

Xia Wang, Dorcas Washington and Georg F. Weber\*

# Complex systems analysis informs on the spread of COVID-19

<https://doi.org/10.1515/em-2021-0019>

Received May 11, 2021; accepted August 31, 2021; published online September 17, 2021

## Abstract

**Objectives:** The non-linear progression of new infection numbers in a pandemic poses challenges to the evaluation of its management. The tools of complex systems research may aid in attaining information that would be difficult to extract with other means.

**Methods:** To study the COVID-19 pandemic, we utilize the reported new cases per day for the globe, nine countries and six US states through October 2020. Fourier and univariate wavelet analyses inform on periodicity and extent of change.

**Results:** Evaluating time-lagged data sets of various lag lengths, we find that the autocorrelation function, average mutual information and box counting dimension represent good quantitative readouts for the progression of new infections. Bivariate wavelet analysis and return plots give indications of containment vs. exacerbation. Homogeneity or heterogeneity in the population response, uptick vs. suppression, and worsening or improving trends are discernible, in part by plotting various time lags in three dimensions.

**Conclusions:** The analysis of epidemic or pandemic progression with the techniques available for observed (noisy) complex data can extract important characteristics and aid decision making in the public health response.

**Keywords:** autocorrelation; average mutual information; complex systems; COVID-19; epidemiology; fractal dimension; new infections; wavelet analysis.

## Introduction

The spread of infectious diseases depends on pathogen factors (virulence), host factors (immunity), and – on the population level – on countermeasures taken by the community. Such measures cover a broad spectrum of possible engagements, and they may be highly consequential for the course of an epidemic or a pandemic (Christakis 2020). The analysis of acute infectious progression in a society is critical for gauging the effectiveness of public health responses, but it is made difficult through the non-linear nature of the underlying process. Conventional approaches of reductionist research or common linearization techniques are not meaningfully applicable. The differential equations of the frequently applied SIR (susceptible, infectious, recovered individuals) and SEIR models (susceptible, exposed, infectious, recovered individuals) do not capture the complex nature of epidemiologic progression, even when additional variables are taken into account (Al-Raei, El-Daher, and Solieva 2021; Stehlé et al. 2011), such as the inclusion of quarantine (Groendyke and

---

\*Corresponding author: Georg F. Weber, Academic Health Center, University of Cincinnati, Cincinnati, OH, USA, 231 Albert Sabin Way, Cincinnati, OH 45267-0514, USA, Phone: +1 513 558 0947, E-mail: georg.weber@uc.edu

Xia Wang, Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, USA

Dorcas Washington, Health Science Library, University of Cincinnati, Cincinnati, OH, USA

Combs 2021). Stochastic transition models and newer mathematical models to characterize imported escaper and asymptomatic patients have displayed similar limitations (Sun and Wang 2020).

Various strategies have been employed to account for the complexity of infectious propagation. The spread of COVID-19 has been modeled with machine learning (Mehta et al. 2020), networks of compartments (Wang et al. 2020) and cellular automata (Bin, Sun, and Chen 2019). Power laws have been inferred (Blasius 2020). Such investigations are of value, even though they are inevitably based on idealizing assumptions regarding the mechanisms of progression. These assumptions generate algorithms, which are then refined to match the real curves as closely as possible. In addition to modeling approaches, the analysis of actually observed data is of critical importance. This latter strategy does not require mechanistic premises or hypotheses at the outset. It seeks to extract patterns and characteristics from the numbers obtained in the field. The challenge has been that the numbers in such data sets are noisy, and they are eminently non-linear (also described as “complex data” or “observed chaotic data” (Abarbanel 1995)). Complex systems research has increasingly made techniques and algorithms available to extract relevant information from observed non-linear data series.

The manifestations of the COVID-19 pandemic have varied widely among geographic areas, when compared across countries (Bertacchini, Bilotta, and Pantano 2020; Chakraborty and Maity 2020; Wang et al. 2020) as well as across US states (White and Hébert-Dufresne 2020), depending on when the virus reached them, what the population characteristics were at the time of onset, and what actions were taken in response to the infectious spread. Here, we set out to investigate underlying patterns. We apply basic tools of complex systems research to compare the spread of COVID-19 in distinct countries, characterized by their varying approaches to the pandemic, from its beginning stages through early or late October 2020. Further, we compare various regions within the USA, which has left major decisions to the individual states. Patterns are discernible in Fourier and wavelet analyses. Order can be detected in time-lagged plots. Therefrom, quantitative measurements are obtainable, including autocorrelation, average mutual information (AMI), fractal dimension, and embedding dimension, which inform on the pandemic progression.

## Methods

### Source data

Here we analyze the new infections per day, either as absolute numbers or as rates per 10,000 inhabitants. The source data utilized for the present analysis came from Bing COVID-19 Tracker ([www.bing.com/covid](http://www.bing.com/covid)).

### Fourier spectrum and univariate wavelet analysis

Fourier analysis evaluates the spectral density by relative numbers of new infections (case rates per 10,000 inhabitants) vs. frequency or vs. period. Wavelet analysis does not assume stationarity in the time-series. Thus, it allows the study of localized periodic behavior. In particular, we look for regions of high-power in the frequency-time plot. The calculations for wavelet analyses of new infections were done in R, a free software environment for statistical computing and graphics (R Core Team 2020). In WaveletComp, the null hypothesis, that there is no periodicity in the series, is tested via  $p$ -values obtained from simulation, where the model to be simulated can be chosen from a range of options (Roesch and Schmidbauer 2018). The algorithm analyzes the frequency structure of uni- or bivariate time series using the Morlet wavelet. The time series to be analyzed was standardized, after detrending, in order to obtain a measure of the wavelet power, which is relative to unit-variance white noise and directly comparable to results of other time series. Detrending is accomplished using polynomial regression. Where indicated, all graphs are normalized to the same  $y$ -axis scale.

### Bivariate wavelet analysis

Wavelet methodology transforms the time series from the time domain to the frequency domain just as the spectrum analysis does. It can be applied to a single time series (univariate wavelet analysis) or two time series simultaneously (bivariate wavelet analysis). As an important analysis tool to study periodic phenomena in time series, wavelets are particularly useful in the presence

of potential frequency changes across time. We conducted bivariate analysis of lagged data ( $t$  vs.  $t + 7$  or  $t + 14$  or  $t + 21$ ) for joint periodicity. The concepts of cross-wavelet analysis provide tools for comparing the frequency contents by two time series as well as for drawing conclusions about their synchronicity at certain periods and across certain ranges of time. While cross-wavelet power corresponds to covariance in the time domain, wavelet coherence is a time-series measure similar to correlation. Two waves are coherent if they have a constant relative phase. The bivariate analysis results include the cross-wavelet power plot, the wavelet coherence plot, the average power plot and the phase difference image. The cross-wavelet power and coherence plot contain arrows showing the area of significant joint periods (significance level = 0.05). The direction of these arrows indicating the direction of phase differences. Up-right pointing arrows indicate that the two series are in phase and  $x(t)$  series leads, while down-right pointing arrows indicate the two series are in-phase and  $x(t + n)$  series leads. Similarly, up-left pointing arrows express that the two series are out of phase and  $x(t + n)$  series leads, while down-left pointing arrows express that the two series are out of phase and  $x(t)$  series leads. The arrows are only plotted within white contour lines indicating significance at the 10% level. A more explicit global view of the phase difference can be produced with  $(\pi/2, \pi)$  and  $(-\pi, -\pi/2)$  for out of phase and  $(-\pi/2, \pi/2)$  for in-phase. The time-averaged cross-wavelet power provides a summarized view on the shared periods, the corresponding power and the statistical significance. Cross-wavelet plots may mark areas significant due to one series swinging widely, rather than two series sharing a joint period. To avoid this false positive readout, it is more appropriate to examine wavelet coherence plots, like the coefficient of correlation. It has a value range between 0 and 1 and it shows statistical significance only in areas where the two series actually share jointly significant periods.

### Return plots

From the total numbers of new infections, we generated return plots with increasing lags, plotting daily changes  $x(t + 1), \dots, x(t + 7)$  vs.  $x(t)$  and weekly changes  $x(t + 14), \dots, x(t + 49)$  vs.  $x(t)$ . Short time lags tend to cluster around the 45° angle, whereas increasing time delays reveal the structure of the oscillations. When graphed in 3 dimensions, these diagrams can aid in reconstructing the underlying attractor.

### Autocorrelation

A time series sometimes repeats patterns or has other properties, whereby earlier values display some relation to later values. The autocorrelation statistic (serial correlation statistic) measures the degree of that affiliation as it refers to linear dependence. The magnitude of its dimensionless number reflects the extent of similarity. The formula for autocorrelation  $R_m$  is comprised of terms for autocovariance and variance

$$\text{autocorrelation} = \frac{\text{autocovariance}}{\text{variance}}$$

$$R_m = \frac{\frac{1}{N} \sum_{t=1}^{N-m} (x_t - \bar{x})(x_{t+m} - \bar{x})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2}$$

Autocorrelation coefficients range from  $-1$  to  $+1$ , with  $+1$  indicating perfect synchrony and  $-1$  reflecting exact mirror images. An absence of any correlation yields  $R_m = 0$ .

### Box counting dimension

The dimension of a fractal is best described as a non-integer. The dimension is a quantitative measure for the evaluation of geometric complexity by objects. A general relationship assumes

$$\text{dimension} \propto \frac{\log(\text{number of increments})}{\log\left(\frac{1}{\text{scale size}}\right)}$$

Here, the characteristic of dimension is that it specifies the rate, at which the number of increments varies with scale size. We calculated the box counting dimension after binning into  $16 \times 16$  squares of 2-dimensional return plots with various lags.

### Average mutual information

The AMI represents a non-linear correlation function, which indicates how much common information is shared by the measurements of  $x(t)$  and  $x(t + n)$ . The AMI was calculated with the mutual function R package tseriesChaos. It estimates the mutual information index for a specified number of lags. The joint probability distribution function is estimated with a simple bi-dimensional density histogram.

