



A kernel-modulated SIR model for Covid-19 contagious spread from county to continent

Xiaolong Geng^a, Gabriel G. Katul^{b,c}, Firas Gerges^{a,d}, Elie Bou-Zeid^e, Hani Nassif^f, and Michel C. Boufadel^{a,1}

^aDepartment of Civil and Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102; ^bNicholas School of the Environment, Duke University, Durham, NC 27710; ^cDepartment of Civil and Environmental Engineering, Duke University, Durham, NC 27708; ^dDepartment of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102; ^eDepartment of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08540; and ^fDepartment of Civil and Environmental Engineering, Rutgers University–New Brunswick, Piscataway, NJ 08854

Edited by Andrea Rinaldo, École Polytechnique Fédérale, Lausanne, Switzerland, and approved March 27, 2021 (received for review November 9, 2020)

The tempo-spatial patterns of Covid-19 infections are a result of nested personal, societal, and political decisions that involve complicated epidemiological dynamics across overlapping spatial scales. High infection “hotspots” interspersed within regions where infections remained sporadic were ubiquitous early in the outbreak, but the spatial signature of the infection evolved to affect most regions equally, albeit with distinct temporal patterns. The sparseness of Covid-19 infections in the United States was analyzed at scales spanning from 10 to 2,600 km (county to continental scale). Spatial evolution of Covid-19 cases in the United States followed multifractal scaling. A rapid increase in the spatial correlation was identified early in the outbreak (March to April). Then, the increase continued at a slower rate and approached the spatial correlation of human population. Instead of adopting agent-based models that require tracking of individuals, a kernel-modulated approach is developed to characterize the dynamic spreading of disease in a multifractal distributed susceptible population. Multiphase Covid-19 epidemics were reasonably reproduced by the proposed kernel-modulated susceptible–infectious–recovered (SIR) model. The work explained the fact that while the reproduction number was reduced due to nonpharmaceutical interventions (e.g., masks, social distancing, etc.), subsequent multiple epidemic waves still occurred; this was due to an increase in susceptible population flow following a relaxation of travel restrictions and corollary stay-at-home orders. This study provides an original interpretation of Covid-19 spread together with a pragmatic approach that can be imminently used to capture the spatial intermittency at all epidemiologically relevant scales while preserving the “disordered” spatial pattern of infectious cases.

coronavirus disease | susceptible–infectious–recovered approach | multifractals | Fourier analysis | population agglomeration

Covid-19 has become a serious global health threat due to its rapid spatial spread. In the United States alone, as of February 2021, over 480,000 deaths and nearly 28,000,000 infected cases have been confirmed since the outbreak of the disease in March 2020 (1). The tempo-spatial evolution of Covid-19 cases involves complex epidemiological dynamics as a result of nested personal, societal, and political decisions. Numerous studies have identified spatially heterogeneous distributions of Covid-19 cases and deaths over a wide range of scales (1–3). Thus, an understanding of the Covid-19 outbreak from a spatial perspective is necessary (though not sufficient) to evaluate the nationwide spread of disease and delineate the correlations across different regions. Geographic information system–based spatial analyses have been performed, aiding in the visualization of how the spatial structure and interconnections between the US cities or states impact the incidence of the disease (4–7). However, these studies are primarily focused on relating spatial variability of the disease incidence to exogenous environmental, socioeconomic, and demographic factors. Endogenous mechanisms of the epidemic spreading across the continental United States over a wide range of scales still remains to be uncovered.

