gwangju	1488	501.2	2968.9	2518	30	294	1301	679	22
Daejeon	1500	539.6	2779.8	1835	40	232	656	730	20
Ulsan	1140	1062.4	1073.0	2247	43	97	774	1290	38
Sejong	349	465.3	750.1	386	46	28	114	190	1
Gyeonggi	13405	10193.9	1315.0	35852	668	3161	14413	15819	608
Gangwon	1515	16833.3	90.0	2843	53	155	1420	1128	50
Chungbuk	1632	7418.2	220.0	2708	45	105	1388	1124	65
Chungnam	2204	8254.7	267.0	3175	141	309	1459	1137	36
Jeonbuk	1792	8072.1	222.0	2033	18	112	876	979	58
Jeonnam	1765	12335.7	143.1	1178	16	140	560	418	11
Gyeongbuk	2655	18964.3	140.0	4352	1364	173	1434	1315	79
Gyeongnam	3350	10534.6	318.0	4268	117	139	1696	2243	19
Jeju	670	1850.8	362.0	816	13	33	463	270	1

Figure 1. Spatial distribution of the cumulative confirmed and deaths of COVID-19 as of May 10, 2021. (**a**) A map of South Korea. South Korea is divided into 17 first-tier administrative divisions: 7 metropolitan cities (Seoul, Busan, Daegu, Incheon, Gwangju, Daejeon, and Ulsan), 1 special self-governing city (Sejong), and 9 provinces. The metropolitan area refers to Seoul, Incheon, and Gyeonggi. (**b**) Cumulative confirmed cases. (**c**) Cumulative deaths. Geographical descriptions such as population, area, and population density of each region; and COVID-19 profiles are in (**d**). Population density between metropolitan cities and non-metropolitan areas is extremely polarized, except Gyeonggi. The total population of three metropolitan areas is about 26 million as of May 2021, which is more than 50% of the South Korean population.

Results

Spatial–temporal characteristics of COVID-19 in South Korea. In this section, we present an overview of COVID-19 data collected in South Korea (see Fig. 1 for more description). Daily confirmed cases and deaths of COVID-19 from January 20, 2020 to May 10, 2021, were obtained from the Korea Centers for Disease Control and Prevention (KCDC) and each provincial website[12]. As of May 10, 2021, there were a total of 127,772 COVID-19 confirmed cases and 1875 deaths in South Korea. To analyze the spatiotemporal patterns of COVID-19, the spatial distribution of COVID-19 confirmed cases is refined in 17 first-tier administrative divisions of South Korea. Figure 1 shows a South Korea map (a) with spatial distributions of the cumulative number of COVID-19 confirmed cases (b) and the cumulative number of COVID-19 deaths (c). As displayed in b, c, d of Fig. 1, South Korea shows a high level of spatial and temporal heterogeneity in 17 regions. We can observe that the main characteristics of the temporal patterns of South Korea can be placed into the particular four stages, i.e., three big waves and the last stage. More precisely, the first window is from January 20, 2020 to April 26, 2020, the second window is from July 28, 2020 to October 12, 2020, the third window is from November 3, 2020 to February 1, 2021, and the period after the third wave is February 2, 2021, to May 10, 2021. These are chosen as four windows and represented by different colors in Fig. 2a.

The first case of COVID-19 in South Korea was a 35-year-old Chinese woman who traveled from Wuhan, China, and was confirmed on January 20, 2020. She entered the Incheon international airport and she was isolated at a hospital upon entry. After the index case, only 30 confirmed cases have occurred until February 17, 2020. However, there was an explosive outbreak in Daegu due to the superspreading events from the Shincheonji Church-related clusters from February 18 to March 23, 2020[12]. As a result, the first wave (January 20, 2020–April 26, 2020, see Fig. 2a) was focused on the Daegu and Gyeongbuk area with almost 80% of a total of 10,298 cases (Daegu 6846 and Gyeongbuk 1364). Since March 2020, the epicenter of COVID-19 has begun to move from the Daegu and Gyeongbuk area to Seoul and Gyeonggi regions. A few sporadic clusters of COVID-19 continued in

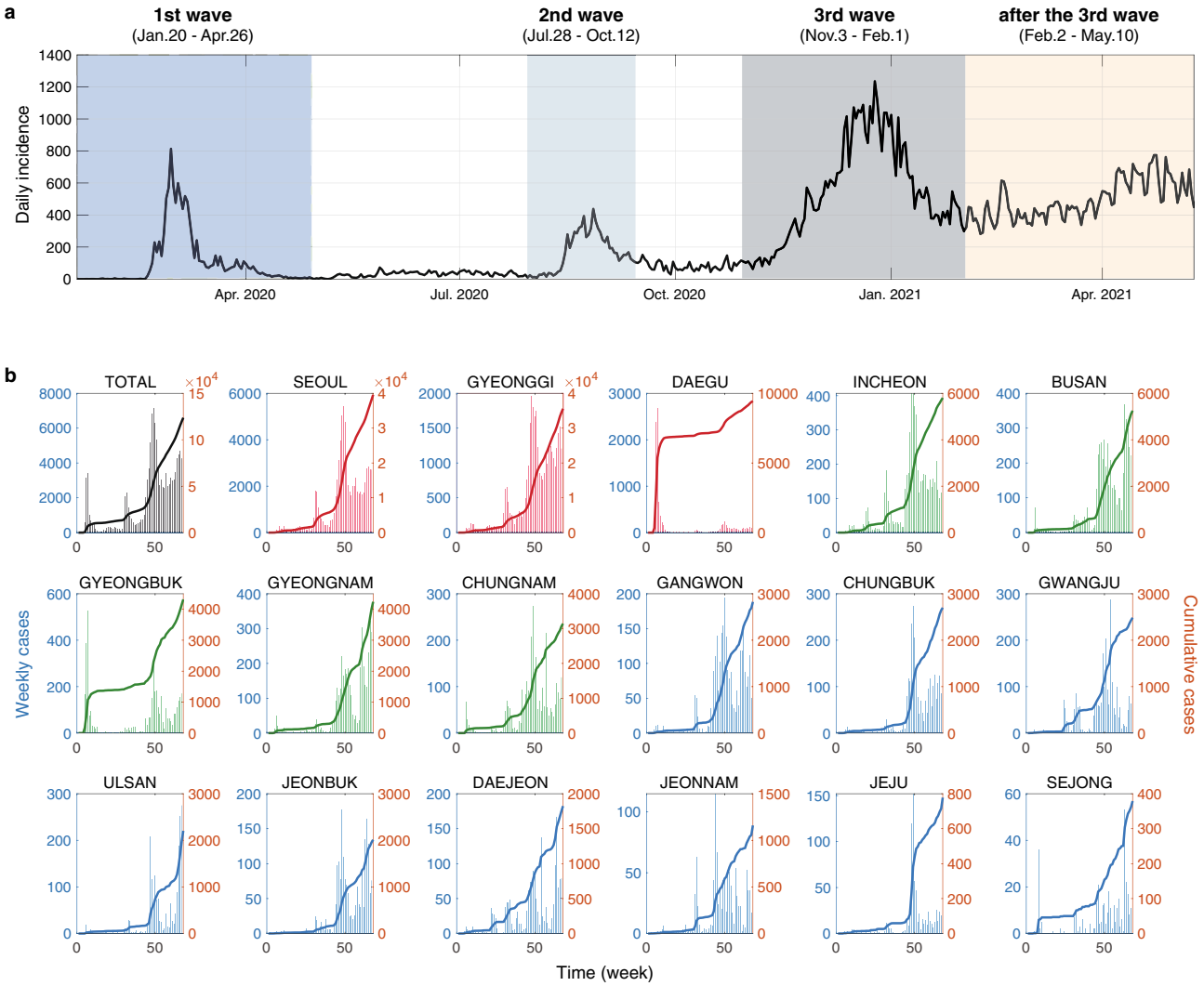


Figure 2. Time series of COVID-19 outbreak in South Korea. (**a**) Daily incidence of COVID-19 in South Korea. South Korea went through three big waves, after the third wave, the incidence has been maintained with no significant increase or decrease. The four windows of main interest were colored and given as; (1) the first wave (January 20, 2020–April 26, 2020); (2) the second wave (July 28, 2020–October 12, 2020); (3) the third wave (November 3, 2020–February 1, 2021); and (4) after the third wave (February 2, 2021–May 10, 2021). (**b**) Weekly incidence and cumulative cases in 17 regions, plotted as the bars and as a curve, respectively. The three highest cumulative cases, the next five highest cases, and the rest cases are marked with red, green, and blue, respectively.

Seoul including the Guro call center and the Itaewon club cluster in May 2020. From July 28 to October 12, 2020, the second wave started in Seoul and Gyeonggi Province (see Fig. 1d). The main cause of the second wave was the rally held at Gwanghwamun Square in Seoul. Seoul city has the highest in the confirmed cases and Gyeonggi Province has the second-highest in the confirmed cases and the highest in deaths. The largest wave was the third wave from November 3, 2020 to February 1, 2021. This was partly due to the winter seasons, which results in a favorable condition for close contact between people staying indoors. After the third wave, the constant level of COVID-19 cases has been maintained nationwide from February 2, 2021, to May 10, 2021.

Analysis of spatial–temporal COVID-19 in South Korea. In this section, we shall present the analysis of spatial–temporal COVID-19 in South Korea. Figure 2 displays confirmed cases of COVID-19 in South Korea from January 20, 2020, to May 10, 2021. Panel a of Fig. 2 shows the daily confirmed cases while the panel b of Fig. 2 illustrates region-specific COVID-19 weekly confirmed cases (bars on the left) and cumulative cases (solid curves on the right) for 17 first-tier administrative divisions of South Korea. The three highest cumulative cases, which include Seoul, Gyeonggi, and Daegu are marked in red, the next five highest cases are marked in green, and the rest of the cases are marked in blue.

First of all, Supplementary Fig. S1 displays the evolution of spatial distributions in 17 regions; the top panels show the cumulative number of COVID-19 cases per 100,000 on the last day of each period. The bottom panels show the cumulative number of COVID-19 cases per 100,000 during each period. The bottom panels indicate

that the hot spots were moving from Daegu to Seoul and Gyeonggi while Jeonnam remained the cold spot in the first, third, and last periods. Interestingly, Daegu and Gyeongbuk were the cold spots during the second wave after the severe first outbreak.

The chosen four windows are then used to apply CwDMD, which results in the discrete DMD modes and eigenvalues in each window. Supplementary Figs. S2–S5 compare the results of the DMD data fitting with the region-specific COVID-19 data for each window. There is a perfect agreement between the COVID-19 data (red dot) and the DMD output (black solid) in all 17 regions.