## Embedding dimension

Here by R package `nonlinearTseries`, we first use the `timeLag` function to decide the optimal time lag  $\tau$  based on the AMI and then by the `estimateEmbeddingDim` function to assess the optimal embedding dimension  $m$ . Then the optimal set of regressors related to  $x(t)$  is  $x(t - \tau), \dots, x(t - (m - 1)\tau), x(t - m\tau)$ . The function to determine the minimum embedding dimension from a scalar time series applies the algorithm by Cao (1997), which recommends using two functions,  $E1(d)$  and  $E2(d)$ , to determine the minimum embedding dimension, where  $d$  denotes the dimension.  $E1(d)$  stops changing when the dimension  $d$  is greater than or equal to the embedding dimension ( $d_0$ ), staying close to 1.  $E2(d)$  distinguishes deterministic signals from stochastic signals for stagnating  $E1(d)$ .

## Surrogate data analysis

For non-linear data sets it is important to distinguish between complexity and randomness. To investigate the complexity inherent in the time series, we carried out a collection of non-linearity tests on the data from 9 countries, including the surrogate data test on a linear stochastic process assumption, as well as additional tests with various null hypotheses. Teraesvirta's neural network test for neglected non-linearity uses a Taylor series expansion of the activation function to arrive at a suitable test statistic. It uses either the  $F$ -statistic or the  $\chi^2$  statistic. The White neural network test for neglected non-linearity of the time series is consistent against arbitrary non-linearity in mean. It uses either the  $F$ -statistic or the  $\chi^2$  statistic. The McLeod–Li test assesses autoregressive conditional heteroscedascity (ARCH) among various time lags. The Keenan test is designed to have optimal local power against departure from the linear autoregressive function in the direction of the square of the linear autoregressive function. The null hypothesis of the Tsay test is that the true model is an autoregressive (AR) process. The AR order, if missing, is estimated.

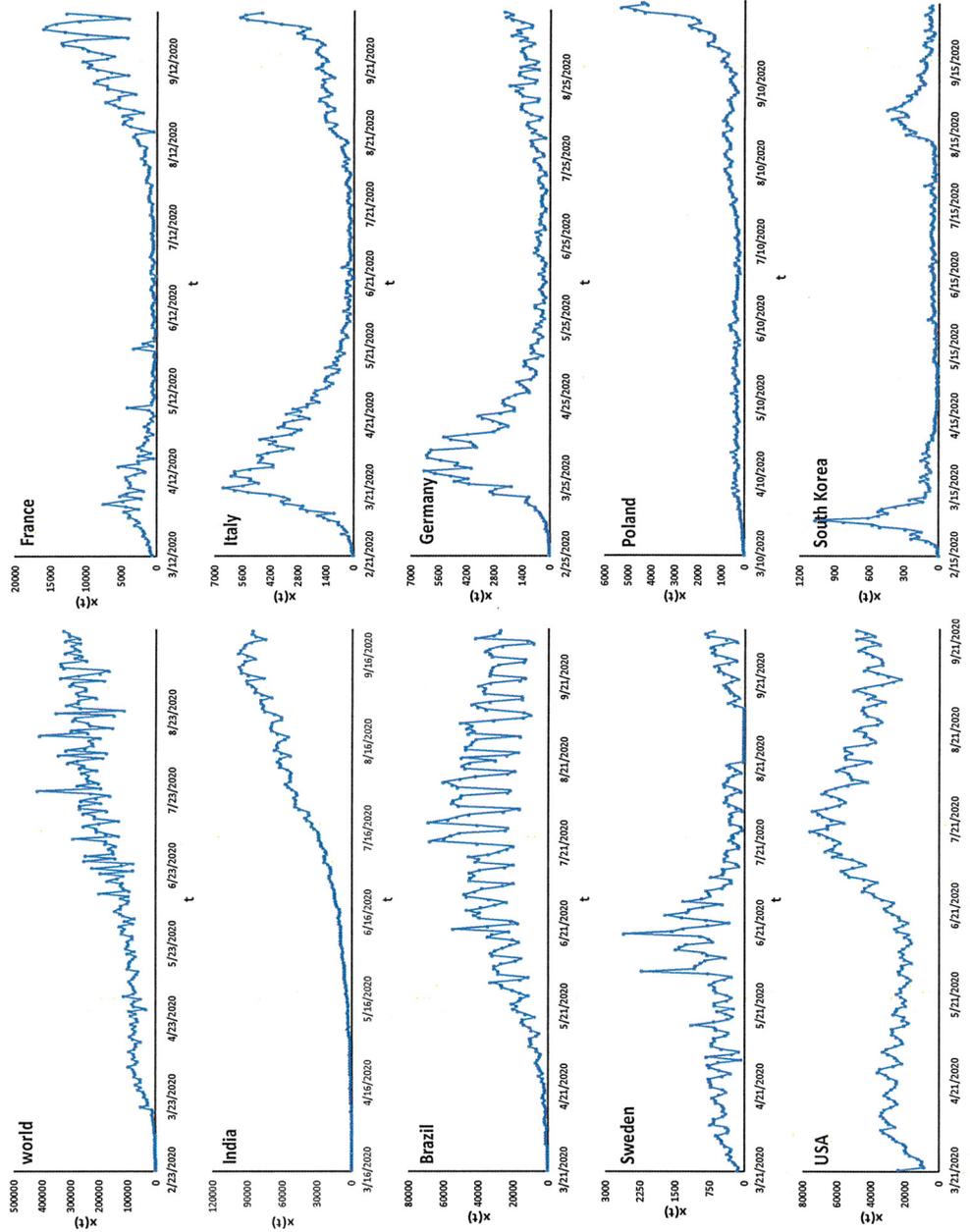
# Results

## Comparison across countries

Across countries, a wide spectrum of measures was taken to curb the spread of SARS-CoV2. This resulted in a range of very different progression curves when graphing the numbers of new infections over time (Figure 1). India, Brazil, Sweden, Italy and the United States have been considered as hard-hit for their own internal reasons. France, Germany, over a long period Poland, and South Korea had tighter control and a less aggressive spread. All curves display close to linear ramp-up phases, followed by more or less irregular oscillations. The levels of success at suppressing the new infection rates diverged among countries, and several are experiencing a second peak. Surrogate data analysis showed that these time series are complex, not random (Table 1).

Wavelet methodology aids in studying periodic phenomena in time series, particularly in the presence of potential frequency changes over time. For cross-country evaluations, all graphs were plotted on the same scale (Figure 2A). Each country was also plotted on its own scale (Figure 2B). The univariate analysis of the time course for the countries under study shows prominence of the recent upswing in France (heat intensity on the right margin of the graph). By contrast, there is comparatively more successful management by Italy, Germany, Poland and South Korea through October 2020. India, Brazil, Sweden, and the United States display cyclical fluctuations of various durations, none of which have been contained. A period of 7 days is prominent in the fluctuations of most countries, which may reflect real cyclicity or weekly reporting habits. The worldwide data are displayed in Figure S1.

For cross-country comparisons, we converted the new infection total numbers to new infection rates by relating them to 10,000 members of the population (Figure 3A). Complex systems can be analyzed with Fourier analysis. We first plotted Fourier power spectra vs. frequency for the rates of new infections (Figure 3B). Spectral density range (high in Brazil, low in South Korea) and frequency distribution provide a readout for infectious spread. The spectral density of the normalized rates (identically scaled y-axes) (Figure 3C) confirmed good management of the pandemic spread in Germany, Poland, and South Korea (and to some degree in Italy). Despite the progressive increase in the numbers of infections in India, on a population basis,

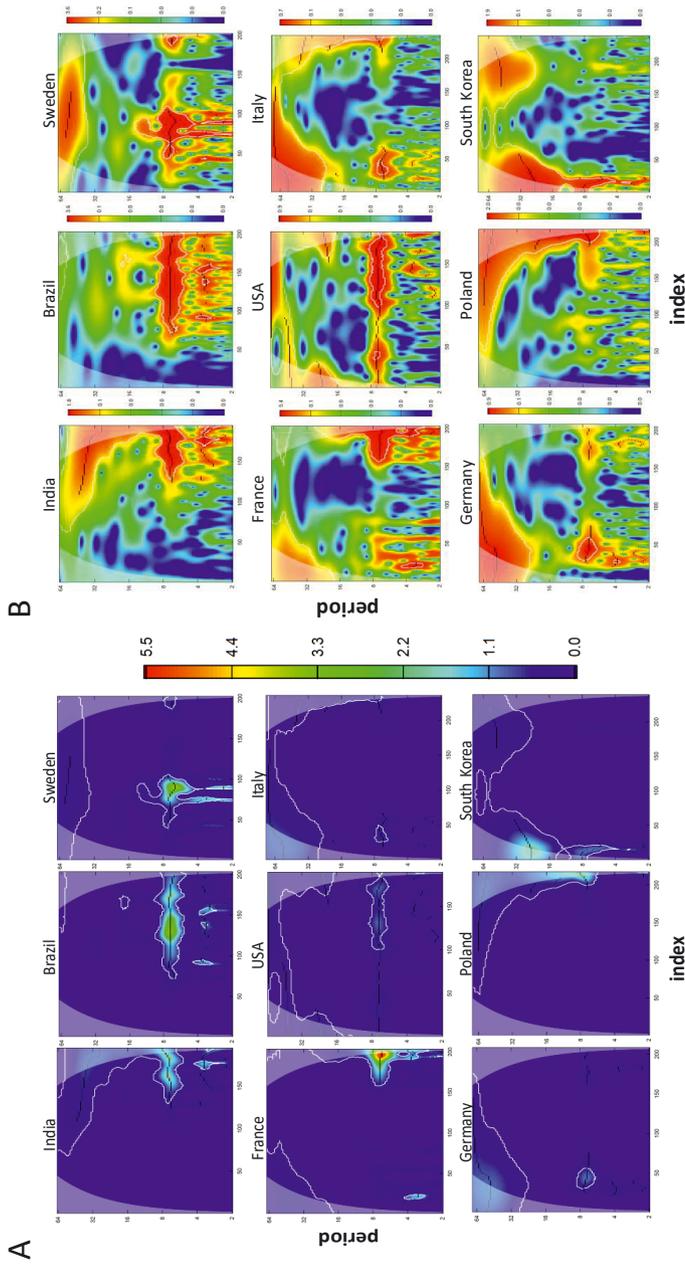


**Figure 1:** Time-course of disease spread by country. Numbers of new cases,  $x(t)$ , per day vs. time ( $t$ , indicating the date). Shown are the curves for (top to bottom, left to right) the globe, India, Brazil, Sweden, Italy, USA, France, Germany, Poland, and South Korea. Note the different scales of the y-axes.