The risk of inoculation against smallpox marked the birth of “static” epidemiological models, which have been developed by Jean-Baptiste le Rond d’Alembert and Daniel Bernoulli in the 18th century (8). Moving from a static compartment framework toward a dynamic representation of disease spread has been a long and tortuous path achieved during the early 1920s (9). This framework is now routinely used to describe disease transmission for planning effective control strategies (10–15). The backbone of this framework is a dynamical system labeled as the susceptible–infectious–removed, or recovered (SIR), approach, in which susceptible individuals could become infected and are eventually “removed” from the pool through recovery or death (15). A key parameter of such models is a reproduction number, which has been identified as an essential measure of the risk of an infectious agent with respect to epidemic spread. There are multiple reproduction numbers in use depending on the particulars of model formulation: in a closed system, the basic reproduction number (R_0) defines the average number of secondary infections produced by a typical case of an infection in a population where all the individuals are treated as susceptible. However, due to immunity from past exposures or vaccination, or any deliberate intervention to curtail disease transmission, the actual average number of secondary infections per infectious case would be lower than the basic reproduction number; this is often referred to as the effective reproduction number (R_e). Numerous studies have been conducted to estimate the reproduction number since the outbreak of coronavirus disease. An uncontrolled $R_0 = 4.5$ across many nations was identified in the early phases before any interventions when

Significance

Spatial analysis of daily Covid-19 cases at the US county scale revealed a dynamic multifractal scaling of infections, spanning from 10 to 2,600 km and consistently trending toward that of the susceptible population. A susceptible–infected–recovered model was expanded to include spatial spread across counties using a spatial kernel. The reproduction number R_b (average number of persons infected by an infected person) decreased because of interventions (masks, social distancing). The model shows that reducing R_b in isolation is not sufficient to stem the spread of the disease and concomitant measures such as curfews and lockdowns may be needed. The R_b of 2.0 estimated here in July to October 2020 is large, hinting at super-spreaders and super-spreader events.

Author contributions: X.G., G.G.K., and M.C.B. designed research; X.G. and H.N. performed research; F.G. and E.B.-Z. contributed new reagents/analytic tools; X.G., G.G.K., F.G., H.N., and M.C.B. analyzed data; and X.G., G.G.K., E.B.-Z., and M.C.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: boufadel@gmail.com.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2023321118/-DCSupplemental>.

Published May 6, 2021.

setting the mean recovery time to 14 d (16). Time-varying estimation of R_e has also been conducted, which often requires information on the distribution of the serial interval and number of daily cases (17–20). A daily updated estimation of R_e shows that the reproduction number is still larger than 1.0 in most of the US states till the end of 2020 regardless of continuous precaution and prevention of the disease (<https://rt.live/>). Modeling the Covid-19 spread has been conducted and published since March 2020 (21–23). These studies fitted time series of infectious individuals for specific locations following the outbreak of Covid-19 in China. Movement of individuals between cities was introduced by linking the SIR framework to agent-based network models (24, 25). Empirical models such as the one adopted by the University of Washington were also routinely employed, which often need to be regularly calibrated, including recalibration of R_e , to reproduce the newly acquired data and anticipated policies (26). Some of the models are developed for representing Covid-19 data at a large

scale (e.g., a metropolitan or continental scale) and therefore become impractical as the spatial resolution is increased to below a certain scale (e.g., 100 km at the county scale) due to a proportional relationship between the connectivity and the squared number of communities. In addition, there are challenges in acquiring sufficient traveler’s data for such agent-based network models due to confidentiality concerns with the tracking and usage of an individual’s mobility information (27, 28). These models might also encounter equifinality problems in which different combinations of parameter values generate equal results when the number of connections between locations increases. Diffusion dynamics have a long and proven tradition to represent human mobility and the spatial transmission of the disease among local-scale neighboring regions. Their success is featured by predicting a spread velocity representing an epidemic wave front (29, 30). Finally, neither of these models incorporate the intrinsic properties of the population nor the spatial structure of the disease,