CwDMD has been used to investigate the spatiotemporal pattern of COVID-19 in 17 regions, whose discussions are presented in the following four subsections. Note that a few important DMD modes selected in each window are categorized into three regimes, oscillatory, growing, and decaying. These are then used for the phase and magnitude analysis of each window.

The first wave. The first wave is chosen as the total of 14 weeks and so, the spatial vs temporal resolution is 17 to 14. This is compatible as discussed in the section for "Methods".

In Fig. 3, we show the power of DMD modes in a, i.e., the measure of the scaled size of the DMD modes (see the section of "Methods" in details). The power is used for the selection of dominant DMD mode and the selected DMD mode is then utilized for both magnitude and phase analysis.

We note that the first three DMD modes of the highest power were chosen and they are denoted by #1, #2, and #3 as shown in Fig. 3a. In fact, the selected three DMD modes correspond to the growing, the oscillatory, and the decaying modes, respectively in the discrete dynamical system for the first window. The magnitude analysis has been performed using these dominant DMD modes. We observe that all three DMD modes show that Daegu and Gyeongbuk have the largest magnitude and Seoul and Gyeonggi are next. These are indicated by gray bars in b, c, d, respectively, and consistent with the cumulative confirmed cases of the first wave given in e and f of Fig. 3.

Next, we explored the phase analysis from the three selected DMD modes. We note that phase or phase difference can be interpreted as the time (in week) between peak to peak of the region-specific COVID-19 outbreak. Namely, the smaller the phase difference of two different regions is, the closer the peaks of these regions will be. We find that in all three DMD modes, the phases of Busan, Gyeongnam, and Chungnam are similar. Note that these three regions are close to the epicenter. Consequently, we find a strong correlation between the phase of the southern part of South Korea and the distances from the epicenter, i.e., Daegu and Gyeongbuk. This is consistent with the data presented in f of Fig. 3.

On the other hand, the phase of DMD mode #1, shows that there is a time lag of 2–3 weeks between the peaks of Seoul and Gyeonggi and those from Busan, Gyeongnam, Gyeongbuk, and Daegu. In particular, from the fact that the DMD mode #1 is a growing mode, the above conclusion indicates that there was definitely a different cause for the COVID-19 outbreak of Daegu and Gyeongbuk from that of Seoul and Gyeonggi. Note that it can be clearly identified in the graph of e and f in Fig. 3. More precisely, the weekly confirmed cases of Seoul and Gyeonggi are similar to those of other regions from weeks 5 to 7. However, the confirmed cases began to increase from week 8 to 13, while those of other regions decreased. We later found that this peculiar behavior could be associated with a few large workplace-related clusters such as the Guro-Gu call center in Seoul and Gyeonggi from March 2020 to April 2020[36,37].

The second wave. The second wave is chosen as the total of 11 weeks and so, the spatial vs temporal resolution is 17 to 11. This is compatible as discussed in the section for "Methods".

In Fig. 4, we show the power of DMD modes in a. The power is used for the selection of dominant DMD mode and the selected DMD mode is then utilized for both magnitude and phase analysis. We note that the first two DMD modes of the highest power were chosen and they are denoted by #1 and #2 as shown in Fig. 4a. The selected two DMD modes correspond to the oscillatory and the growing modes, respectively. The magnitude analysis has been performed using these dominant DMD modes. The weekly confirmed cases of the total of six selected regions are then shown in d of Fig. 4.

The magnitude analysis using both of these selected DMD modes shows that Seoul and Gyeonggi have a significantly large magnitude of confirmed cases. This is in fact, consistent with the data shown as Fig. 4d. The main drive behind this large magnitude can be correlated with the outbreak from the rally held at Gwanghwamun Square in Seoul on August 15, 2020. Note that this rally was organized by SarangGeil Church in Seoul and people from all regions of South Korea participated. We observe that, unlike the first wave, the magnitude of Daegu and Gyeongbuk are relatively small. This can be attributed to the continued strict mitigation interventions in the Daegu and Gyeongbuk area since the first wave. See also the COVID-19 cases shown in Fig. 1 as well as in Supplementary Fig. S1, which are consistent with our magnitude analysis for Daegu and Gyeongbuk.

We now explore the phase analysis from the two selected DMD modes. First, we begin with the following facts; (1) the maximum phase difference in the DMD mode #1 is between Sejong and Ulsan and its value is 1.54 weeks; (2) the maximum phase difference is 1.04 weeks and it is between Busan and Jeju in the DMD mode #2. The relatively short phase difference indicates that the second wave can be characterized as an almost simultaneous nationwide spread. This can be attributed to the fact that all participants from all regions who attended the rally in Seoul returned to their home region within a few days, i.e., less than a week[38].

The third wave. The third wave is chosen as the total of 13 weeks and so, the spatial vs temporal resolution is 17 to 13. This is compatible as discussed in the section for "Methods".

In Fig. 5, we show the power of DMD modes in a. The power is used for the selection of dominant DMD mode and the selected DMD mode is then utilized for both magnitude and phase analysis. We note that the first two DMD modes of the highest power were chosen and they are denoted by #1 and #2 as shown in Fig. 5a. The

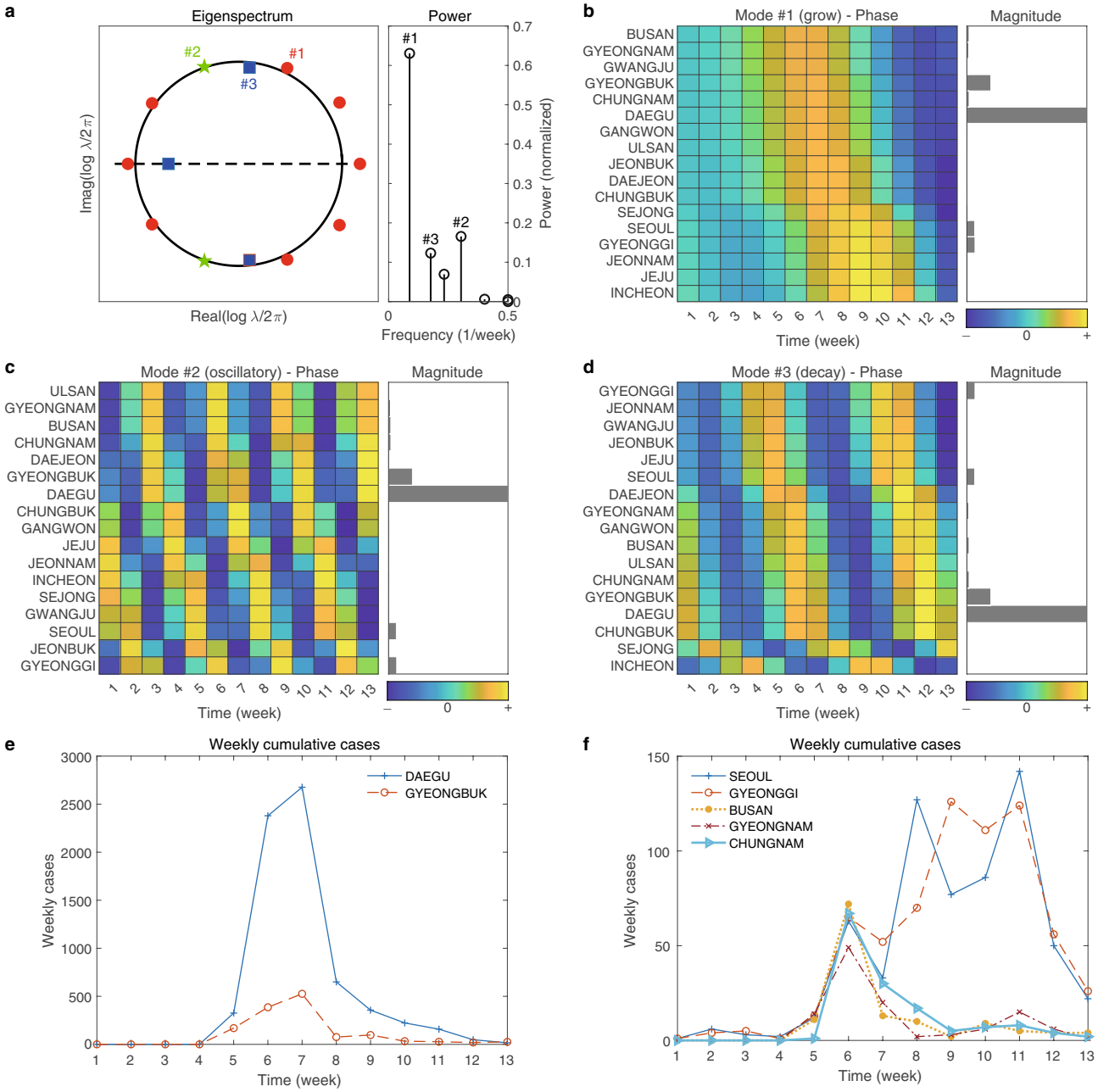


Figure 3. The first wave: DMD eigenvalues and modes. (**a**) shows eigenspectrum $\{\lambda_j\}_{j=1,\ldots}$ in the left and powers, defined as $\{|\lambda_j^p|\|\alpha_j\phi_j\|_F\}_{j=1,\ldots}$, in the right. The first three DMD modes that represent growing, oscillatory and decaying modes are enumerated as #1, #2, and #3. (**b**–**d**) Show the phase and magnitude of the selected DMD modes, #1, #2, and #3, respectively. (**e**) and (**f**) show time series of weekly cumulative cases for some selected regions. (**e**) is for high transmission areas, Daegu and Gyeongbuk, while (**f**) is for other relatively low transmission areas.

selected two DMD modes correspond to the growing and the oscillatory modes, respectively. The magnitude analysis has been performed using these dominant DMD modes. The weekly confirmed cases of the total of eight selected regions are then shown in d of Fig. 5.

The magnitude analysis using both of these selected DMD modes shows that Seoul and Gyeonggi have a significantly large magnitude of confirmed cases, similar to the second wave. This is in fact, consistent with the data shown as Fig. 5d. This is due to the cold winter seasons, as people favorably stayed indoors for close contacts, which is enhanced by the higher population density in Seoul and Gyeonggi; the South Korean population is highly disproportionate and the metropolitan area has more than 50% of the total South Korean population.