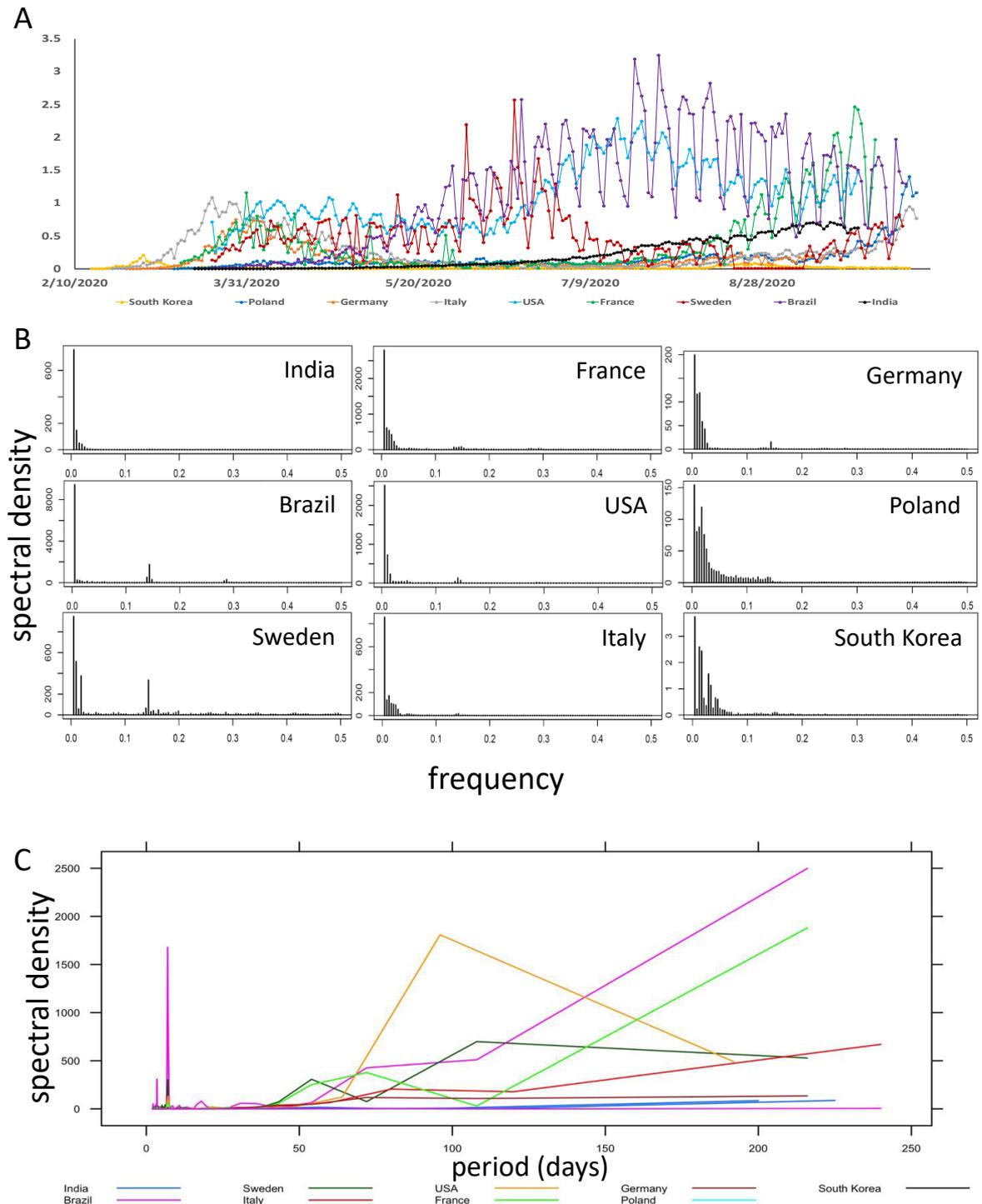
Table 1: Surrogate data analysis by country.

Test	Null hypothesis	India	Brazil	Sweden	Italy	USA	France	Germany	Poland	South Korea
Surrogate data test ( $p = 0.05$ )	Data come from a linear stochastic process	Reject	Reject	Reject						
Terasvirta	Linearity in “mean”	0.0014	<0.001	0.0035	0.022	0.456	0.4644	0.0032	0	<0.001
White	Linearity in “mean”	0.0021	<0.001	0.0042	0.0313	0.6599	0.4712	0.0145	<0.001	0.0019
McLeod – Li	Time series follows some ARIMA process	0	0	<0.001	0	0	0	0	0	0
Keenan	Time series follows some AR process	0.007	0.0283	0.3606	0.003	0.5315	0.8856	0.2188	0.0789	<0.001
Tsay	Time series follows some AR process	<0.001	NA	NA	<0.001	0.0158	<0.001	<0.001	<0.001	<0.001

The left column displays the test applied (as described in Methods), followed by the null hypothesis. The results are shown by country in the ensuing columns. ARIMA = autoregressive integrated moving average. AR process = autoregressive model representation of a type of random process.



**Figure 2:** Univariate wavelet analysis. Cross-wavelet power spectrum in the time-period domain. The x-axis (index) displays the time progression, whereas the y-axis depicts the length of the period. White contour lines indicate significance of periodicity on the 0.1 level for probability of error. Lines represent the ridge of cross-wavelet power. The color bar reveals the power gradient. (A) All countries on the same scale. (B) Each country on its own scale.



**Figure 3: Fourier analysis.**

(A) New infection rates. Daily reported new numbers of infections divided by 10,000 inhabitants. The x-axis shows the calendar date. (B) Power spectrum. Fourier power spectra vs. frequency for new infections per 10,000 inhabitants per day in each of 9 countries. (C) Normalized power spectrum. Spectral density (y-axis) vs. period (in days) for infection rates per 10,000 inhabitants (x-axis). The curve shows the smoothed spectral density estimates. All y-axes have the same scale.

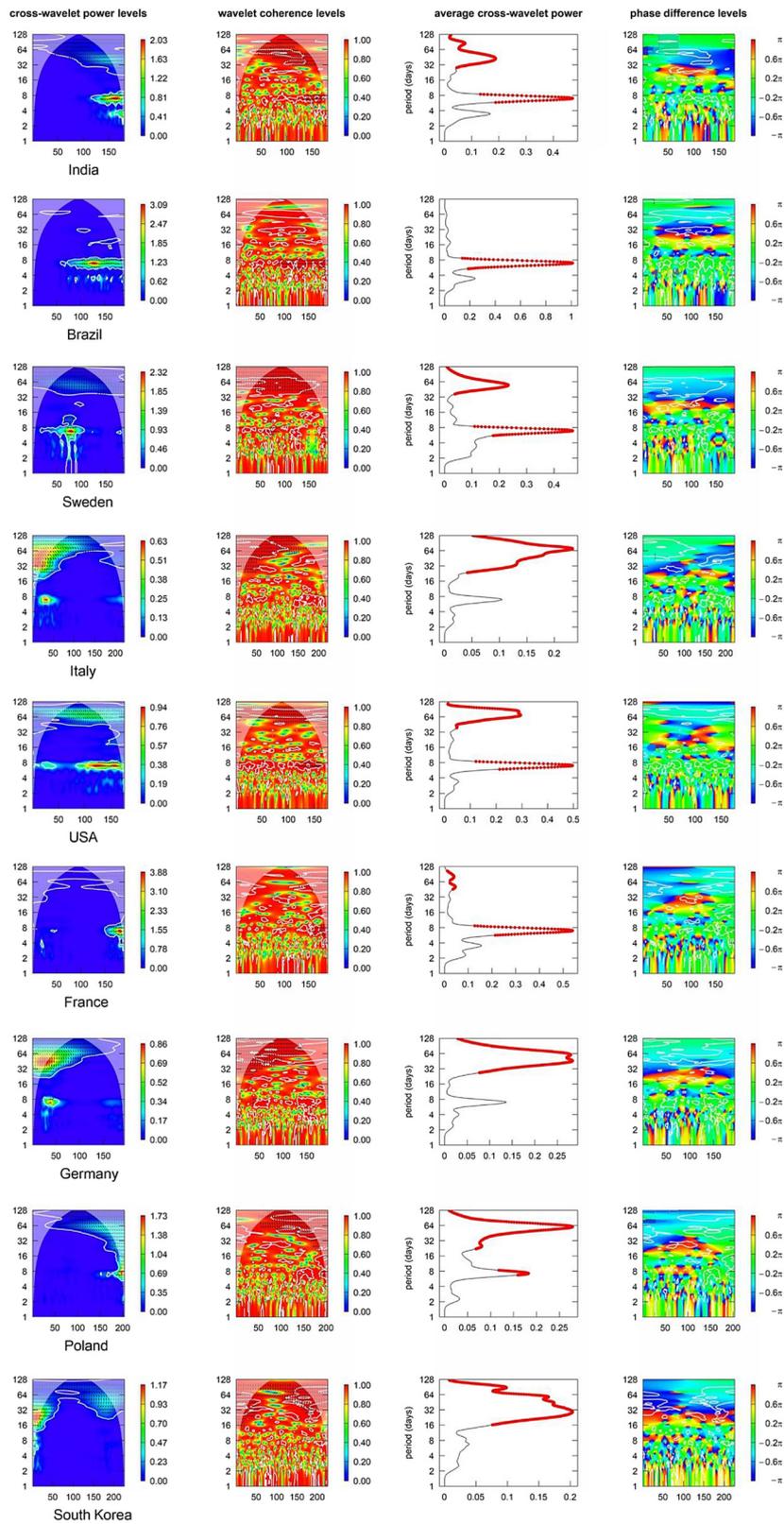
control has apparently not been lost through October 2020. By contrast, the power spectra for Brazil, Sweden, and France are reflective of potentially adverse developments. The United States display an anomaly with a periodic behavior that has a prominent cycle around 100 days.