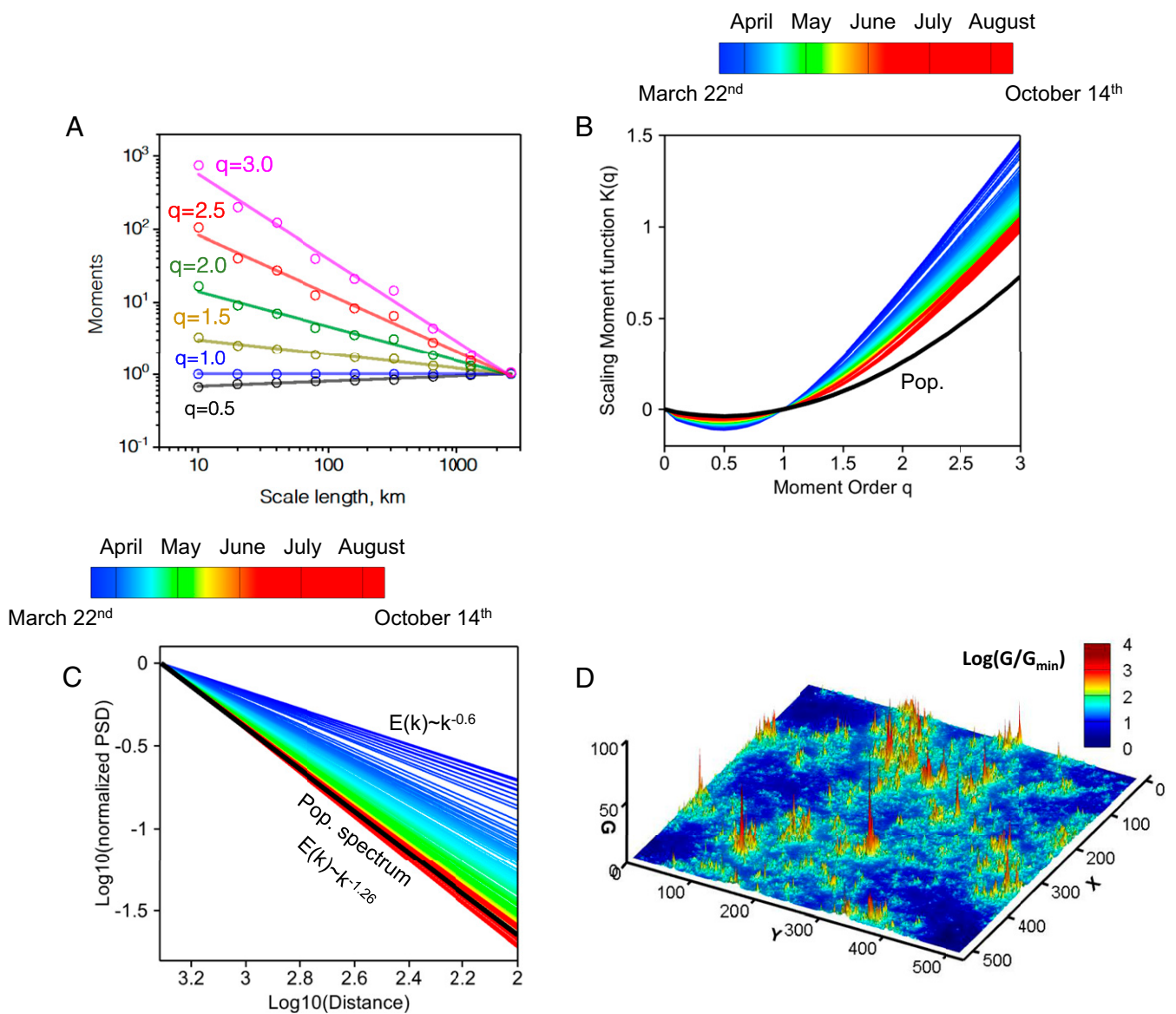


Fig. 1. (A) The q moments of the Covid-19 cases and their scaling from 10 up to 2,600 km used in the determination of $K(q)$. (B) The moment scaling function $K(q)$; the convex shape indicates multifractality. (C) The fitted spectral slope at various dates. (D) Example multifractal field generated using the UM model using the parameter values estimated from the population. Notice that both $K(q)$ and Fourier spectral slope approach that of the underlying population, indicating increased spatial correlation that trends toward a “saturated” threshold set by the population agglomeration.

both attributes important for stability of prediction and for intervention planning.