The phase analysis in this wave shows that the maximum phase difference is larger than that of the second wave for both modes. Namely, the maximum phase difference in the DMD mode #1 is 4.02 weeks, which is between Busan and Jeonnam, while the maximum phase difference in the DMD mode # 2 is 4.13 weeks, which

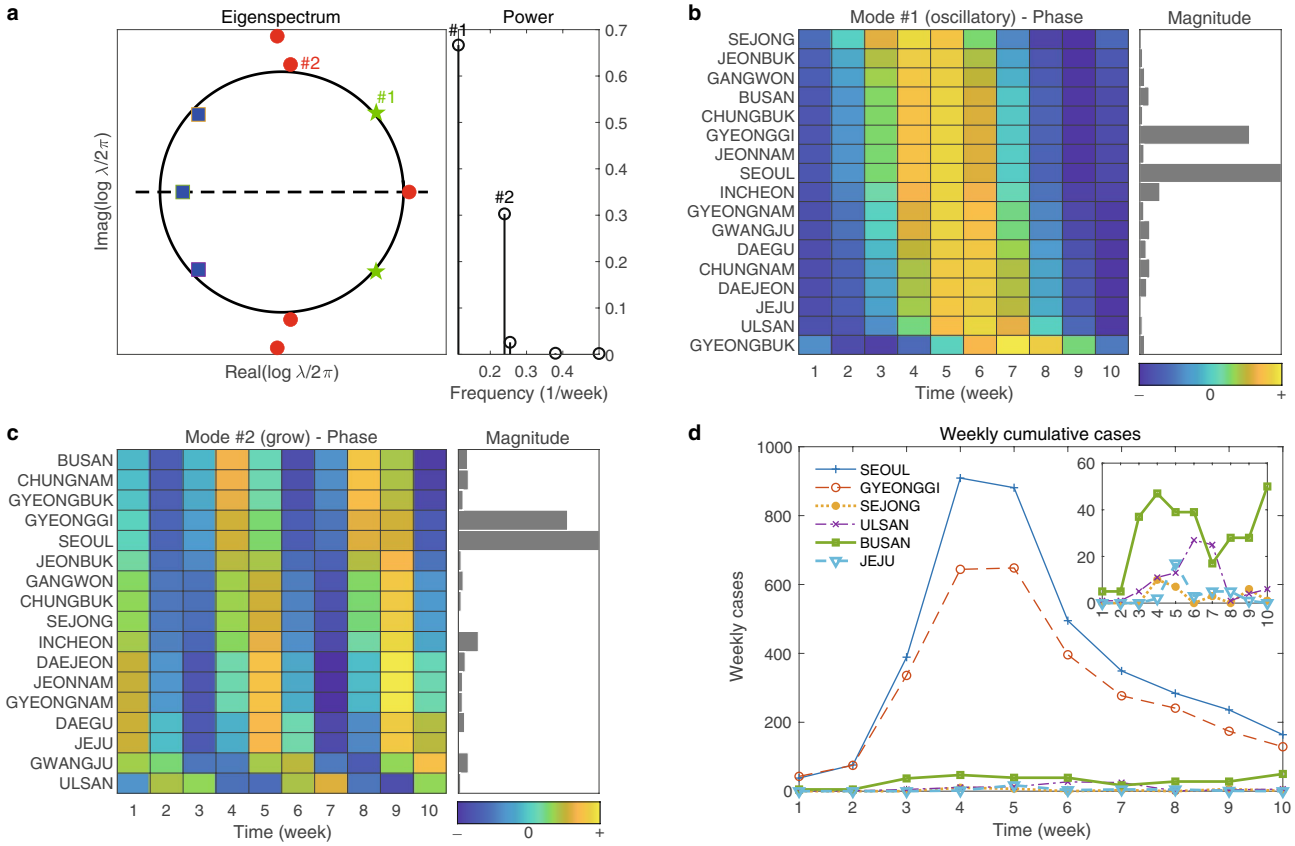


Figure 4. The second wave: DMD eigenvalues and modes. (**a**) shows eigenspectrum $\{\lambda_j\}_{j=1,\ldots}$ in the left and powers, defined as $\{|\lambda_j^p|\|\alpha_j\phi_j\|_F\}_{j=1,\ldots}$, in the right. The first two DMD modes of highest powers are enumerated as #1 and #2. (**b**) and (**c**) show phase and magnitude of the #1 and #2 DMD modes, respectively. In the phase diagram, regions, whose phases are similar are gathered. (**d**) is the time series for weekly cumulative cases for some selected regions.

is between Gyeonggi and Jeonnam. In particular, regions grouped according to the higher phase similarity are (1) Busan, Gyeongnam, and Ulsan, which are all located in the southeast area, (2) Seoul, Gyeonggi, and Incheon, which are all located in the northwest area, and (3) Daegu and Gyeongbuk, which are at the central area. This analysis identifies that there are strong spatial correlations in the third wave. This seems to be natural. But, to our surprise, we observe that there is more or less independent phase behavior between Gwangju and Jeonnam in DMD mode # 2. This means that Jeonnam is not much affected by the outbreak of COVID-19 in Gwangju, even if Jeonnam surrounds Gwangju. In fact, it is in this way throughout the whole time when COVID-19 data is collected. This indicates that the expected spatial correlation is sometimes misleading. Additionally, the similar phenomenon is also observed between in Daejeon and Chungnam.

The period after the third wave. The period after the third wave is chosen as the total of 13 weeks and so, the spatial vs temporal resolution is 17 to 13 again like the third wave. The main feature of this period is that the weekly incidence is relatively large all over South Korea.

In Fig. 6, we show the power of DMD modes in A. The power is used for the selection of dominant DMD mode and the selected DMD mode is then utilized for both magnitude and phase analysis. We note that a single DMD mode shows the dominant power and so, only this DMD mode is chosen and denoted by #1 as shown in Fig. 6a. The selected DMD mode corresponds to the oscillatory mode.

The magnitude analysis using this selected DMD mode shows that Seoul and Gyeonggi have the largest magnitude, which is consistent with the highest cumulative COVID-19 cases during the period after the third wave in these regions as shown in Fig. 2b. This consistency also holds for the next largest magnitudes or cumulative cases occurring in the southeast areas, which include Busan, Ulsan, and Gyeongnam.

The phase analysis in this period shows that the maximum phase difference is 11.8 weeks, which is from Incheon and Gwangju. Furthermore, the phase difference between neighboring regions such as Seoul and Gyeonggi, Daejeon and Chungnam, and Daegu and Gyeongbuk is also more than three weeks, which is relatively large. This indicates that overall large weakly incidence in each region is local in nature. Namely, the outbreaks in each region are mainly due to local outbreaks within the region and the inter-regional correlation of outbreaks seems to be irrelevant in this period. This has been further justified by investigating the spatial variations using the estimation of so-called the coefficient of variation below.

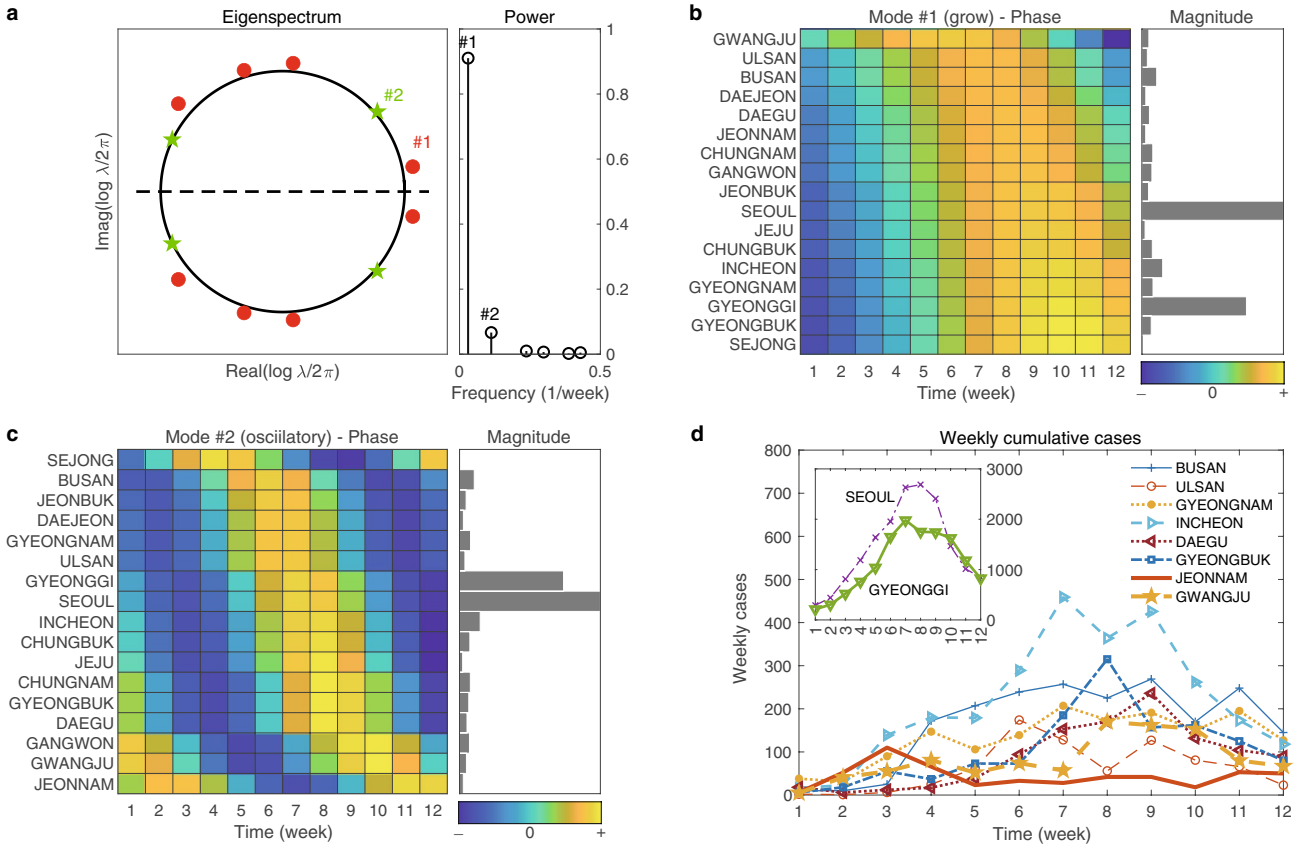


Figure 5. The third wave: DMD eigenvalues and modes. (**a**) shows eigenspectrum $\{\lambda_j\}_{j=1,\cdots}$ in the left and powers, defined as $\{|\lambda_j^p|\|\alpha_j\phi_j\|_F\}_{j=1,\cdots}$, in the right. The first two DMD modes of highest powers are enumerated as #1 and #2. (**b**) and (**c**) show the phase and magnitude of the #1 and #2 DMD modes, respectively. In the phase diagram, regions, whose phases are similar are gathered. (**d**) is the time series for weekly cumulative cases for some selected regions.