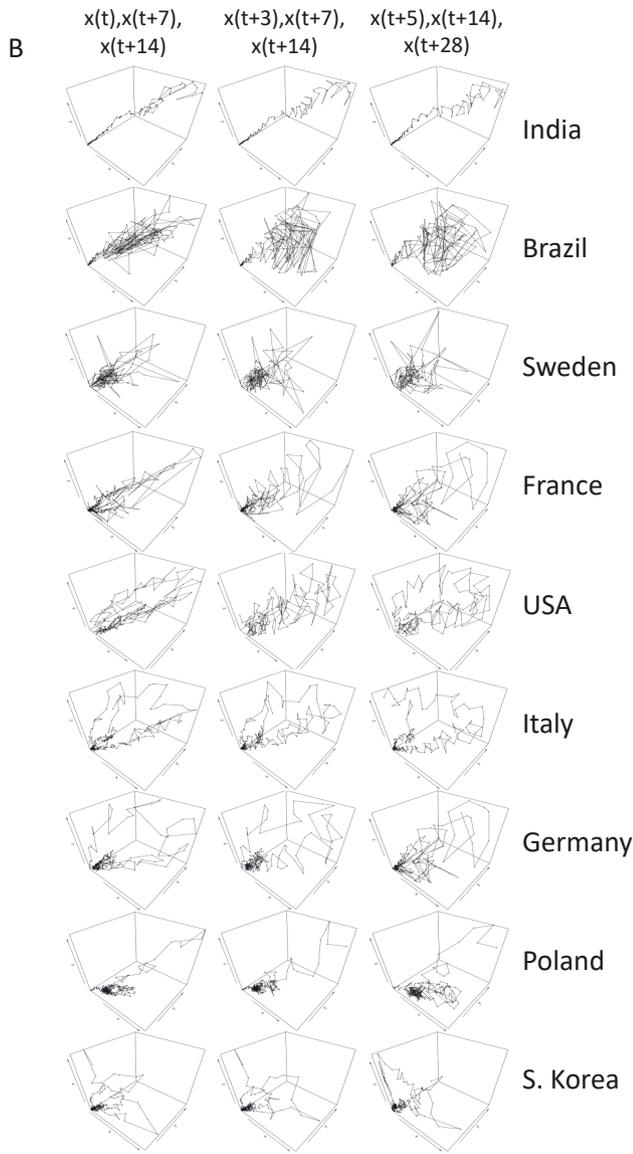
To gain a better understanding of the dynamics, with which disease spread occurs, we investigated progressive numbers of new infections in comparison to their increasing time lags. This approach may reveal periodicities or aid in the visualization of attractors. Expectedly, short time delays were associated with little change. With a lag time of about 7 days onward, distinct patterns emerged among countries. According to bivariate wavelet analysis for time-delayed data series (including the cross-wavelet power plot, wavelet coherence plot, average power plot and phase difference image), Italy, Germany and South Korea shared significantly joint periods of 1–2 months in the comparison  $x(t)$  vs.  $x(t + 7)$ . South Korea has comparatively high power and significant shared periods around 3 weeks at the early stage and later the significant shared periods are also 1–2 months. The remaining countries all have segments of shorter periods (around 7 days) and longer periods shared. For India, Brazil, France, USA and Poland, the shared 7-day period only appears significant in the later part of the series. Similar results are observed in the analyses for  $x(t)$  vs.  $x(t + 14)$  and  $x(t)$  vs.  $x(t + 21)$ . The phase difference plots show that in the shared longer periods,  $x(t)$  are mostly in phase with  $x(t + 7)$ , while they gradually become out of phase in  $x(t)$  vs.  $x(t + 14)$  and  $x(t)$  vs.  $x(t + 21)$ , thus making longer lags more discriminating and informative (Figure 4A and Figure S2A and B). A reduction in cross-wavelet power levels is apparent in Italy, Germany and South Korea. Poland and France are experiencing recent increases. India, Brazil and the USA have had protracted periods of high cross-wavelet power levels. Containment is associated with longer periodicity in the distribution of cross-wavelet power. This is the case for South Korea, Germany and Italy. High cross-wavelet power around a periodicity of 7 days is reflective of poor control.

To generate informative return plots, we utilized 3 dimensions, which allows for the visualization of two lags from  $x(t)$  (or from a later start point) and may reveal the pattern of an attractor. In this depiction, a rapid increase or decrease in new infections is reflected in a close-to straight line, oscillations generate a near-toroid attractor, while successful management shrinks the torus and moves it closer to the origin. Initially, we evaluated multiple time delays. Most discriminating were  $x(t)$ ,  $x(t + 7)$ ,  $x(t + 14)$  and  $x(t + 3)$ ,  $x(t + 7)$ ,  $x(t + 14)$  and  $x(t + 5)$ ,  $x(t + 14)$ ,  $x(t + 28)$  (Figure 4B). The progressive increase in new cases over the time period in India is reflected in a predominantly linear curve on each scale. The wide fluctuations in Brazil generate a largely disordered appearance. Disorder is also apparent in Sweden. France initially managed the pandemic well, but is experiencing a dramatic upswing, which obscures order. Cyclical patterns, suggesting the outlines of attractors, are apparent in USA, Italy, Germany, and South Korea (where most data points are concentrated near the origin). Poland initially displayed a well-contained attractor, but the recent substantial upswing in new infections is reflected in a linear progression from there (for separate analyses of the two phases, see Figure S3). We also calculated the embedding dimensions for the lagged data (Figure 4C). Germany has the highest embedding dimension of 10, followed by Poland with 9. Several countries have an embedding dimension of 7, including Brazil, Sweden, USA and South Korea. Italy and France have the embedding dimension equal to 5. India is unusual due to its longer lag period of 24 days. When the lag period is set at 7 days, the embedding dimension of India is also equal to 7. For the worldwide data, the calculated embedding dimension is 7 with a time lag of 1 (not shown).

The autocorrelation of two data strings with short time lags is expected to be high (approaching 1.0) because there is little opportunity for dramatic change (high infection rates on day  $t$  likely produce similarly high numbers on the consecutive day  $t + 1$ , while low numbers are followed by few new infections on the next day). Autocorrelation may remain high for extended lags in the initial ramp-up and at the oscillatory stage, depending on the regularity of the fluctuations. A society that succeeds in curbing the disease spread will leave the highly correlated initial ramp-up and consecutive oscillatory phases, thus displaying a gradual decrease in values at the longer lags. The decline in the autocorrelation numbers of progressively lagged data by country appeared to be reflective of the stringency, with which the pandemic was addressed (Table 2A). From a lag of 6 onward, Poland and South Korea have substantially declining values (although due to the recent steep upswing in new infections, Poland deviates from the trend at very long lags), Germany shows a

A





**Figure 4:** Time-lagged data analysis.

(A) Bivariate wavelet analysis. Shown are cross-wavelet power plot, wavelet coherence plot, average power plot and phase difference image (from left to right in each row) time-lagged data were used for  $x(t), x(t + 14)$  (for the lags  $x(t), x(t + 7)$  and  $x(t), x(t + 21)$  see Figure S2). White contour lines indicate significance for joint periodicity, black arrows depict the phase difference in the areas with significant joint periods. The solid red dots on the average power plot (the third from the left) depict significant joint periods at a probability of error of 0.1. Where shown, the color bars reveal the ranges of cross-wavelet power levels. (B) Return plots in 3 dimensions. Time-lagged return plots in 3 dimensions are shown (from left to right) for  $x(t), x(t + 7), x(t + 14)$  and  $x(t + 3), x(t + 7), x(t + 14)$  and  $x(t + 5), x(t + 14), x(t + 28)$ . Each country of interest has its own row. (C) Embedding Dimension. The plots show how Cao’s algorithm uses 2 functions in order to estimate the embedding dimension from the time series (the  $E1(d)$  and  $E2(d)$  functions), where  $d$  denotes the dimension.

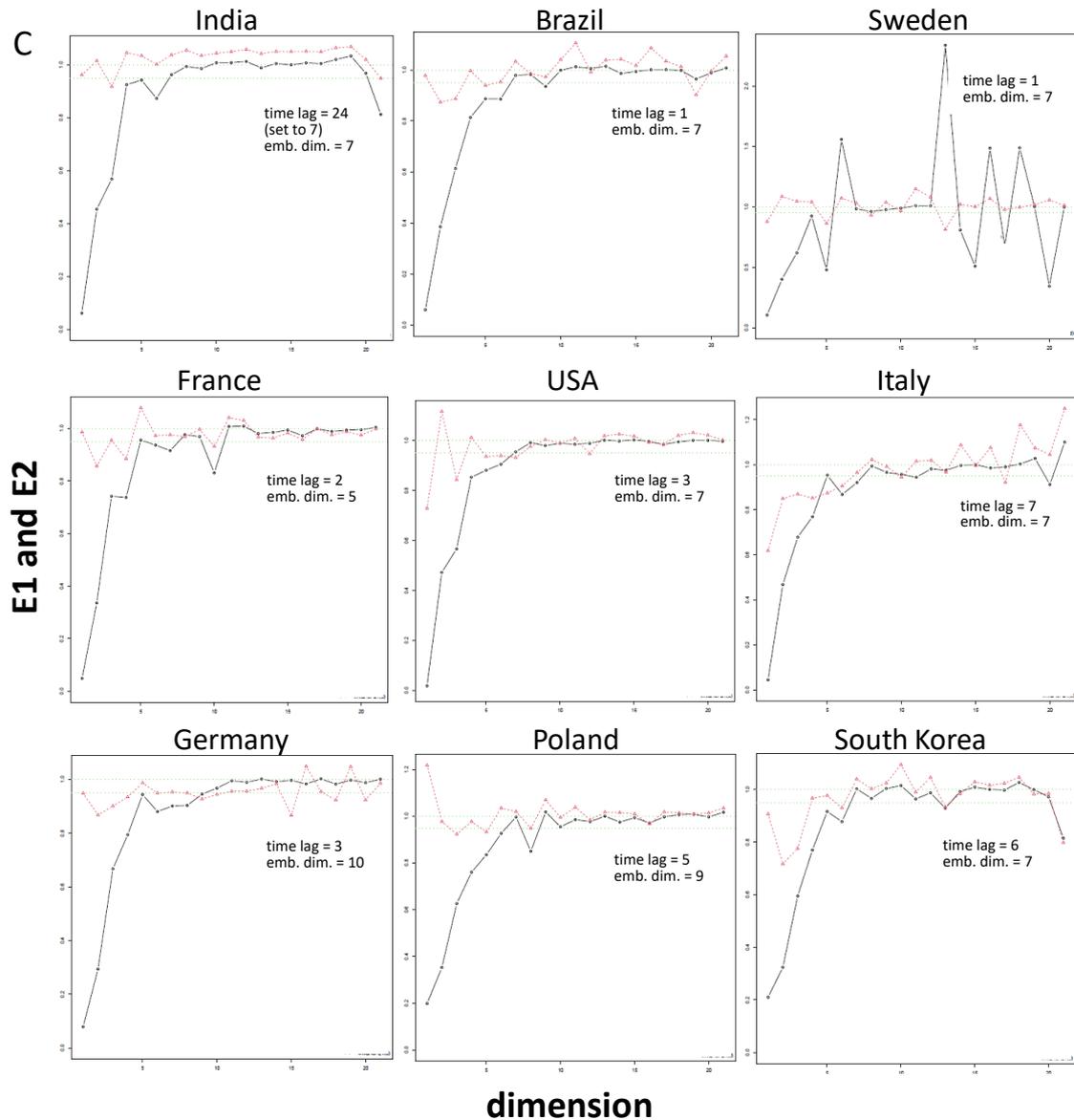


Figure 4: (continued)

dramatic lowering at a lag of 21 and above. By contrast, India and Brazil stay uniformly high. So do the global numbers, which are inherently heterogeneous.