Herein, spatial analysis of the Covid-19 infectious cases was performed in the United States at the county scale over time, revealing a dynamic multifractal scaling in the spatial correlation. This scaling spans from 10 to 2,600 km, consistently trending toward that of the population (Fig. 1). A kernel-modulated SIR model mapped onto a multifractal population framework is developed to explore tempo-spatial variation patterns of the infections across the continental United States (Fig. 2). The susceptible population was viewed as consisting of multiple segregated sets, representing quarantine and mobility, to reflect the multistage conditional reopening policy implemented late April to early June 2020 in the United States. The model also reliably reproduces observed long-range spatial correlation of Covid-19 cases evolving with time, compared to diffusion-modulated and immobile (i.e., each county is assumed to be a closed compartment following early infections) SIR models (Fig. 3). Sensitivity tests indicate that the evolution patterns of Covid-19 cases for all the US states could be generally retrieved from the model (Fig. 4). Of significance is that the proposed model preserves the so-called “disordered” infectious patterns featured by boundaries delineating hotspots of infections from disease-free regions. Such a representation unravels the interaction among spatial scales responsible for the spread of the disease across the continental United States.

Results

Daily moment analysis of the infected Covid-19 cases during March 22 to October 14 2020 reveals a multifractal scaling that spanned from 10 to 2,600 km (Fig. 1A and *SI Appendix, Fig.*

S1–S8). The calculated orders of the moments demonstrate large divergence early in the outbreak through March to April and become nearly identical at all spatial scales after May, which suggests that the spatial correlation by then approached a stationary set. This is because after travel restrictions and corollary stay-at-home orders, the spatial evolution of infections was primarily controlled by the spatial patterns of the population. As the spatial patterns of the Covid-19 cases approach that of the population, the underlying statistics of the spatial spread of the disease tends to become stationary, set by the spatial distribution of the population. The human population was also identified to be multifractal distributed (*SI Appendix, Fig. S1*), which is consistent with prior findings (31). We revealed herein that the multifractal scaling of the population and Covid-19 cases operated over the same range of spatial scales. Power-law scaling was also reported for several properties of urban systems such as city sizes (32). Signatures of multifractal characteristics are routinely detected using the convexity of the so-called moment scaling function $K(q)$ for moment order q . Three parameters describe the mathematical properties of $K(q)$ in the universal multifractal (UM) model (33, 34): α , the Levy index measuring departure from monofractal behavior defined by $\alpha=0$; $c1$, the codimension parameter measuring the degree of intermittency; and H , the Hurst exponent linked to the degree of scale invariant smoothing. For the US population, $K(q)$ shows a convex shape (upward looking), indicating a multifractality of the population variation over space (*SI Appendix, Fig. S1*). The population α , $c1$, and H are estimated to be 1.78, 0.15, and 0.21, respectively. The convex shape of the observed daily moment scaling exponents for the Covid-19 cases is consistently calculated over the entire period of our study from