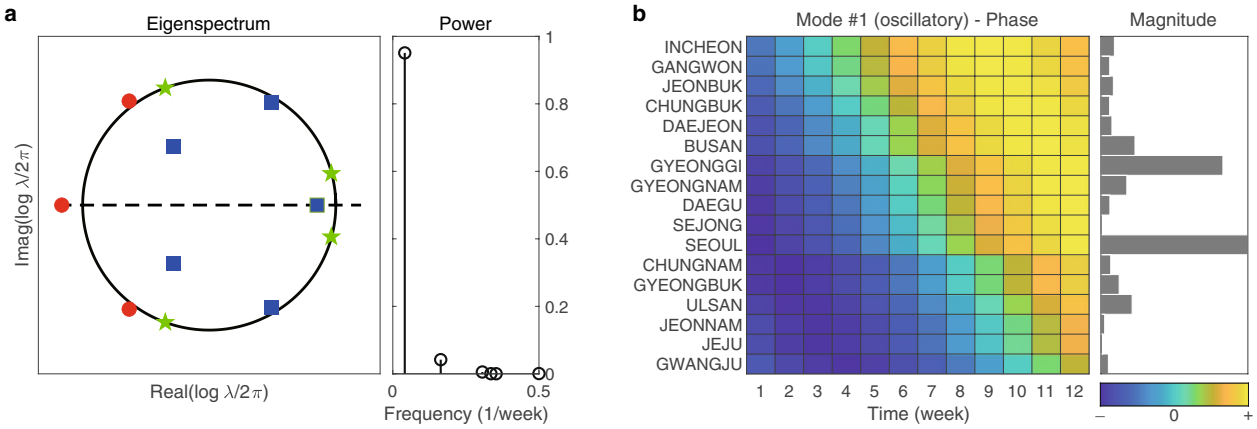


Figure 6. After the third wave: DMD eigenvalues and modes. (**a**) shows eigenspectrum $\{\lambda_j\}_{j=1,\cdots}$ in the left and powers, defined as $\{|\lambda_j^p|\|\alpha_j\phi_j\|_F\}_{j=1,\cdots}$, in the right. The first two DMD modes of highest powers are enumerated as #1 and #2. In this wave, the only one DMD mode of the dominant power is selected. (**b**) shows the phase and magnitude of the selected DMD mode. In the phase diagram, regions, whose phases are similar are gathered.

Time dependent spatial variation of COVID-19 in South Korea. In this section, we further investigate the data to quantify the time-dependent spatial variation of COVID-19 in South Korea over the period of interest.

We investigate the rate of incidence per 100,000 people in each region for the first wave, the second wave, the third wave, and the period after the third wave and plot this in Fig. 7a–d, respectively. This shows that the regional variation in weekly incidence is gradually decreasing over time. We observe that in the first wave (see

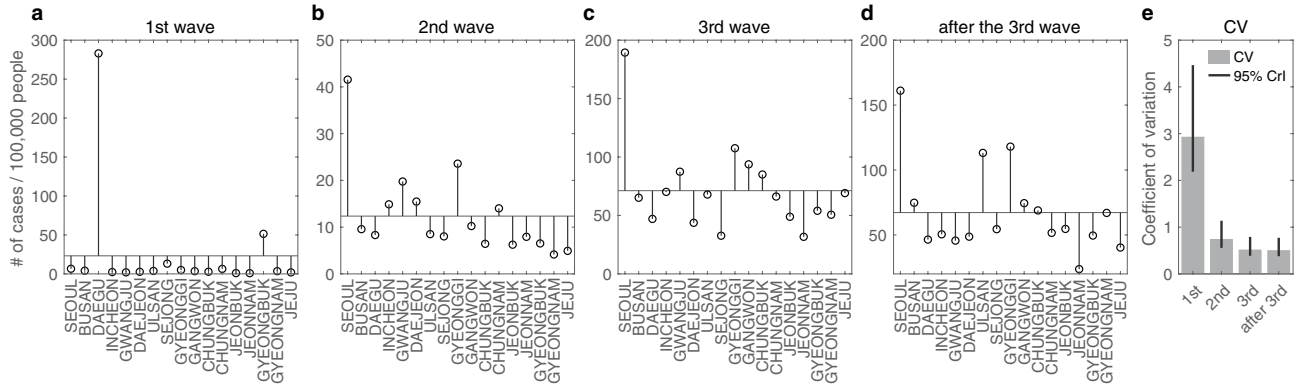


Figure 7. Regional variation in infection rate relative to the average rate. (**a**–**d**) show the rate of incidence per 100,000 people relative to the average rate for the first wave, second wave, third wave and after the third wave, respectively. The only incidence rates of Daegu and Gyeongbuk are shown to be higher than average. After the first wave, the incidence rates of Daegu and Gyeongbuk decreased to be under the average. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean is plotted in (**e**). Grey bar and black vertical line represent the CV of each period and its 95% credible interval (CrI), respectively. The CV estimated for each period is found to be 2.93 (95% CrI: 2.19–4.47), 0.75 (0.56–1.14), 0.52 (0.39–0.79), and 0.51 (0.38–0.77), respectively. This result shows that regional variation in the rate of incidence per 100,000 population becomes gradually uniform over time.

Fig. 7a), only the rate of incidence for Daegu and Gyeongbuk is shown to be higher than average. After the first wave, the rate of incidence for Daegu and Gyeongbuk becomes below the average, whereas that of Seoul and Gyeonggi stays higher than the average. Even if it is not definitely clear, as time proceeds, the regional differences seem to get smaller. To quantify this observation on the time-dependent regional difference in the incidence rate, we compute so-called the coefficient of variation (CV) for the rate of incidence per 100,000 people. The CV is defined by the ratio of the standard deviation to the mean[39]. This is a dimensionless number that can be used to compare the dispersion of groups with different means or different units. Similar to the standard deviation, the larger the CV is, the more over-dispersed the data will be. The computed CV is presented in Fig. 7e, in which we find that the CV decreases in time. More precisely, we have 2.93 CV (95% credible interval (CrI): 2.19–4.47 CV) for the first wave, 0.75 CV (95% CrI: 0.56–1.14 CV) for the second wave, 0.52 CV (95% CrI: 0.39–0.79 CV) for the third wave, and 0.51 CV (95% CrI: 0.38–0.77 CV) for the period after the third wave. This result clearly demonstrates that the first drastic reduction in CV occurred during the second wave, and the regional variation of weekly incidence tends to decrease over time. Namely, the spatiotemporal incidence pattern tends to be homogeneous, thereby indicating that the local outbreaks are dominant in most of the regions for the period after the third wave.

Novel compatible window-wise dynamic mode decomposition. Our data analysis using CwDMD has clearly shown the usefulness of the method to identify patterns of the spatially and temporally correlated nonlinear data. It is shown as well that some hidden patterns could be identified. The standard DMD, however, has a limitation in that it may provide misleading analysis generally for the inconsistent data[32]. The inconsistent data, equivalent to the nonlinear data can be interpreted as the data in which spatial resolution, the amount of spatial detail is given incompatible with the temporal resolution, the amount of temporal detail. Precise condition for the compatibility is obtained in section for "Methods". In Fig. 8 we have considered the COVID-19 time series data collected in a total of 17 regions. The standard DMD operator is shown to be able to fit the data perfectly in case a total of 18 or smaller temporal data is selected. The number 18 is the maximal time resolution for which the compatibility between spatial and temporal resolutions is valid. As the temporal resolution increases, the data fitting quality by DMD deteriorates significantly. This is unequivocally interpreted that DMD is inadequate to provide meaningful data analysis for these cases. To quantify the inadequacy, we investigate the phase and magnitude analysis from the selected DMD mode. For 19-week data from December 27, 2020–May 8, 2021, there is an evident disagreement between the COVID-19 data (black solid) and the DMD output (orange bar). The actual data indicates that the number of confirmed cases is higher in Gyeonggi and Seoul and it is relatively lower in Ulsan. However, DMD data analysis indicates otherwise that the number of confirmed cases in Ulsan is higher than in Seoul. This implies that the selected DMD mode does not represent the data pattern adequately. Thus, the direct and reliable DMD analysis of large time-series data is concluded not to be feasible unless it is linear.

We, therefore, arrive at the need of introducing a novel compatible window-wise dynamic mode decomposition. The main issue in DMD for large time-series lies in the nonlinearity of the data. The point of CwDMD is that for any given nonlinear data, it is proven to be possible to select an adequate set of representative subdomains called windows, each containing moderate-sized linear data. For example, Fig. 2a, shows specially chosen windows for COVID-19 data in South Korea we analyze. The total size-times duration of all the windows serving a given system depends only on local situations that can arise in the full-time series data. We then apply the standard DMD for each window. This strategy is called the compatible window-wise dynamic mode decomposition

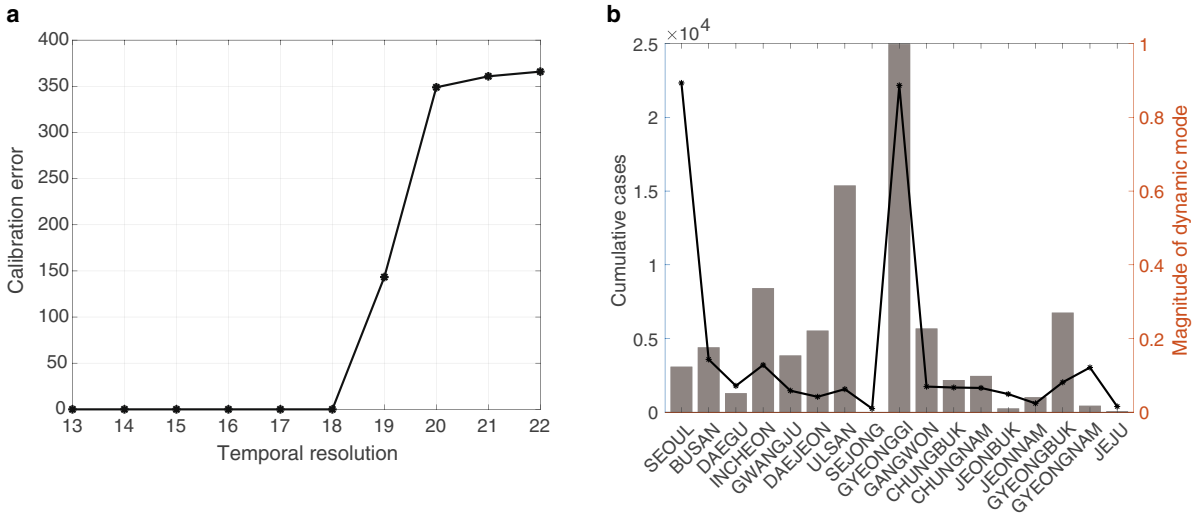


Figure 8. Results showing the inadequacy of the standard DMD applied to incompatible data. (**a**) shows the calibration error as a function of temporal resolutions. The spatial resolution is fixed as 17 and we see that as the temporal resolution becomes larger than 18, the calibration errors start to increase. In (**b**), we consider the total of 19 weeks' time series data from Dec. 27, 2020–May 8, 2021, which is incompatible with 17 spatial resolution. We selected DMD mode consistently and analyzed its magnitude. Clearly, (**b**) shows that the magnitude does not adequately represent that of data. Recall Figs. 5 and 6 are for the compatible windows, in which such erroneous result do not occur.