The AMI reflects information shared by the measurements of  $x(t)$  and  $x(t + n)$ . Expectedly, it declines with increasing lag. Poland starts with a relatively low value (1.15 at  $t$  vs.  $t + 1$ ) and shows a rapid decrease with longer lag. It then stays around at a low level of 0.15 from lags of 21–49 days. France displays a gradually decreasing trend with the AMI starting at 1.60 and ending at 0.34 at the lag of 49 days. India shows a similar pattern as France but with much higher AMI (due to the constant uptick in numbers), ranging between 2.61 and 1.37. Four other countries, including Germany, USA, Sweden and Brazil, all express relatively flat AMI values, staying around levels of 2.20 for the USA and Brazil, 1.5 for Germany, and 1.3 for Sweden. Reflecting progressively improved control, Italy and South Korea also have decreasing trends, but much flatter at 1.96–1.36 for Italy and 1.26 to 0.66 for South Korea, respectively (Table 2B).

Table 2: Readouts of complexity for lagged data on COVID-19 spread by country.

A	$t + 1$	$t + 2$	$t + 3$	$t + 4$	$t + 5$	$t + 6$	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$	$t + 42$	$t + 49$
World	0.805	0.848	0.782	0.818	0.843	0.815	0.829	0.761	0.694	0.660	0.626	0.588	0.535
India	0.990	0.980	0.947	0.969	0.962	0.957	0.949	0.867	0.779	0.701	0.631	0.556	0.470
Brazil	0.816	0.633	0.573	0.557	0.618	0.796	0.898	0.851	0.800	0.709	0.629	0.515	0.412
Sweden	0.725	0.537	0.427	0.470	0.565	0.687	0.704	0.562	0.383	0.171	0.137	0.096	0.065
USA	0.947	0.885	0.841	0.823	0.842	0.881	0.895	0.766	0.576	0.395	0.201	0.043	-0.068
France	0.882	0.800	0.769	0.767	0.768	0.791	0.795	0.606	0.445	0.285	0.152	0.036	-0.044
Italy	0.970	0.931	0.863	0.862	0.848	0.839	0.825	0.613	0.408	0.233	0.101	0.025	-0.143
Germany	0.943	0.894	0.860	0.849	0.841	0.838	0.835	0.517	0.192	-0.048	-0.148	-0.159	-0.162
Poland	0.915	0.827	0.633	0.595	0.657	0.486	0.467	0.269	0.136	0.077	0.103	-0.381	0.167
South Korea	0.922	0.845	0.772	0.703	0.642	0.575	0.495	0.088	0.065	0.043	-0.038	-0.106	-0.128
B	$t + 1$	$t + 2$	$t + 3$	$t + 4$	$t + 5$	$t + 6$	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$	$t + 42$	$t + 49$
World	2.145	2.134	2.111	2.113	2.104	2.111	2.065	1.951	1.817	1.859	1.784	1.653	1.609
India	2.609	2.537	2.480	2.481	2.473	2.539	2.513	2.356	2.132	1.925	1.718	1.532	1.365
Brazil	2.311	2.284	2.218	2.230	2.244	2.203	2.301	2.227	2.250	2.189	2.121	2.129	2.083
Sweden	1.416	1.272	1.287	1.315	1.293	1.272	1.387	1.252	1.251	1.259	1.312	1.422	1.368
USA	2.234	2.204	2.145	2.155	2.173	2.124	2.170	2.195	2.192	2.219	2.221	2.157	2.116
France	1.597	1.543	1.484	1.497	1.451	1.463	1.476	1.230	1.007	0.811	0.584	0.479	0.342
Italy	1.960	1.908	1.859	1.813	1.772	1.792	1.775	1.561	1.549	1.460	1.489	1.426	1.367
Germany	1.746	1.660	1.576	1.611	1.565	1.550	1.642	1.539	1.461	1.507	1.440	1.419	1.501
Poland	1.155	1.101	0.942	0.892	0.893	0.829	0.835	0.535	0.235	0.133	0.143	0.169	0.146
South Korea	1.270	1.197	1.145	1.086	1.003	1.053	1.035	0.929	0.920	0.880	0.799	0.687	0.667
C	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$								
World	1.445	1.532	1.532	1.542	1.489								
India	1.292	1.202	1.292	1.250	1.282								
Brazil	1.464	1.494	1.537	1.532	1.494								
Sweden	1.339	1.373	1.365	1.373	1.357								
USA	1.506	1.537	1.542	1.562	1.567								
France	1.418	1.389	1.373	1.339	1.339								
Italy	1.500	1.547	1.547	1.506	1.489								
Germany	1.432	1.494	1.527	1.527	1.527								

Table 2: (continued)

C	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$
Poland (con.)	1.537	1.547	1.542	1.537	1.567
Poland	1.000	1.000	1.022	1.062	1.098
South Korea	1.282	1.272	1.131	1.098	1.131

(A) Autocorrelation. Tabulation of the autocorrelation in COVID-19 spread with each row representing a different country. The selected time lags are indicated in the top row, all are calculated vs.  $x(t)$ . (B) Average mutual information. Values of AMI in COVID-19 spread with each row representing a different country. The selected time lags as indicated in the top row are all calculated vs.  $x(t)$ . (C) Fractal dimensions. Box counting dimensions are calculated for 2-dimensional return plots of increasing lags,  $x(t + 7)$  vs.  $x(t)$  through  $x(t + 35)$  vs.  $x(t)$ . 9 countries are evaluated, and the worldwide numbers are included. Poland is represented twice, over the entire evaluation period through 12 October 2020 (which contains a steep incline) and over the shorter phase of containment through 18 September 2020 (cont. = contained period).

A rapid increase in new infections is reflected in a small fractal dimension (practically approximated by the box counting dimension with values between 1 and 2) of the 2-dimensional return plots with progressive lags. Intermediate phases are characterized by higher fractal dimensions (approaching 2), depending on the nature of the oscillations. Conversely, successful management through the reduction in new infections should be reflected in a contraction of the attractor on the return plot, which is assessable through the box counting dimension. A trend is displayed in the comparisons from shorter to longer lag periods. Distinct management strategies across different countries generate a heterogeneous pattern worldwide, rendering the fractal dimension high regardless of the lag in  $x(t+n)$  vs.  $x(t)$  plots. Steep increases in new infections (Poland, India) have dimensions close to 1. Intermediate phases are characterized by higher numbers. Successful fights against the pandemic (South Korea) are causative for declining size dimensions with increasing lag (Table 2C).

## Comparison across US states

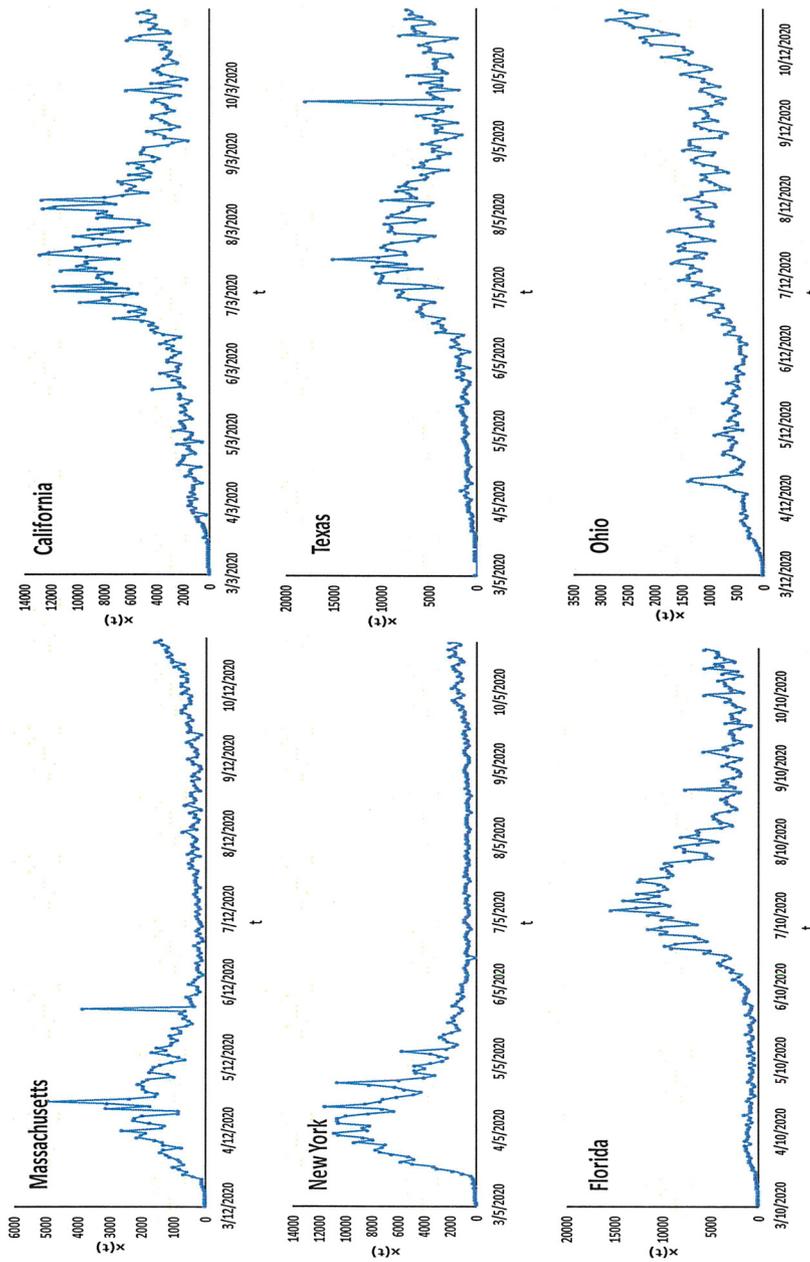
Within the USA, individual states have encountered a rather wide range of progression phenotypes in the spread of new COVID-19 infections (Figure 5). This is due to variations in international connectedness and population density (reflected in the early peaks in the Northeastern states New York and Massachusetts), holiday travel (Florida), policy decisions and other factors.