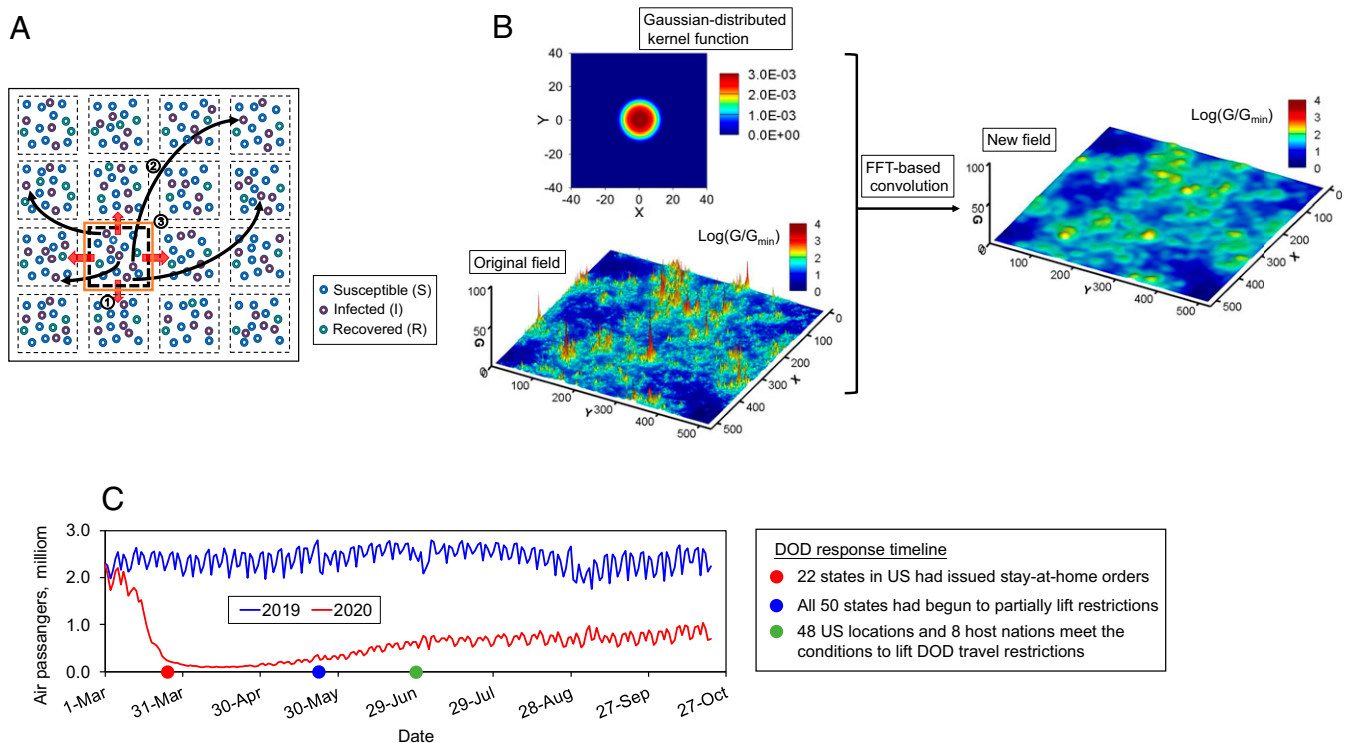


Fig. 2. (A) Schematics of kernel-based model, diffusion-based model, and immobile model. Label 1 represents the spreading mechanism of the diffusion-based model that is driven by the concentration gradient between two adjacent cells. Label 2 represents a more realistic spreading mechanism of the kernel-based model that allows individuals to “travel” beyond one cell with a Gaussian-distributed probability. Label 3 represents the spreading mechanism of the immobile model that each cell is well mixed but isolated from other cells. (B) Example multifractal field and the new field postprocessed by the Fast Fourier Transform–based convolution using the Gaussian-distributed kernel function. (C) TSA check point travel numbers for 2020 and 2019 (<https://www.tsa.gov/coronavirus/passenger-throughput>) and major US Department of Defense response timeline (<https://www.defense.gov/Explore/Spotlight/Coronavirus/DOD-Response-Timeline/>).