(CwDMD). Basically, CwDMD is a collection of DMD for a specially selected set of consistent windows. In each window, we choose the most significant DMD modes, and the reconstructed data in its dimension, from the selected DMD modes are constructed and investigated to understand the actual data.

Discussion

In this study, we have developed a novel data-driven framework: *compatible window-wise dynamic mode decomposition* (CwDMD). Using the CwDMD, we have identified the spatiotemporal transmission patterns of COVID-19 in South Korea from January 20, 2020 to May 10, 2021. It is generally very challenging to uncover COVID-19 transmission dynamics since there exists a complex interplay among various time-varying factors such as virus, human, mobility, socio-economic infrastructures, and public health policies. However, our CwDMD analysis successfully elucidates how spatial correlations among 17 regions evolve in the presence of such complex features.

The first wave was focused on the Daegu and Gyeongbuk area, which was mainly caused by the superspreading events from the Shincheonji Church-related clusters[12,17]. It spread to several regions nearby, but this was quickly contained. This was due to aggressive interventions such as drive-through or walk-through rapid PCR testing, contact tracing, isolation, mask distribution, and social distancing[40]. Most of all, the behavior and awareness of South Koreans were the most crucial reasons for great success. As a result, the substantially largest outbreaks from the Shincheonji Church-related clusters did not last for more than a month. Our analysis also confirmed that the local outbreaks were kept in the Daegu and Gyeongbuk areas. Towards the end and after the first wave, a few major large-scale outbreaks occurred in the metropolitan regions including Seoul, Gyeonggi, and Incheon. For example, in Seoul, there were a few sporadic large outbreaks, which include the Guro-call center, Itaewon Club, and Richway (Seoul-based health product retailer) between March 2020 and June 2020.

Execution of online school, which was initiated in the middle of the first wave, and ongoing intensive interventions contributed to maintaining a low level of COVID-19 outbreaks nationwide until the rally held in Seoul. On August 15, 2020, the rally led by SarangGeil Church caused 641 cases in Seoul and thus initiated the second wave. Our phase analysis for the second wave captured that COVID-19 spread rapidly throughout the nation. This is linked to the fact that people participated in the rally and returned to their home regions in a few days[38], which arose local outbreaks in every region as well.

The local outbreaks became dominant compared to inter-regional outbreaks during the winter season from November 2020 to February 2021. The large-scale local spread of COVID-19 led to the third wave with the largest cumulative cases nationwide. Since the outbreaks were significantly severe in the third wave and the majority of the cases were focused on the metropolitan area, region-specific public health policies were first implemented and risk assessment level for social distancing interventions was refined from Level 3 to Level 5, as of November 7, 2020[41]. Moreover, region-specific restrictions of large gatherings, such as prohibiting gatherings of more than four people and closing shops after 10 pm, have been imposed during the third wave[42] as well. Additionally, a special quarantine period was imposed on Thanksgiving and the New Year's holidays nationwide. These strict interventions combined with vaccination have slowed down the third wave from February 2021. Vaccination started from February 26, 2021, with a slow rate at the early stage; 7.1% of the primary dose; 1.1% of the second

dose, as of May 10, 2021[2]. COVID-19 has then been maintained without major outbreaks for more than four months after the third wave, between February 2, 2021 to May 10, 2021.

Overall, cumulative cases and deaths of COVID-19 in South Korea seem not that large compared to those of other countries with similar population densities, and the duration of each wave seems not too long either. For example, as of July 9, 2021, a total of 814,533 cumulative cases and 14,933 deaths of COVID-19 in Japan were reported while a total of 165,344 cumulative cases and 2036 deaths in South Korea[4] were reported. Japan's vaccination rate (2.1 % of the primary dose and 1.0% of the second dose as of May 10, 2021) and population density ($337/km^2$) are similar to those of South Korea. However, the fourth big wave occurred in Japan, from March 2021 to May 2021 with a maximum daily number of confirmed cases of more than 6000. This can be associated with the fact that Japan imposes voluntary social distancing policy, while South Korea continues to enforce compulsory social distancing policies even after the third wave. Japan has invoked a number of COVID-19 State of Emergencies, but compulsory policies such as forced suspension or lockdown was not imposed[43]. On the other hand, policies in South Korea such as prohibiting gatherings of more than four people and closing shops after 10 p.m. forcibly prevent further infections from occurring. It is worth mentioning that there are data-related issues in this study. First, the official (reported) data could be different from the real ones due to the selective biases of various kinds[5,6]. Next, other factors such as temperatures, seasonality, UV radiation, pollution, etc.[44–46] are not included in the analysis.

South Korea is one of the most successful countries for mitigating and preventing the COVID-19 pandemic. Since South Korea has learned a valuable lesson from the MERS-CoV outbreak, which was the largest outbreak originated from the Middle Eastern countries in 2015, various preparedness plans have been initiated for emerging infectious diseases including medical infrastructure and transparent data disclosure through daily briefings[47]. Real-time infection transmission notification through mobile phone applications or websites, and a real-time alarm system through mobile phone (including location-specific risk notification through GPS) have been newly developed during the COVID-19 pandemic. In addition, South Koreans were quickly alert and carried out voluntary preventing activities such as wearing a mask and prohibiting gatherings. With such an ensemble of national infrastructure and citizens' voluntary participation in quarantine, South Korea demonstrates its superiority in handling COVID-19 outbreaks through successful mitigation strategies.

DMD has been successful to extract spatial–temporal coherent patterns in a specific form of periodic, growing, and decaying dynamical spectrum decomposition[34]. On the other hand, it is shown that balance between spatial and temporal resolutions has to be taken into account since otherwise, DMD mode analysis can result in erroneous data interpretation for highly nonlinear time series data. This balance is mathematically identified as the linearity of data in this paper, which means that DMD can in general make sense only for the appropriate selection of windows from the full temporal data sets so that spatial resolution is larger than the temporal resolution. This clearly generates the limitation of the use of classical DMD and/or its variants[22,25] since oftentimes it is useful to extract spatiotemporal patterns for rather long data sets. To overcome this issue, one can select a special set of the time series data with certain labels as discussed in[48] or more generally, one can use a certain multiscale temporal representation of the data. Namely, one can decompose the temporal steps, from fine to coarse so that in coarse level, the global data makes the linearity, while the fine-scale is handled only in several local windows. Somewhat similar but different idea, named as multiresolution DMD can be found at[49]. Overall, a systematic method or mathematical modeling for forecasting COVID-19 data is an open and challenging issue. The multiscale approach briefly described above is potentially useful to generate the prediction operator. Lastly, if we can identify the data related to external controls and interventions to stop spreading COVID-19, then we may be able to apply DMD with control, presented in[50] for analysis, which is yet to be investigated.

Methods

Compatible window-wise dynamic mode decomposition (CwDMD). In this section, we shall describe the compatible window-wise Dynamic Mode Decomposition (CwDMD), a novel dynamic mode decomposition method that respects the compatibility of the data set. A detailed statement of compatibility will be presented as well. Basically, we present a new observation that the consistent data is a linear data and suggest that DMD has to be applied for the consistent or linear data. A compatibility condition is a way of achieving this consistency or linearity of the data set. We shall show that certain windows of the given time series data has to be selected so that a balance between the spatial and temporal resolution of the data set is made. This balance will then lead to the linearity of the selected windows. The application of DMD for each window is shown to result in accurate data analysis.

Throughout this section, for the sake of convenience, we denote $\mathbb{C}^{n\times\ell}$ by the space of complex matrices of size $n\times\ell$. For $n=1$ or $\ell=1$, we shall omit writing it. Namely, for $\ell=1$, we set $\mathbb{C}^n:=\mathbb{C}^{n\times 1}$, that of which is sets of complex vectors of size n. For any element $c\in\mathbb{C}$, we shall denote $\bar{c}$ by its complex conjugate. We shall denote $\underset{\sim}{\cdot}$ by the vector and $\underset{\approx}{\cdot}$ by the tensor. For $\underset{\approx}{M}\in\mathbb{C}^{n\times\ell}$, its null and range will be denoted by $\mathcal{N}(\underset{\approx}{M})$ and $\mathcal{R}(\underset{\approx}{M})$, respectively. We denote $\underset{\approx}{M}^*$ by its complex adjoint matrix, and also denote $\underset{\approx}{M}^\dagger$ by the pseudoinverse of $\underset{\approx}{M}$. The symbol $\underset{\approx}{\delta}$ denotes the identity matrix. Note that $\underset{\approx}{M}^\dagger$ satisfies the following conditions:

$$\underset{\approx}{M}\,\underset{\approx}{M}^\dagger\,\underset{\approx}{M}=\underset{\approx}{M},\quad \underset{\approx}{M}^\dagger\,\underset{\approx}{M}\,\underset{\approx}{M}^\dagger=\underset{\approx}{M}^\dagger,\quad (\underset{\approx}{M}\,\underset{\approx}{M}^\dagger)^*=\underset{\approx}{M}\,\underset{\approx}{M}^\dagger,\quad\text{and}\quad (\underset{\approx}{M}^\dagger\,\underset{\approx}{M})^*=\underset{\approx}{M}^\dagger\,\underset{\approx}{M}.$$

In particular, if $\underset{\approx}{M}$ has a linearly independent columns, it holds that $\underset{\approx}{M}^\dagger=(\underset{\approx}{M}^\dagger\,\underset{\approx}{M})^{-1}\,\underset{\approx}{M}^*$.