Wavelet analysis of new infections (one scale across all states) shows good control (right side of the graph) after initial affliction (left area) for Massachusetts and New York, which having had early spikes in new infections have achieved good success in containment. Through the observation period, control has not been maintained in Ohio. The periodicity in individual states (each on their own scales) is poorly defined, except for Florida and Ohio, where 7 days yield a prominent signal (Figure 6A and B).

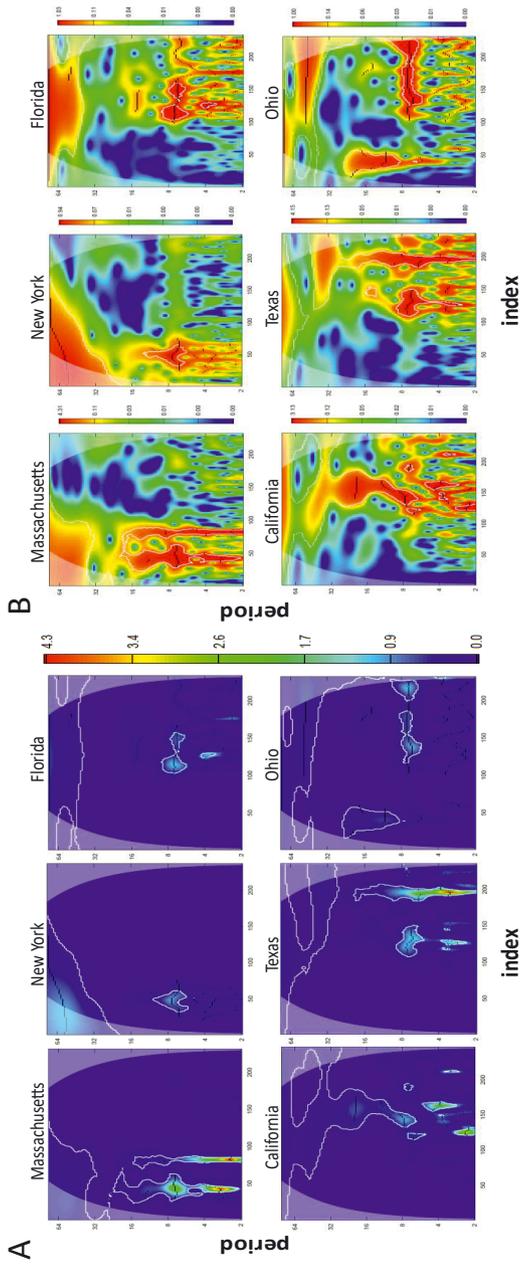
We normalized the new infection numbers to rates by relating them per 10,000 inhabitants (Figure 7A). Figure 7B shows the periodogram for the 6 states under investigation with frequencies between 0 and 0.10 (the graph is almost flat for the higher frequencies). There exist clear heterogeneous patterns in the comparison among these states. New York and Massachusetts display steadily decreasing spectral density values from the longest period to around 1–2 weeks (corresponding to a frequency range around 0.07–0.14). Florida and Texas share similar patterns with a few low spikes in their periodograms after the first 3 highest ones. The graph for California flattens out after the lowest three frequencies, with the longest period (the whole series) having the highest value. Ohio's pattern is quite unique with fluctuating values from the longest periods through around 5–6 weeks. The Fourier power spectrum for the infection rates (Figure 7C) indicates similar periodic patterns as in the periodograms of Figure 7B. These patterns are less prominent due to the adjustment to the same y-axis scale (the scale reflects the magnitude of the positive rates, the shape shows the evolution of the disease).

We conducted bivariate wavelet analysis on the time-lagged data (Figures 8A and S4). The shared synchronicity segments between  $x(t)$  and  $x(t+n)$  can be grouped into shorter periods (around 7 days) and longer periods (approximately 3 weeks, 1 month, 2 months). New York does not display substantial joint short periods. Ohio and Texas mainly have correlation at the end of the series around the 7-day period. Massachusetts experiences joint periodicity around the 7-day period at the early stage of the series. Florida and California have joint periods in the middle of the observation time frame. The levels of average cross-wavelet power are higher in states with poor control ( $x$ -axes scales for Florida, Ohio). The peak power shifts toward higher periodicity with improved control ( $y$ -axes scales for New York, Massachusetts). The return plots in 3 dimensions, utilizing the same time lags as for the countries, seemed to reflect contraction of the attractor in Massachusetts, cyclicity in New York, Florida and California, no containment in Texas, and an ejecting diagonal in Ohio which may reflect exacerbation (Figure 8B). The embedding dimensions varies among states, such that the most contained states (New York, Massachusetts) have the lowest embedding dimension (Table 3).

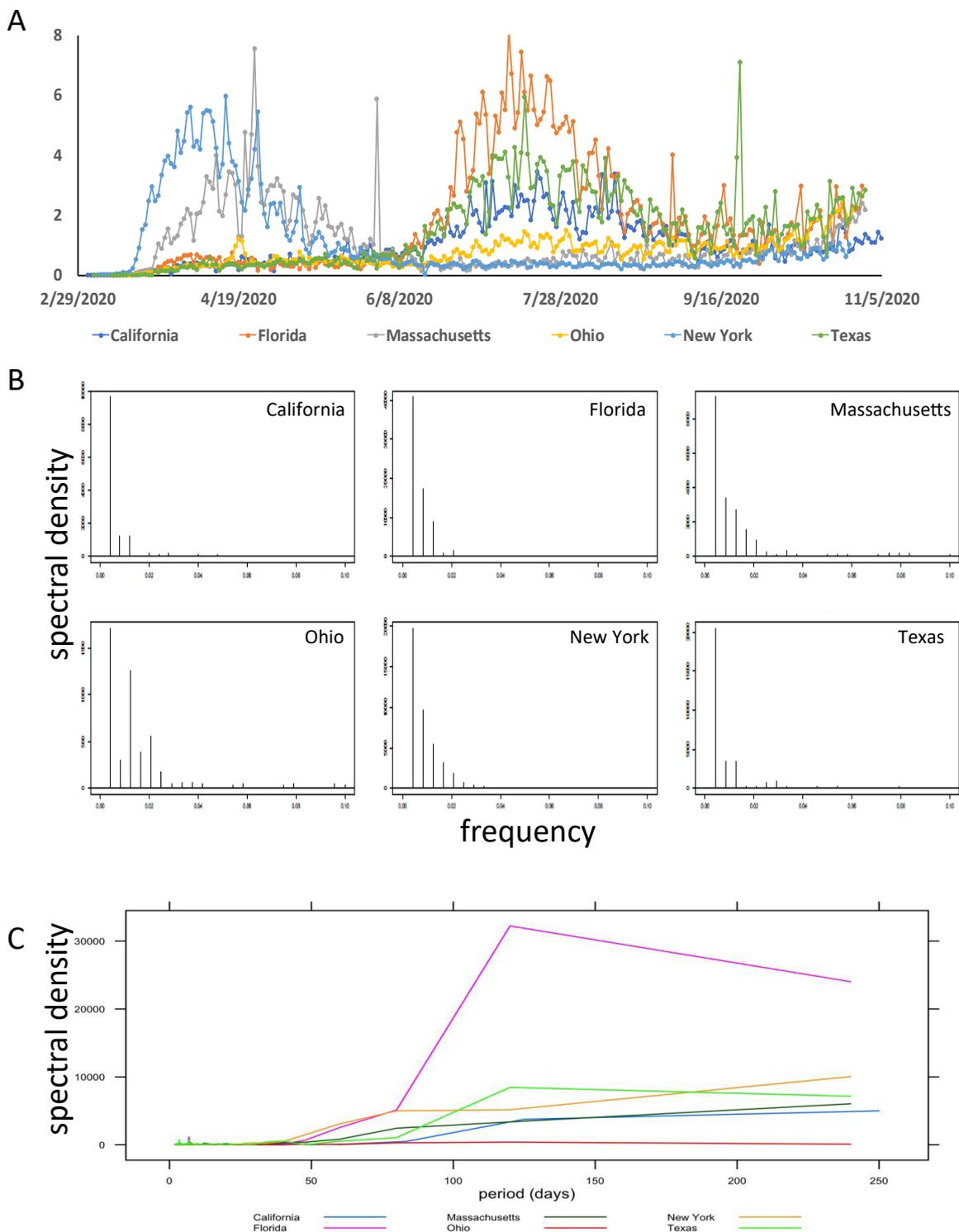
The autocorrelation for return plots of increasing lags show a progressive decline in the numbers of New York and Massachusetts, which implemented strong containment measures after having been afflicted early.



**Figure 5:** Time-course of disease spread for individual US states. Numbers of new cases,  $x(t)$ , per day vs. time ( $t$ , indicating the date). Shown are the curves for (top to bottom, left to right) Massachusetts, New York, Florida, Texas, California, and Ohio.