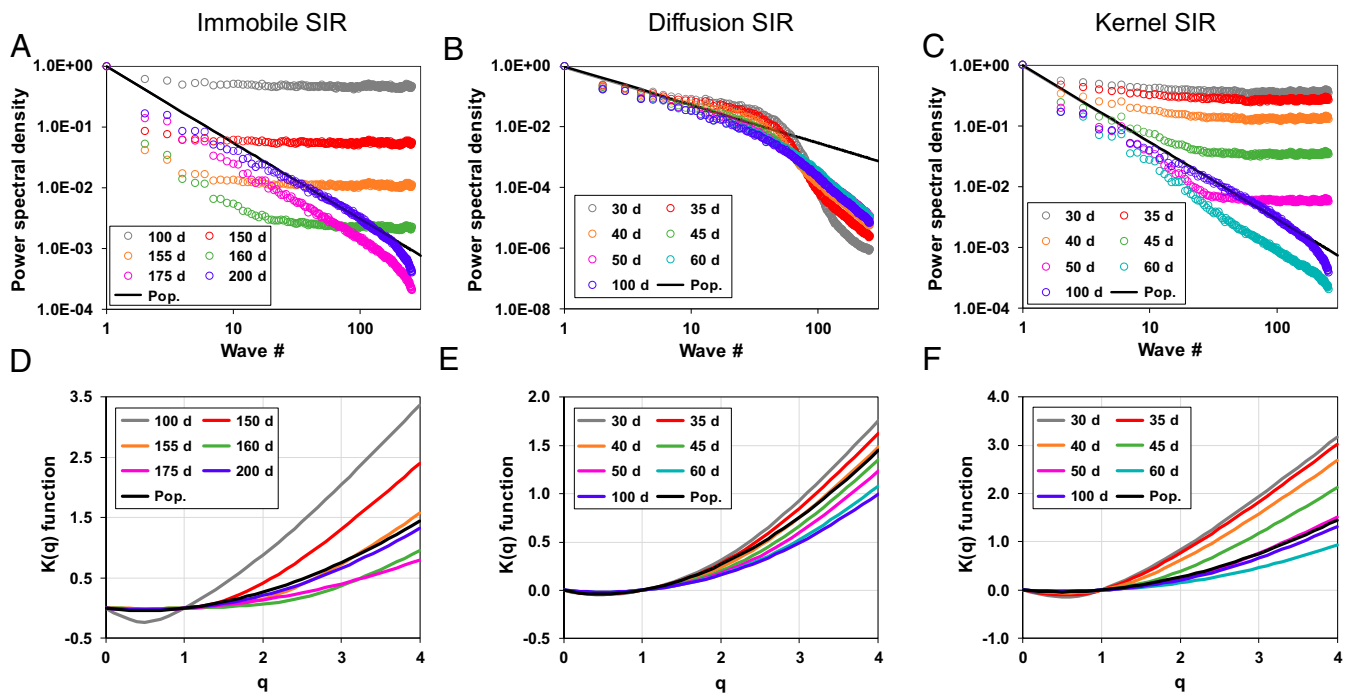


Fig. 3. Temporal evolution of (A–C) power spectral density function and (D–F) moment scaling function $K(q)$ for infected cases derived from the kernel-based SIR model, diffusion-based SIR model, and immobile SIR model, respectively. The kernel-based SIR model (C and F) better reproduced the spatial correlation of geographic spread of Covid-19 evolving to that of the underlying population agglomeration compared to the diffusion-based model and accelerates the spatial correlation trending toward the “saturation” threshold set by the population agglomeration.

late March to early October as well (Fig. 1B). Similar to the moments, the calculated scaling moment function $K(q)$ evolves in time and approaches a nearly stationary set after June ($\alpha = 1.88$, $c1 = 0.18$, and $H = 0.33$). These results underscore the multifractal properties of the spatial spread of the infected cases during the Covid-19 outbreak.

The Fourier power spectrum’s exponent of the Covid-19 cases (κ) temporally increases from 0.54 to 1.31, trending toward that of the population value $\kappa = 1.26$ (Fig. 1C). The exponent’s increase is rapid until mid-May and then slows down. A temporal increase in κ reflects the fact that the Covid-19 cases tended to be more spatially correlated with time and eventually became spatially “distributed” following the spatial correlation structure of the local population. Such increase in κ with time reflects the decrease in $K(2)$ (Fig. 1B) as well as the increase in H (Eq. 5), which experienced an uptick from 0.14 in later March to 0.33 in early October 2020. The multifractality parameter α of the Covid-19 cases remains invariant at a value of 1.78 ± 0.10 (SI Appendix, Fig. S9) and is commensurate with that of the population. In particular, an $\alpha < 2$ indicates that the underlying statistics of the spread of the disease is Levy based and “fatter” than lognormal. The decrease in α implies the increase in the likelihood of occurrence of large numbers of infections. The parameter $c1$ represents the degree of spatial intermittency (i.e., on–off activity patterns in space). For the Covid-19 cases, $c1$ decreases over time from 0.42 to 0.18 on October 14 2020, which is higher than the population $c1 = 0.15$. As expected, the infectious population is more spatially intermittent than the overall population across counties.