Dynamic mode decomposition (DMD). Given a data set in a form of a time series data as follows:

$$\underset{\approx}{T} = \{\underset{\sim}{u}_0, \underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_{m-1}, \underset{\sim}{u}_m\} \in \mathbb{C}^{n\times(m+1)},$$

where $\underset{\sim}{u}_k$ stands for the kth snapshot of the data set for $k \geq 0$ with $m+1$ being the last entry of the data set, we let $\underset{\approx}{X}$ and $\underset{\approx}{Y}$ denote the followings:

$$\underset{\approx}{X} = \{\underset{\sim}{u}_0, \underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_{m-1}\} \quad \text{and} \quad \underset{\approx}{Y} = \{\underset{\sim}{u}_1, \underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_m\}.$$

We shall briefly review the general description of the dynamic mode decomposition (DMD) applied for $\underset{\approx}{T}$. For clarity, we assume an ordered sequence of data separated by a constant sampling time Δt. The idea of DMD lies at the assumption that there exists a linear operator $\underset{\approx}{A}$ that connects at least, approximately two data $\underset{\sim}{u}_k$ and its subsequent data $\underset{\sim}{u}_{k+1}$ for all $k \geq 0$, that is

$$\underset{\sim}{u}_{k+1} \approx \underset{\approx}{A}\,\underset{\sim}{u}_k, \quad \forall k \geq 0 \quad \text{equivalently} \quad \underset{\approx}{Y} \approx \underset{\approx}{A}\,\underset{\approx}{X}. \tag{1}$$

The ambiguity in the approximation $\approx$ will be clarified by defining $\underset{\approx}{A} = \underset{\approx}{Y}\,\underset{\approx}{X}^{\dagger}$ or as the solution to the following optimization problem:

$$\underset{\approx}{A} = \arg\min_{\underset{\approx}{C}} \| \underset{\approx}{Y} - \underset{\approx}{C}\,\underset{\approx}{X} \|_F, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm. We note that the operator $\underset{\approx}{A}$ is a type of dynamic operator that relates two consecutive data set. The goal of the dynamic mode decomposition is to extract the dynamic characteristic of $\underset{\approx}{A}$, not directly to construct the mapping $\underset{\approx}{A}$. More precisely, DMD obtains spectrums or spatial–temporal characteristics of the dynamical process described by $\underset{\approx}{A}$. We note that the spectrums can be used to completely construct the action of the operator $\underset{\approx}{A}$ if needs arise.

The essential algorithmic background lies in singular value decomposition of data, $\underset{\approx}{X}$ and the relationship between eigen-pairs of $\underset{\approx}{A}$ and its representation in principal component modes (see Lemma 1 and Lemma 2, in Supplementary note for Method). These are used to obtain the standard dynamic mode decomposition algorithm, as provided in Algorithm 1[51].

Algorithm 1 Dynamic mode Decomposition

With the expression of $\underset{\approx}{A} = \underset{\approx}{Y}\underset{\approx}{X}^{\dagger} \in \mathbb{C}^{n\times n}$,

1: Apply the singular value decomposition of $\underset{\approx}{X}$ to obtain:

$$\underset{\approx}{X} = \underset{\approx}{U}\underset{\approx}{\Sigma}\underset{\approx}{V}^*.$$

2: Consider the operator $\underset{\approx}{A}$ in the principal component modes[51]:

$$\widehat{\underset{\approx}{A}} := \underset{\approx}{U}^*\underset{\approx}{A}\underset{\approx}{U} = \underset{\approx}{U}^*\underset{\approx}{Y}\underset{\approx}{X}^{\dagger}\underset{\approx}{U} = \underset{\approx}{U}^*\underset{\approx}{Y}\underset{\approx}{V}\underset{\approx}{\Sigma}^{-1}\underset{\approx}{U}^*\underset{\approx}{U} = \underset{\approx}{U}^*\underset{\approx}{Y}\underset{\approx}{V}\underset{\approx}{\Sigma}^{-1}.$$

3: Find the eigen-pairs $(\lambda_i, \underset{\sim}{w}_i)_{i=1,\cdots,n}$ of the operator $\widehat{\underset{\approx}{A}}$, and define the DMD modes $\underset{\approx}{\Phi} = [\underset{\sim}{\phi}_1, \cdots, \underset{\sim}{\phi}_n]$ as follows:

$$\widehat{\underset{\approx}{A}}\,\underset{\sim}{w}_i = \lambda_i \underset{\sim}{w}_i \quad \text{and} \quad \underset{\sim}{\phi}_i = \underset{\approx}{U}\underset{\sim}{w}_i, \quad \forall i = 1, \cdots, n.$$

Generally, the data analysis can be accomplished through the dynamic modes and eigenvalues, as given as $(\lambda_i, \underset{\sim}{\phi}_i)_{i=1,\cdots,n}$. We remark that $\{\underset{\sim}{\phi}_i\}_{i=1,\cdots,n}$'s are called the DMD modes or mode vectors and they provide a rich set of information, especially spatial information about the data set[25]. For example, the modulus of the element of the mode vector provides measure of the spatial region's participation for that mode. On the other hand, the eigenvalues $\{\lambda_i\}_{i=1,\cdots,n}$ are relevant to the time evolution of the data sets and thus, they contain temporal information.

Linearity, consistency, and CwDMD. A loophole in DMD lies in that DMD spectrums are found for an approximate dynamic operator $\underset{\approx}{A}$ for the data set $\underset{\approx}{T}$. It is very much ambiguous and completely unknown theoretically how much the error observed in Eq. (1) results in misleading data interpretation from DMD spectrums. This has been elaborated in Fig. 8 for further clarity. The desired DMD is then not to start with constructing DMD-spectrums for $\underset{\approx}{A}$ that satisfies (1), but, to build DMD spectrums based on $\underset{\approx}{A}$ that satisfies the following relationship:

$$\underset{\sim}{u}_{k+1} = \underset{\approx}{A}\,\underset{\sim}{u}_{k}, \quad \forall 0 \le k \le m, \quad \text{equivalently} \quad \underset{\approx}{Y} = \underset{\approx}{A}\,\underset{\approx}{X}. \tag{3}$$

Thus, we investigate the condition for the existence of an operator A that satisfies the Eq. (3). This is in fact dependent on the data set T. Namely, there must be a condition for $\underset{\approx}{T}$, which leads to the existence of such an operator A. Therefore, we introduce a notion of the linearity. Basically, we say that the data T is linear if and only if there exists an operator $\underset{\approx}{A} \in \mathbb{C}^{n\times n}$ such that $\underset{\approx}{Y} = \underset{\approx}{A}\,\underset{\approx}{X}$ (see the notion of linearity precisely defined for T in Definition 1 of Supplementary note). The compatibility condition is basically the condition for which the data $\underset{\approx}{T}$ is linear. We remark that a relevant notion that states the Eq. (3) for a particular $\underset{\approx}{A}$ of the form $\underset{\approx}{A} = \underset{\approx}{Y}\,\underset{\approx}{X}^{\dagger}$ for the data $\underset{\approx}{T}$ has been provided by Tu et al. in[32], i.e., a notion of linear consistency, stating that the null space of $\underset{\approx}{X}$ is contained in that of $\underset{\approx}{Y}$ ($\mathcal{N}(\underset{\approx}{X}) \subset \mathcal{N}(\underset{\approx}{Y})$) (see the notion of linear consistency defined for T in Definition 2 and also Theorem 1 of Supplementary note). We remark that the linearity is much more intuitive and general than the linear consistency. The notion of the linearity is a certain extension of the existence of line connecting two points in two dimensional Euclidean space consisting of one spatial dimension and one temporal dimension. On the other hand, we observe that these two concepts; linearity and linear consistency are in fact equivalent. Namely, the linear consistency of T holds if and only if the linearity of T holds (see Theorem 2 in Supplementary note for detailed proof). In another words, nonlinear data is inconsistent and inconsistent data is nonlinear. This equivalency is remarkable since these two concepts can be used to derive so-called the compatibility condition, which can be used to easily verify the linearity of T. Note that the linear consistency condition provides an important algebraic condition for the data being linear. However, authors find it difficult to verify that condition in general.

The concept of compatibility is based on the observation that the data T being linear is relevant to the balance between spatial and temporal resolutions. As mentioned, for example, in one spatial dimension, only two points (two temporal data) can be connected in general by a line, unless data consisting of more than two points are collinear. Its extension for higher dimensional case can be understood as a simple inequality: $m \le n$. More precisely, the compatibility condition can be stated as follows:

Definition (*Compatibility Condition*) Compatibility condition is the balance between to the balance between temporal and spatial resolutions, i.e., a data set $\underset{\approx}{T}$ with the temporal resolution $m+1$ and spatial resolution n have the relationship that $m \le n$.

Note that for $m > n$, $\underset{\approx}{T}$ will be in general inconsistent unless it is linear. The compatibility condition is stated to cover very general situations for which DMD can have a meaningful usage. We can show that under the compatibility condition, DMD will provide meaningful results with probability one. To be more precise, we note that the consistency can be easily understood in terms of the linear independency of the data X, i.e., the linear independency of X implies the consistency of T and this can in particular, remove the trivial case that any column of X is the zero vector. Theoretically, it is established that if $\underset{\approx}{T}$ satisfies the compatibility condition, then almost all $\underset{\approx}{X} \in \mathbb{C}^{n\times m}$ with $m \le n$ will consist of columns which are linearly independent[52,53]. This means that $\mathcal{N}(\underset{\approx}{X}) = \{\underset{\approx}{0}\}$. Therefore, the data set T is linear. The compatibility condition thus implies the consistency with probability one. Thus, the compatibility condition implies that the linearity of the data $\underset{\approx}{T}$ is almost always guaranteed in case $m \le n$, which then leads to the meaningful DMD results.

In a very much rare case, when the consistency breaks under the compatibility condition, one can provide a small (arbitrarily small) perturbation to obtain $T_\varepsilon \in \mathbb{C}^{n\times(m+1)}$, which is proven to result in a linear data[54]. Namely, for $m \le n$, let $X_\varepsilon \in \mathbb{C}^{n\times m}$ consist of first m columns of T_ε. Then we consider $\widetilde{X_\varepsilon} \in \mathbb{C}^{m\times m}$ obtained from X_ε by chopping off all rows underneath mth row of X_ε. This square matrix can be proven to be diagonalizable[52,54], i.e., it consists of linear independent columns and thus the columns of X_ε is linearly independent. In view of the spatio-temporal analysis of the data, arbitrarily small perturbation will not change the result significantly. Furthermore, theoretically, such arbitrarily small perturbation will not affect the computation of the DMD-spectrums if they are in particular, Gaussian[55,56]. We remark that our data is generally very nice, i.e., whenever we choose $m \le n$, the data set T is always linear consistent and so, no perturbation was needed.