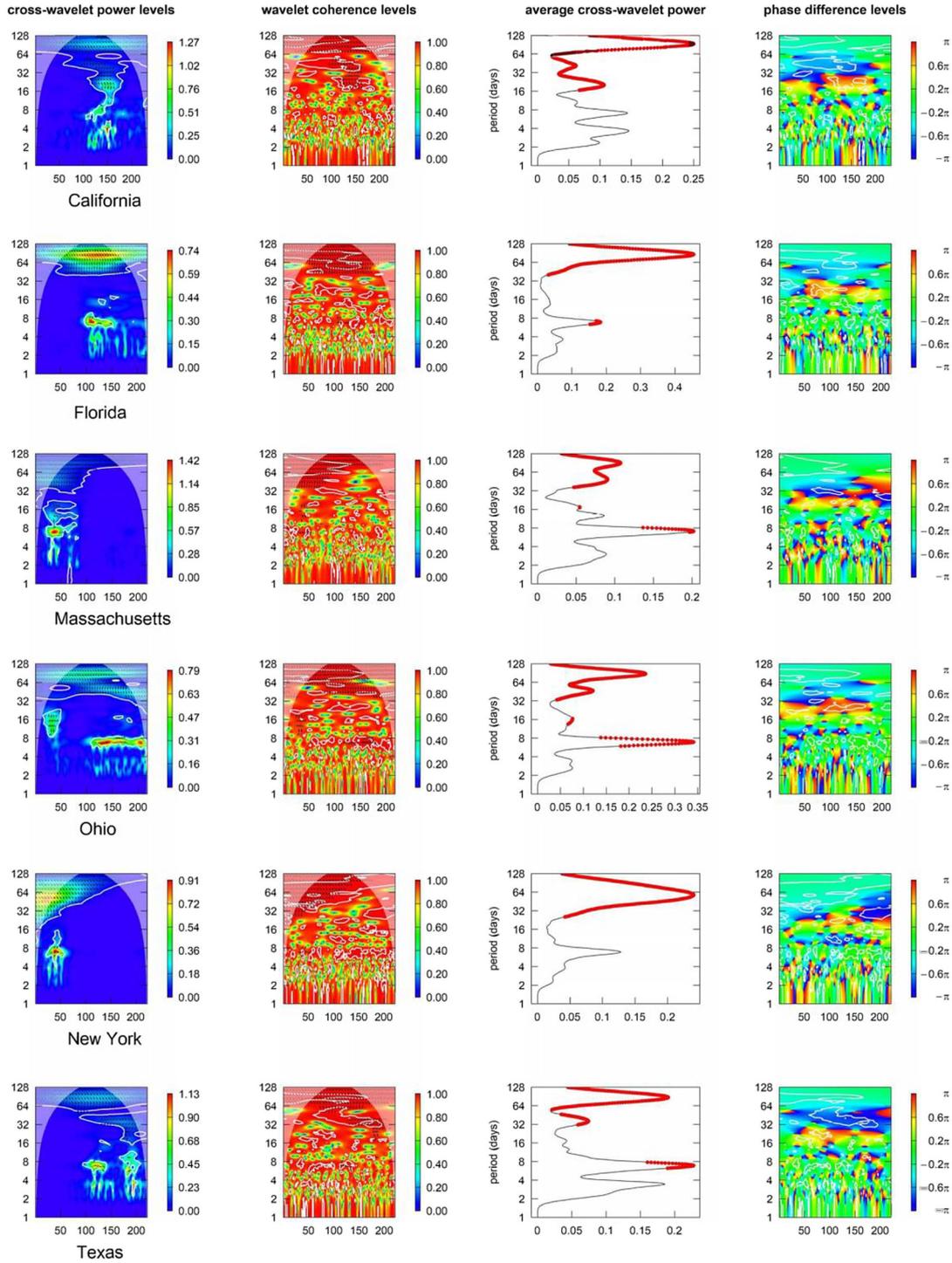
**Figure 6:** Univariate wavelet analysis. Wavelet power spectrum in the time-period domain. Contour lines indicate significance of periodicity with 0.1 significance level. Black lines indicate the ridge of wavelet power. The color bar reveals the power gradient. (A) All states on the same scale. (B) Each state on its own scale.



**Figure 7: Fourier analysis.**

(A) New infection rates. Daily reported new numbers of infections divided by 10,000 inhabitants (infection rates). The x-axis shows the calendar date. (B) Power spectrum. Periodogram plot on the series of the new infection rates. The x-axis is the frequency (per day) and the y-axis represents the spectral density. The y-axis ranges vary among graphs. (C) Normalized power spectrum. Spectral density vs. period (in days) for infection rates. All y-axes have the same scale.

A



**Figure 8:** Time-lagged data analysis by US state.

(A) Bivariate wavelet analysis. Shown are cross-wavelet power plot, wavelet coherence plot, average power plot and phase difference image (from left to right on each row). Time-lagged data were used for  $x(t)$ ,  $x(t + 14)$  (for the lags  $x(t)$ ,  $x(t + 7)$  and  $x(t)$ ,  $x(t + 21)$  see Figure S4). White the contour lines indicate significance of joint periodicity, black arrows indicate the phase difference in the areas with significant joint periods. The solid red dots on the average power plot (the third from the left) reflect significant joint periods at a significance level of 0.1. (B) Return plots in 3 dimensions. Time-lagged return plots in 3 dimensions are shown (from left to right) for  $x(t)$ ,  $x(t + 7)$ ,  $x(t + 14)$  and  $x(t + 3)$ ,  $x(t + 7)$ ,  $x(t + 14)$  and  $x(t + 5)$ ,  $x(t + 14)$ ,  $x(t + 28)$ . Each state under investigation has its own row.

B

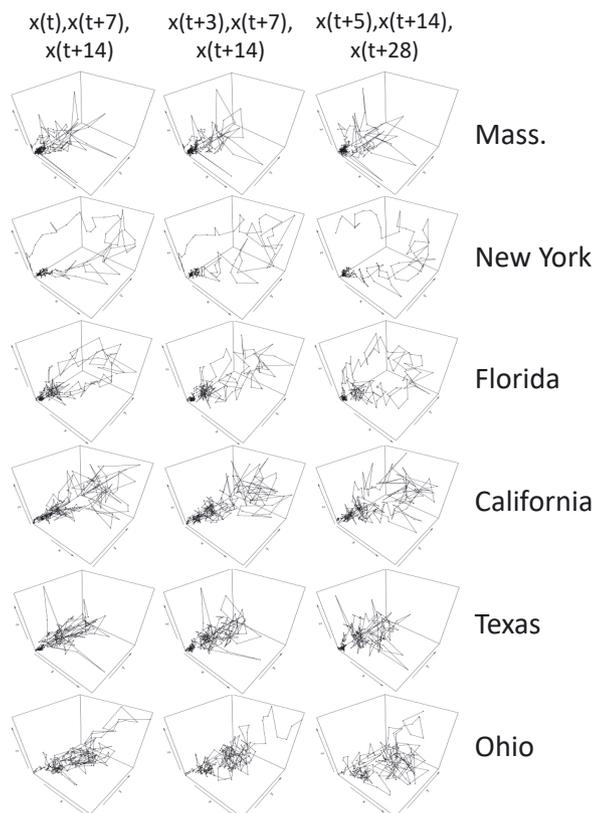


Figure 8: (continued)

Table 3: Embedding dimension for time-lagged data by U.S. state.

State	California	Ohio	Florida	Texas	New York	Massachusetts
Time lag	11	3	3	3	10	7
Embedding dimension	8	8	7	7	6	6

Embedding dimensions were calculated according to Cao's algorithm, which uses 2 functions in order to estimate the embedding dimension from the time series. The table shows the calculated time lags and embedding dimensions for each U.S. state under study.

The values decline less steeply for Texas and California. Ohio displays an anomaly with increasing values for very long lags. The state, while not heavily afflicted on a per capita basis, never achieved containment, only a stationary level, and has since experienced another wave (Table 4A). Up to a maximum lag of 49 days, the AMI for the 6 US states under study ranges between 1.0 and 2.0. Overall, all states show a slightly decreasing pattern except for California, which is relatively leveled at a value of 2.0 (Table 4B). Unexpectedly, the box counting dimension (Table 4C) is less discerning than it was for the evaluation across countries. This may be due to the much lower power conveyed by smaller population sizes.

Table 4: Readouts of complexity for time-lagged data by U.S. state.

A	$t + 1$	$t + 2$	$t + 3$	$t + 4$	$t + 5$	$t + 6$	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$	$t + 42$	$t + 49$
Massachusetts	0.753	0.678	0.672	0.622	0.631	0.663	0.668	0.540	0.314	0.157	0.008	-0.126	-0.136
New York	0.958	0.918	0.896	0.890	0.887	0.887	0.877	0.657	0.426	0.212	0.090	0.019	-0.027
Florida	0.928	0.897	0.879	0.880	0.877	0.894	0.897	0.780	0.583	0.382	0.183	-0.003	-0.151
Ohio	0.910	0.835	0.787	0.750	0.738	0.737	0.733	0.554	0.050	0.319	0.280	0.315	0.303
California	0.876	0.852	0.838	0.841	0.816	0.793	0.837	0.748	0.648	0.515	0.388	0.248	0.107
Texas	0.819	0.739	0.749	0.737	0.723	0.773	0.802	0.693	0.544	0.526	0.414	0.284	0.164
B	$t + 1$	$t + 2$	$t + 3$	$t + 4$	$t + 5$	$t + 6$	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$	$t + 42$	$t + 49$
Massachusetts	1.353	1.269	1.231	1.248	1.177	1.170	1.235	1.060	1.016	1.065	1.058	1.046	1.175
New York	1.689	1.618	1.617	1.580	1.566	1.542	1.582	1.470	1.420	1.423	1.366	1.401	1.480
Florida	1.892	1.857	1.809	1.784	1.785	1.778	1.760	1.590	1.578	1.515	1.466	1.418	1.423
Ohio	1.933	1.823	1.801	1.744	1.717	1.694	1.709	1.501	1.391	1.379	1.366	1.367	1.416
California	2.127	1.993	2.004	2.039	1.976	2.005	2.064	2.000	1.947	2.010	1.960	1.928	1.975
Texas	1.752	1.657	1.709	1.625	1.660	1.743	1.771	1.631	1.603	1.559	1.556	1.391	1.395
C	$t + 7$	$t + 14$	$t + 21$	$t + 28$	$t + 35$								
Massachusetts	1.373	1.411	1.439	1.389	1.396								
New York	1.458	1.464	1.445	1.432	1.439								
Florida	1.489	1.506	1.511	1.594	1.576								
Ohio	1.464	1.511	1.489	1.522	1.522								
California	1.589	1.619	1.623	1.639	1.639								
Texas	1.452	1.464	1.494	1.464	1.494								

6 US states have been evaluated. (A) Autocorrelation. Values of the autocorrelation in COVID-19 spread with each row representing a different US state. The selected time lags are indicated in the top row, all are calculated vs.  $x(t)$ . (B) Average mutual information. Values of AMI in COVID-19 spread with each row representing a different state. The selected time lags are indicated in the top row, all are calculated vs.  $x(t)$ . (C) Fractal dimensions. Box counting dimensions are calculated for 2-dimensional return plots of increasing lags,  $x(t + 7)$  vs.  $x(t)$  through  $x(t + 35)$  vs.  $x(t)$ .