Fig. 1D illustrates an example multifractal field created using the UM model, taking the multifractality of the spatial population distribution as a template for the susceptible groups; the existence of local-scale spikes implies that the infectious field behaves non-Gaussian. The simulation results conducted using the SIR model integrated with a multifractal population framework

demonstrate that the kernel-modulated approach could reproduce the spectral slope of the total infected cases evolving to that of the population with time at all spatial scales (Fig. 3). For the kernel-modulated model, the spectral slope tends to be flat at early stages of the disease spread (e.g., within the first 40 d) (Fig. 3C). Infected cases were initially randomly distributed in space (random attack points on the multifractal susceptible population across counties) and therefore are expected to be less correlated over space. As the disease outbreak progresses by spatial spreading across counties, the spectral slope increases and eventually approaches that of the population. Similar behavior is also observed for simulated $K(q)$ function (Fig. 3F). By comparison, the spectral slope for the diffusion-based model converges to that of the population faster at large scales but becomes steeper than that of the population at small scales (Fig. 3B). Such a break in the spectral slope is attributed to the diffusion mechanism that allows complete mixing among adjacent regions and therefore elevates the spatial coherence at small scale (whose signature is a steep spectral slope). In addition, the convergence behavior of $K(q)$ to that of population is unlikely to be observed from the diffusion-modulated model (Fig. 3E). The simulation results from the immobile model exhibiting the convergence behavior for both spectral slope and $K(q)$ but with a slow pace compared to the kernel-modulated model (Fig. 3A and D). This finding is expected because, for the immobile model, the disease only spreads within each compartment, and the spread of the disease across compartments (here, counties) is prohibited.

Temporal evolution of the Covid-19 cases for each US state along with simulated results are reported in Fig. 4A, spanning the period from February 15 to October 14 2020. The simulation was conducted state by state using the kernel-modulated SIR model, taking into account the multiphase population release mechanism. The calibrated population flow and model parameters are reported in SI Appendix, Fig. S10 and Table S1. The state-level evolution of the cases exhibits distinct variations, which can be