We are in a position to introduce our new algorithm, so-called a compatible window-wise dynamic mode decomposition (CwDMD). Our observation is that for $m > n$, $\underset{\approx}{T}$ will be in general inconsistent unless it is linear. As such, the direct and reliable DMD analysis of large time series data is not feasible in general. The strategy is to choose an adequate set of representative subdomains called windows, each containing a moderate size of time-series data that satisfies the compatibility. The total size-times duration of all the windows serving a given

system depends only on local situations that can arise in the full time series data. For example, Fig. 2, A shows a class of windows for the COVID-19 data in South Korea. Namely, given a data set $\{\underset{\sim}{u}_0, \underset{\sim}{u}_1, \ldots, \underset{\sim}{u}_k, \ldots, \underset{\sim}{u}_m\}$, we consider the following windows that are consistent:

$$(\underset{\approx}{X}_k, \underset{\approx}{Y}_k), \text{ with } \underset{\approx}{X}_k := \{\underset{\sim}{u}_{k_s}, \ldots, \underset{\sim}{u}_{k_{e-1}}\} \quad \text{and} \quad \underset{\sim}{Y}_k := \{\underset{\sim}{u}_{k_s+1}, \ldots, \underset{\sim}{u}_{k_e}\}.$$

for which $\underset{\approx}{X}_k$ and $\underset{\approx}{Y}_k$ are consistent for $k = 0, 1, \ldots, \ell$. The compatible window-wise dynamic mode decomposition is to apply the dynamic mode decomposition locally for each compatible window $(\underset{\approx}{X}_k, \underset{\approx}{Y}_k)$. Note that these windows can be constructed so that they may overlap or non-overlap depending on the situations. Therefore, choices of window can be made without too much restriction other than the condition of compatibility. This can be summarized as in the Algorithm 2.

Algorithm 2 Compatible Window-wise Dynamic mode Decomposition

Given a data set $\underset{\approx}{T} = \{\underset{\sim}{u}_0, \underset{\sim}{u}_1, \cdots, \underset{\sim}{u}_{m-1}, \underset{\sim}{u}_m\}$,

1: We generate the consistent data sets $\bigcup_{\nu=0,1,\cdots,\ell}(\underset{\approx}{X}_\nu, \underset{\approx}{Y}_\nu)$.

2: For $\nu = 0, 1, \cdots$, set $\underset{\approx}{X} = \underset{\approx}{X}_\nu$ and $\underset{\approx}{Y} = \underset{\approx}{Y}_\nu$, $m = m_\nu$.

3: We apply the Algorithm 1 to obtain $(\lambda_i^\nu, \phi_i^\nu)_{i=1,\cdots,n}^{\nu=0,1,\cdots,\ell}$.

Data fitting, dimensional reduction, frequency and phase analysis. In this section, we discuss the data fitting using the DMD operator and choice of modes for the dimensional reduction and their uses for the phase analysis of each window. Throughout this section, we assume that $\underset{\approx}{T} \in \mathbb{C}^{n\times(m+1)}$ is consistent and the DMD operator $\underset{\approx}{A}$ is given in terms of eigen-pairs $(\lambda_i, \underset{\sim}{\phi}_i)_{i=1,\cdots,n}$. We would also like to mention that the precise action of the operator $\underset{\approx}{A}$ may not be found solely from these eigenspectrums. Namely, the data $\underset{\approx}{X}$ has to be represented in terms of DMD modes, which requires to solve certain optimization problem. In a prior work, this has been accomplished by taking into account the whole data $\underset{\approx}{X}$. We shall show that this can be done taking into account any single snapshot data in $\underset{\approx}{X}$ under the consistency condition, thereby achieving a significant computational reduction. We begin our discussion with the fact that almost all complex matrices over complex fields are diagonalizable[52,54]. Namely, geometric and algebraic multiplicities of almost all complex matrices over complex fields are identical. This means that the DMD modes make a full set of eigenvectors for almost all data set satisfying the compatibility. Some list of a couple of equivalent conditions to the fact that algebraic and geometric multiplicities agree for a matrix $\underset{\approx}{A} \in \mathbb{C}^{n\times n}$ can be found at[57] and Theorem 3 in Supplementary note. Therefore, in general, we have that $\mathbb{C}^n = \text{span}\{\underset{\sim}{\phi}_i\}_{i=1,\cdots,n}$. Having a full set of eigenvectors of $\underset{\approx}{A}$, we can represent for example, the data $\underset{\sim}{u}_\eta$ of $\underset{\approx}{T}$ with $0 \le \eta \le m+1$, as follows:

$$\underset{\sim}{u}_\eta = \sum_{i=1}^{n} \alpha_i \underset{\sim}{\phi}_i \quad \text{or} \quad \underset{\sim}{\alpha} = \underset{\approx}{\Phi}^{-1} \underset{\sim}{u}_\eta,$$

where $\Phi = [\underset{\sim}{\phi}_1 \ \ldots \ \underset{\sim}{\phi}_n]$. With $\underset{\sim}{\alpha}$ given above, we can obtain the action of the DMD operator $\underset{\approx}{A}$ as follows: for $-\eta \le \tilde{k} \le -\eta + m + 1$,

$$\underset{\sim}{u}_k = \sum_{i=1}^{n} \alpha_i e^{k\,\Re(\log(\lambda_i))} e^{\hat{i}k\Im(\log(\lambda_i))} \underset{\sim}{\phi}_i, \tag{4}$$

where $\hat{i}$ is the pure imaginary number such that $\hat{i}^2 = -1$. We remark that it is standard to choose $\eta = 0$, which is also our choice. Oftentimes DMD is argued to be biased to the initial data[24], our observation is that it is not really the case, for the consistent data. We recall that the framework of the optimized DMD[22] is also designed to obtain the same $\underset{\sim}{\alpha}$ for fitting, $\underset{\approx}{X}$, by solving the following optimization problem:

$$\underset{\sim}{\alpha} = \underset{\underset{\sim}{\mu}=(\mu_i)_{i=1,\cdots,n}}{\arg\min} \left\| \underset{\approx}{X} - \underset{\approx}{\Phi}\, \underset{\approx}{D_\mu}\, \underset{\approx}{V_{m-1}} \right\|_F,$$

where

$$\underset{\approx}{D_\mu} = \mathrm{diag}(\mu) \quad \text{and} \quad \underset{\approx}{V_m} = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^m \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^m. \end{pmatrix}$$

It is clear that the consistency of data leads to a significant reduction of the computational effort.

We now can consider a discrete to continuous extension of the action of DMD operator. We remark that from the discrete represent of $\underset{\sim}{u}_k$ in (4), a continuous extension can be achieved as follows: for all $t \geq t_0 = 0$,

$$\underset{\sim}{u}(t) := \sum_{i=1}^{n} \alpha_i (\lambda_i)^{t-t_0} \underset{\sim}{\phi}_i = e^{(t-t_0)\,\Re(\log(\lambda_i))} e^{\hat{i}(t-t_0)\Im(\log(\lambda_i))} \underset{\sim}{\phi}_i . \tag{5}$$

We now discuss the mode choice for the phase analysis, which will be used to obtain the dimensional reduction of the data. The most natural guide to choose the important DMD mode is to find the DMD mode which contributes most significantly to the data both temporally and spatially. This leads us to choose the index of DMD mode for which the following quantity, product of the temporal and spatial contribution in each window is maximized:

$$\arg\left\{ \max_k \{ |\lambda_k|^p \| \alpha_k \underset{\sim}{\phi}_k \|_F, 1 \leq k \leq n. \} \right\}, \tag{6}$$

where p is the temporal resolutions for the window. We call the quantity $|\lambda_k|^p \| \alpha_k \underset{\sim}{\phi}_k \|_F$ the power of the kth DMD mode and observe that in general one or two dominant powers exist. These are then chosen to form a dimensionally reduced data. For example, $\underset{\sim}{\phi}_k$ is the DMD mode whose power is the largest. Then it is used to form a dimensionally reduced data: for all $t \geq t_0 = 0$,

$$\underset{\sim}{\widetilde{u}}(t) = \alpha_k (\lambda_k)^{t-t_0} \underset{\sim}{\phi}_k = e^{(t-t_0)\,\Re(\log(\lambda_k))} e^{\hat{i}(t-t_0)\Im(\log(\lambda_k))} \underset{\sim}{\phi}_k , \tag{7}$$

which is used for the data interpretation such as phases and magnitudes. In literature, DMD modes are chosen based on their norms or weighted norm by the corresponding DMD eigenvalues[32]. For example, the use of weighted norm by DMD eigenvalues, can be interpreted as to penalize spurious modes with large norms but quickly decaying contributions to the dynamics[29]. In our choice, we incorporate $\underset{\sim}{\alpha}$, the coordinate of data in the frame of DMD modes as a special scale for DMD modes. These measurements are meaningful especially for highly nonlinear data, since coordinates given in terms of DMD modes can much affect the dynamics of data. We remark that the frequency of the solution for the mode k, can be defined through $\Im(\log(\lambda_k))/2\pi$ and thus the period is given by the reciprocal of the frequency. The identified DMD mode can be categorized as periodic, growing or decaying modes depending on the magnitude of λ_k. Namely, for eigenvalues on (or close), outside or inside the unit circle, the corresponding modes are considered as oscillatory, growing, and decaying modes, respectively. In the present work, we give a tolerance $\epsilon = 5.E{-}2$ and denote $N_o = \{i : ||\lambda_i| - 1| \leq \epsilon\}$, $N_g = \{i : |\lambda_i| > 1 + \epsilon\}$, $N_d = \{i : |\lambda_i| < 1 - \epsilon\}$ by the set of oscillatory modes, the set of growing modes, and the set of decaying modes, respectively. We first select the DMD modes of large powers, and then measure the magnitude of its eigenvalues and determine whether they are oscillatory, growing or decaying mode.

Data availability

The map of South Korea was obtained in the form of a shapefile from the website of the Statistical Geographic Information Service (panel A of Fig. 2)[58]. Population, area, and density by the 17 regions as of April 9, 2021, were obtained from the website of e-indicatior in South Korea (panel B of Fig. 2)[59]. The daily incidence of COVID-19 by regions from January 20, 2020, to May 10, 2021, was obtained on the website of Seoul National University Asia Regional Information Center (SNUARIC) (panel B of Fig. 2)[60].