## Discussion

In the present investigation, we find that the analysis tools for observed complex data can aid in the interpretation of pandemic spread across communities. Difficulties in analyzing the non-linear patterns of infectious disease progression may be tamed by applying the methods of complex systems research. The approach can reveal patterns, where a simple time course of new cases does not (Supplement Figure S6 (note added in proof)), and it does not require the formulation of models and algorithms upfront. Fractal dimension can be used to measure the complexity of the dynamics in a time series. Its graphic depiction may inform on rhythmicity/periodicity, and confinement is reflected in a shrinking/movement by the attractor toward the origin of the graph. Autocorrelation and AMI find repeating patterns in the time series. Harmonics can be an indicator of exacerbation. The operational approximation of Lyapunov exponents may be meaningful, although they were largely uninformative for the present study (Supplemental Figure S5). Fuzzy Logic is meaningfully added to represent the uncertainty in the process of making a forecast (Castillo and Melin 2020; Castillo and Melin 2021). Further, non-linear analysis allows the study into various facets of the process, depending on whether the starting data are new cases, hospitalizations, deaths or other readouts. Maps can be generated and evaluated for their fractal dimensions (Păcurar and Necula 2020). Computer programs are available or can be written to update pertinent readouts whenever new data are received. We expect that the tracking of infectious progressions with these tools of complex systems analysis will provide improved readouts for the state of the crisis management and will enable more informed decision making.

Among the countries analyzed, South Korea has had the most successful control of the pandemic spread according to low intensity in univariate wavelet analysis, low spectral density range in Fourier analysis, low spectral density of the normalized rates, a reduction in cross-wavelet power levels according to bivariate wavelet analysis and longer periodicity in the distribution of cross-wavelet power. Further, declining box counting dimensions, autocorrelation values with increasing time lag, and decreasing trends (at a low slope) in AMI confirm containment. Cyclical patterns in return plots, suggesting the outlines of attractors, are apparent and most data points are concentrated near the origin of the graph. Germany exhibited good management through October 2020 according to univariate wavelet analysis, spectral density in the power spectrum of the normalized rates, a reduction of cross-wavelet power levels in bivariate wavelet analysis, longer periodicity in the distribution of cross-wavelet power, a dramatic lowering of autocorrelation values at a lag of 21 and above, and relatively flat AMI values, staying around levels of 1.5. Cyclical patterns in return plots suggest the outlines of an attractor. Good control by Italy consecutive to the early impact and through October 2020 is reflected in low intensity and fluctuation when applying univariate wavelet analysis, in a reduction of cross-wavelet power levels for bivariate wavelet analysis of time-delayed data, longer periodicity in the distribution of cross-wavelet power, and decreasing trends (at a low slope) in AMI. Cyclical patterns in return plots, suggesting the outlines of an attractor, are apparent. Poland had two distinct phases. By univariate wavelet analysis and density in the power spectrum of normalized rates, there was indication of good management through October 2020. According to bivariate wavelet analysis for time-delayed data series and return plots, the recent substantial upswing in new infections is reflected, which also results in box counting dimensions close to 1. From a lag of 6 onward, Poland has substantially declining autocorrelation values, although due to the recent steep upswing in new infections, the trend reverses at very long lags. The AMI starts with a relatively low value (1.15 at  $t$  vs.  $t + 1$ ) and shows a rapid decrease with longer lag, staying level from lags of 21–49 days. In the United States, univariate wavelet analysis displays cyclical fluctuations of various durations, none of which have been contained. According to bivariate wavelet analysis for time-delayed data series, there have been protracted periods of high cross-wavelet power levels. Cyclical patterns in return plots, suggesting the outlines of attractors, are apparent. The USA expresses relatively flat AMI values, staying around levels of 2.20. In France, univariate wavelet analysis of the time course shows prominence of the recent upswing (heat intensity on the right margin of the graph), the power spectrum is reflective of potentially adverse developments. The second wave of infections is apparent in bivariate wavelet analysis and in the obscured order in return plots. France displays a gradually decreasing trend of AMI. India expresses cyclical fluctuations

of various durations in univariate wavelet analysis, none of which have been contained. On a population bases, the spectral density suggests that control has not been lost through October 2020. Bivariate wavelet analysis shows protracted periods of high cross-wavelet power levels, return plots reflect the progressive increase in new cases over the time period in a predominantly linear curve on each scale, box counting dimensions are close to 1, and autocorrelation values stay uniformly high with increasing time lag. India displays a gradually decreasing trend of AMI. Brazil experiences cyclical fluctuations of various durations in univariate wavelet analysis, none of which have been contained. By Fourier analysis, the spectral density range is high. The power spectrum is indicative of potentially adverse developments. According to bivariate wavelet analysis, there have been protracted periods of high cross-wavelet power levels. In return plots, the wide fluctuations generate a largely disordered appearance. The autocorrelation values stay uniformly high. Brazil expresses relatively flat AMI values, staying around levels of 2.20. Sweden shows cyclical fluctuations of various durations in univariate wavelet analysis, none of which have been contained. The power spectrum is reflective of potentially adverse developments. In return plots, disorder is apparent. Sweden expresses relatively flat AMI values.

Prima facie, the curves of new infections vs. time for three Western European countries, France, Italy, and Germany, appear similar. Complex systems analysis reveals the upswing in France to be much more perilous than the increases in the curves of new infections by the other two countries. The management of infectious spread also requires improvements in the United States, Sweden and Brazil. The selection of the observation period can dramatically influence the results. Poland was initially very successful in containing the pandemic, but then experienced a substantial upswing. Analyzing these two phases individually or in conjunction yields very different data sets, which inform about distinct aspects of the infectious progression.

The fluctuations of new infections in an epidemic or a pandemic pose challenges to the evaluation whether a decline reflects true containment (“rounding the corner”) or just the calm before another wave. The readouts of non-linear systems analysis can aid in making such a distinction. A complex occurrence that experiences containment will strive toward a point attractor in phase space and move toward the origin. Such a progression is represented in a declining fractal dimension, and the transition from fluctuations (often associated with a torus attractor, where the time series does not have a strict periodicity, the trajectory is not closed, and the limit cycle acquires the shape of a torus.) toward limitation of new cases is expected to reduce the autocorrelation.

One constraint of complex systems analysis is the need for large data sets. In this regard, the availability of about 230 data points (daily new cases March through October 2020) for each geographic area in this study is somewhat low. The robustness of pertinent studies increases with larger data sets over time. Reporting errors could have a non-trivial impact, and may be reflected in the frequent occurrence of a peak at 7 days in the spectral analysis (possibly indicating weekly totals). This problem can be addressed by utilizing moving averages. The homogeneity or heterogeneity in management by the community under study determines the noise level. The worldwide numbers of new infections have a lot of background due to varying patterns across countries.

**Research funding:** GFW is supported by NIH grant CA224104.

**Author contribution:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Ethical approval:** The local Institutional Review Board deemed the study exempt from review.

## References

- Abarbanel, H. D. I. 1995. *Analysis of Observed Chaotic Data*. Switzerland: Springer Nature.
- Al-Raei, M., M. S. El-Daher, and O. Solieva. 2021. “Applying SEIR Model without Vaccination for COVID-19 in Case of the United States, Russia, the United Kingdom, Brazil, France, and India.” *Epidemiologic Methods* 10: 20200036.
- Bertacchini, F., E. Bilotta, and P. S. Pantano. 2020. “On the Temporal Spreading of the SARSCoV-2.” *PLoS One* 15: e0240777.

- Bin, S., G. Sun, and C.-C. Chen. 2019. "Spread of Infectious Disease Modeling and Analysis of Different Factors on Spread of Infectious Disease Based on Cellular Automata." *International Journal of Environmental Research and Public Health* 16: 4683.
- Blasius, B. 2020. "Power-Law Distribution in the Number of Confirmed COVID-19 Cases." *Chaos* 30: 093123.
- Chakraborty, I., and P. Maity. 2020. "COVID-19 Outbreak: Migration, Effects on Society, Global Environment and Prevention." *The Science of the Total Environment* 728: 138882.
- Christakis, N. A. 2020. *Apollo's Arrow: The Profound and Enduring Impact of Coronavirus on the Way We Live*. New York: Hachette Book Group.
- Cao, L. 1997. "Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series." *Physica D: Nonlinear Phenomena* 110: 43–50.
- Castillo, O., and P. Melin. 2020. "Forecasting of COVID-19 Time Series for Countries in the World Based on a Hybrid Approach Combining the Fractal Dimension and Fuzzy Logic." *Chaos, Solitons & Fractals* 140: 110242.
- Castillo, O., and P. Melin. 2021. "A Novel Method for a COVID-19 Classification of Countries Based on an Intelligent Fuzzy Fractal Approach." *Healthcare* 9: 196.
- Groendyke, C., and A. Combs. 2021. "Modifying the Network-Based Stochastic SEIR Model to Account for Quarantine: An Application to COVID-19." *Epidemiologic Methods* 10: 20200030.
- Mehta, M., J. Julaiti, P. Griffin, and S. Kumara. 2020. "Early Stage Machine Learning—Based Prediction of US County Vulnerability to the COVID-19 Pandemic: A Machine Learning Approach." *JMIR Public Health and Surveillance* 6: e19446.
- Păcurar, C.-M., and B.-R. Necula. 2020. "An Analysis of COVID-19 Spread Based on Fractal Interpolation and Fractal Dimension." *Chaos, Solitons & Fractals* 139: 110073.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Also available at <https://www.R-project.org/>.
- Roesch, A., and H. Schmidbauer. 2018. *WaveletComp: Computational Wavelet Analysis. R Package Version 1.1*. Also available at <https://CRAN.R-project.org/package=WaveletComp>.
- Stehlé, J., N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J. F. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems. 2011. "Simulation of an SEIR Infectious Disease Model on the Dynamic Contact Network of Conference Attendees." *BMC Medicine* 9: 87.
- Sun, T., and Y. Wang. 2020. "Modeling COVID-19 Epidemic in Heilongjiang Province, China." *Chaos, Solitons & Fractals* 138: 109949.
- Wang, K., L. Ding, Y. Yan, C. Dai, M. Qu, D. Jiayi, and X. Hao. 2020. "Modelling the Initial Epidemic Trends of COVID-19 in Italy, Spain, Germany, and France." *PloS One* 15: e0241743.
- White, E. R., and L. Hébert-Dufresne. 2020. "State-Level Variation of Initial COVID-19 Dynamics in the United States." *PloS One* 15: e0240648.

---

**Supplementary Material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/em-2021-0019>).