Received: 16 July 2021; Accepted: 22 November 2021
Published online: 28 December 2021

References

1. World Health Organization. Timeline: WHO's COVID-19 response. http://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline (Accessed 30 Sept 2021).
2. Our World in Data. Coronavirus (COVID-19) vaccinations - statistics and research. http://ourworldindata.org/covid-vaccinations. (Accessed 21 June 2021).
3. Lopez Bernal, J. *et al.* Effectiveness of COVID-19 vaccines against the B. 1.617. 2 (Delta) variant. *N. Engl. J. Med.* 585–594 (2021).
4. World Health Organization. WHO coronavirus (COVID-19) dashboard, situation by region, country, territory & area. http://covid19.who.int/table. (Accessed 7 June 2021).
5. US Food and Drug Administration and others. SARS-CoV-2 viral mutations: impact on COVID-19 tests (2021).
6. Woloshin, S., Patel, N. & Kesselheim, A. S. False negative tests for SARS-CoV-2 infection—challenges and implications. *N. Engl. J. Med.* **383**, e38 (2020).
7. Alwan, N. A. Surveillance is underestimating the burden of the COVID-19 pandemic. *Lancet* **396**, e24 (2020).
8. Modi, C., Böhm, V., Ferraro, S., Stein, G. & Seljak, U. Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. *Nat. Commun.* **12**, 1–9 (2021).
9. Carroll, C. *et al.* Time dynamics of COVID-19. *Sci. Rep.* **10**, 1–14 (2020).
10. Castro, M. C. *et al.* Spatiotemporal pattern of COVID-19 spread in Brazil. *Science* **372**, 821–826 (2021).

11. Institute for Health Metrics and Evaluation COVID-19 Forecasting Team. Modeling COVID-19 scenarios for the United States. *Nat. Med.* **27**, 94 (2021).
12. Korea Disease Control and Prevention Agency. Confirmed cases in Korea (2021). http://ncov.mohw.go.kr.
13. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).
14. Lee, H. *et al.* Risk assessment of importation and local transmission of COVID-19 in South Korea: Statistical modeling approach. *JMIR Public Health Surveill.* **7**, e26784 (2021).
15. Feng, Y. *et al.* Spatiotemporal spread pattern of the COVID-19 cases in China. *PLoS ONE* **15**, e0244351 (2020).
16. Ghosh, P. & Cartone, A. A spatio-temporal analysis of COVID-19 outbreak in Italy. *Reg. Sci. Policy Pract.* **12**, 1047–1062 (2020).
17. Kim, S. & Castro, M. C. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of May 31, 2020). *Int. J. Infect. Dis.* **98**, 328–333 (2020).
18. Sartorius, B., Lawson, A. & Pullan, R. Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England. *Sci. Rep.* **11**, 1–11 (2021).
19. Wang, Y., Liu, Y., Struthers, J. & Lian, M. Spatiotemporal characteristics of the COVID-19 epidemic in the United States. *Clin. Infect. Dis.* **72**, 643–651 (2021).
20. Bag, R., Ghosh, M., Biswas, B. & Chatterjee, M. Understanding the spatio-temporal pattern of COVID-19 outbreak in India using GIS and India's response in managing the pandemic. *Reg. Sci. Policy Pract.* **12**, 1063–1103 (2020).
21. Schmid, P. J., Meyer, K. E. & Pust, O. Dynamic mode decomposition and proper orthogonal decomposition of flow in a lid-driven cylindrical cavity. In *8th International Symposium on Particle Image Velocimetry*, 25–28 (2009).
22. Jovanović, M. R., Schmid, P. J. & Nichols, J. W. Sparsity-promoting dynamic mode decomposition. *Phys. Fluids* **26**, 024103 (2014).
23. Erichson, N. B., Mathelin, L., Kutz, J. N. & Brunton, S. L. Randomized dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.* **18**, 1867–1891 (2019).
24. Azencot, O., Yin, W. & Bertozzi, A. Consistent dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.* **18**, 1565–1585 (2019).
25. Kutz, J. N., Brunton, S. L., Brunton, B. W. & Proctor, J. L. *Dynamic Mode Decomposition: Data-driven Modeling of Complex Systems* (SIAM, 2016).
26. Bistrian, D., Dimitriu, G. & Navon, I. Processing epidemiological data using dynamic mode decomposition method. In *AIP Conference Proceedings*, 080002 (AIP Publishing LLC, 2019).
27. Sato, R. C. Disease management with ARIMA model in time series. *Einstein* **11**, 128 (2013).
28. Bistrian, D., Dimitriu, G. & Navon, I. Modeling dynamic patterns from COVID-19 data using randomized dynamic mode decomposition in predictive mode and ARIMA. In *AIP Conference Proceedings*, 080002 (AIP Publishing LLC, 2020).
29. Proctor, J. L. & Eckhoff, P. A. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *Int. Health* **7**, 139–145 (2015).
30. Cervellin, G., Comelli, I. & Lippi, G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J. Epidemiol. Glob. Health* **7**, 185–189 (2017).
31. Rovetta, A. Reliability of Google Trends: Analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. *Front. Res. Metrics Anal.* **6**, 28 (2021).
32. Tu, J. H., Rowley, C. W., Luchtenburg, D. M., Brunton, S. L. & Kutz, J. N. On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.* **1**, 391–421 (2014).
33. Arbabi, H. & Mezić, I. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM J. Appl. Dyn. Syst.* **16**, 2096–2126 (2017).
34. Avila, A. & Mezić, I. Data-driven analysis and forecasting of highway traffic dynamics. *Nat. Commun.* **11**, 1–16 (2020).
35. Zhang, J.-M., Zou, L., Sun, T.-Z., Wen, Z.-H. & Yu, Z.-B. Experimental investigation on the propagation characteristics of internal solitary waves based on a developed piecewise dynamic mode decomposition method. *Phys. Fluids* **32**, 082102 (2020).
36. Al-Rousan, N. & Al-Najjar, H. Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. *J. Med. Virol.* **92**, 1603–1608 (2020).
37. Shim, E., Tariq, A., Choi, W., Lee, Y. & Chowell, G. Transmission potential and severity of COVID-19 in South Korea. *Int. J. Infect. Dis.* **93**, 339–344 (2020).
38. Greer, S. L., King, E., Massard da Fonseca, E. & Peralta-Santos, A. *Coronavirus Politics: The Comparative Politics and Policy of COVID-19* (University of Michigan Press, 2021).
39. Everitt, B. S. & Skrondal, A. *The Cambridge Dictionary of Statistics* 4th edn. (Cambridge University Press, 2010).
40. Kim, S. *et al.* Evaluation of COVID-19 epidemic outbreak caused by temporal contact-increase in South Korea. *Int. J. Infect. Dis.* **96**, 454–457 (2020).
41. Ministry of Health and Welfare, Korea. Refined social distance. http://ncov.mohw.go.kr/. (Accessed 9 July 2021).
42. BBC NEWS in Korea. Enhanced social distance. https://www.bbc.com/korean/news-55407911. (Accessed 9 July 2021).
43. Ministry of Health, Labour and Welfare, Japan. Novel coronavirus (COVID-19). http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00079. (Accessed 7 June 2021).
44. Briz-Redón, Á. & Serrano-Aroca, Á. The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modelling techniques. *Prog. Phys. Geogr. Earth Environ.* **44**, 591–604 (2020).
45. Ma, Y., Pei, S., Shaman, J., Dubrow, R. & Chen, K. Role of meteorological factors in the transmission of SARS-CoV-2 in the United States. *Nat. Commun.* **12**, 1–9 (2021).
46. Lolli, S., Chen, Y.-C., Wang, S.-H. & Vivone, G. Impact of meteorological conditions and air pollution on COVID-19 pandemic transmission in Italy. *Sci. Rep.* **10**, 1–15 (2020).
47. Kim, Y., Ryu, H. & Lee, S. Effectiveness of intervention strategies on MERS-CoV transmission dynamics in South Korea, 2015: Simulations on the network based on the real-world contact data. *Int. J. Environ. Res. Public Health* **18**, 3530 (2021).
48. Takeishi, N., Fujii, K., Takeuchi, K. & Kawahara, Y. Discriminant dynamic mode decomposition for labeled spatio-temporal data collections. Preprint at arXiv:2102.09973 (2021).
49. Kutz, J. N., Fu, X. & Brunton, S. L. Multiresolution dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.* **15**, 713–735 (2016).
50. Proctor, J. L., Brunton, S. L. & Kutz, J. N. Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.* **15**, 142–161 (2016).
51. Stewart, G. W. On the early history of the singular value decomposition. *SIAM Rev.* **35**, 551–566 (1993).
52. Hetzel, A. J., Liew, J. S. & Morrison, K. E. The probability that a matrix of integers is diagonalizable. *Am. Math. Mon.* **114**, 491–499 (2007).
53. Elyze, M., Guterman, A., Morrison, R. & Šivic, K. Higher-distance commuting varieties. *Linear Multilinear Algebra* 1–23 (2020).
54. O'Meara, K. & Vinsonhaler, C. On approximately simultaneously diagonalizable matrices. *Linear Algebra Appl.* **412**, 39–74 (2006).
55. Deif, A. Rigorous perturbation bounds for eigenvalues and eigenvectors of a matrix. *J. Comput. Appl. Math.* **57**, 403–412 (1995).
56. Wang, R. Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.* **36**, 158–177 (2015).
57. Ding, J. & Rhee, N. H. On the equality of algebraic and geometric multiplicities of matrix eigenvalues. *Appl. Math. Lett.* **24**, 2211–2215 (2011).
58. Statistics Korea. Statistical geographic information service. http://sgis.kostat.go.kr/jsp/english/index.jsp. (Accessed 5 June 2021).
59. e-Index. e-Indicators in South Korea. http://www.index.go.kr/main.do (Accessed 5 June 2021).
60. Seoul National University Asia Regional Information Center. COVID-19. http://sites.google.com/view/snuaric/COVID-19/COVID-19-data (Accessed 5 June 2021).

Acknowledgements
This work was supported by Samsung Science & Technology Foundation under Project Number SSTF-BA2002-02. The forth author is supported in part, by Brain Pool Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science and ICT (NRF2020H1D3A2A01041079) and by Faculty Development Leave Presidential Award funded by Texas State University.

Author contributions
S.K. conceptualization, validation, writing-original draft preparation; M.K. methodology, simulations; Y.L. methodology, formal analysis, writing-review and editing; S.L. conceptualization, writing-review and editing, supervision; All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03487-2.

Correspondence and requests for materials should be addressed to S.L. or Y.J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.