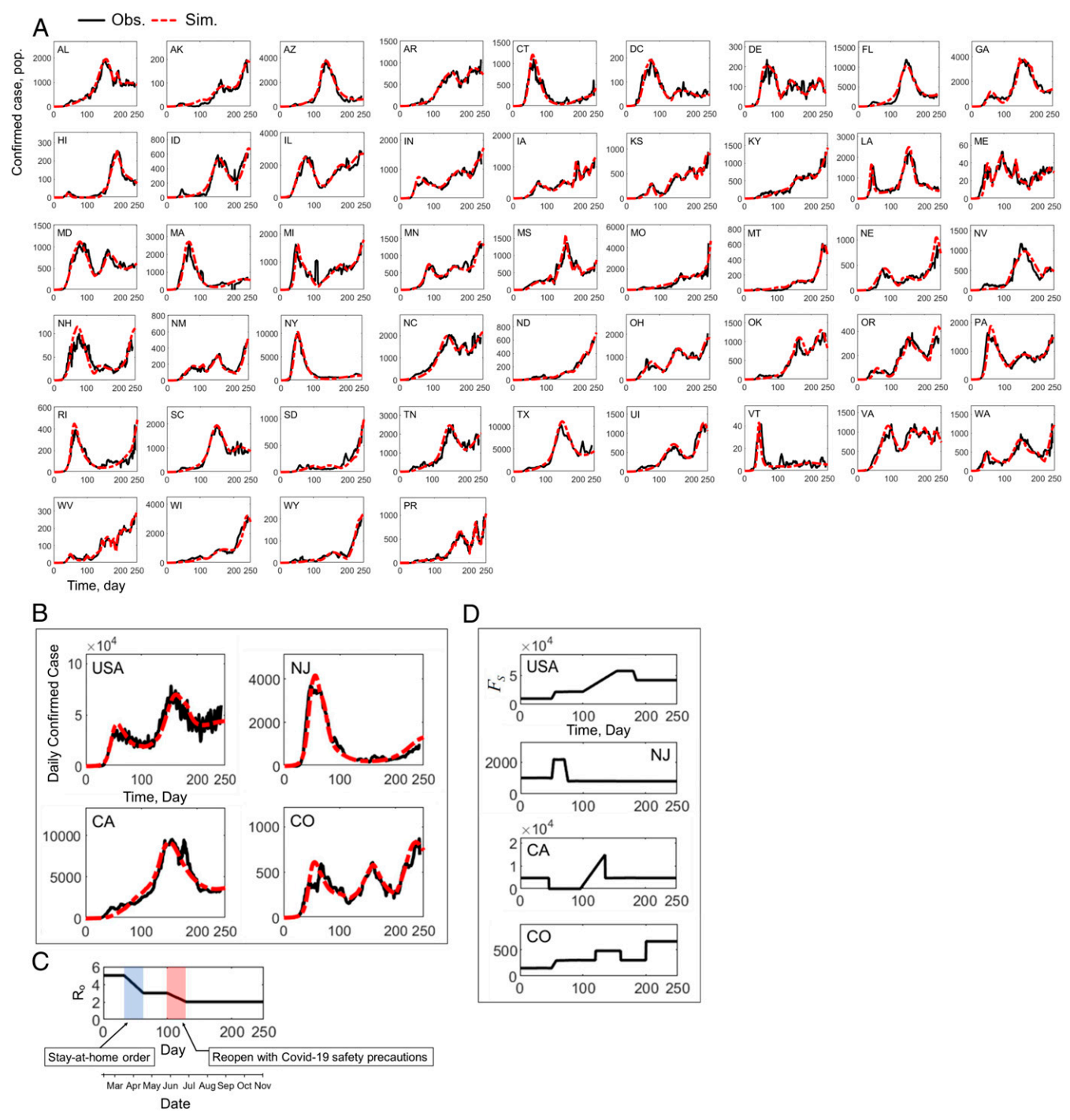


Fig. 4. (A) Observed daily Covid-19 cases for the United States and each US state. Note that the time origin is Feb. 15, 2020, and 7 d moving average is taken for each US state. (B) Simulated Covid-19 cases using the kernel-based SIR model, taking into account the multiphase population release. Here, we selected daily confirmed cases reported in the United States, New Jersey, California, and Colorado as representative scenarios showing two peaks occurring before and after the reopening, with a single peak occurring before the reopening, a single peak occurring after the reopening, and multiple peaks, respectively. Note that the model acceptably reproduces the US Covid-19 cases. (C) Values of R_b used for the simulations. (D) Calibrated piecewise linear function F_s .

loosely characterized by the number of the infectious peaks and the period of their occurrence. For example, the states of Connecticut, Massachusetts, New Jersey, and New York show a single peak within the first 100 d after the outbreak, while the states of Alabama, California, Florida, South Carolina, and Texas show a single peak after the travel restrictions were partially lifted. In contrast, the states of Louisiana, Maryland, New Hampshire, and Utah exhibits two peaks before and after travel restriction was

partially lifted. Some states such as Colorado, Delaware, New Mexico, Ohio, Washington, and Puerto Rico even show multiple peaks, which can be attributed to a multistage conditional reopening policy. These distinct patterns are linked to the timeline of each state's policy and regulation regarding reopening. For instance, the state of Alabama ordered residents to stay home for much of April and began reopening business starting in May. The state of Colorado was among the first states to begin reopening as well, letting

Downloaded at Palestinian Territory, occupied on November 27, 2021

stay-at-home orders expire on April 26; however, in response to continuing outbreaks in neighboring states, the state's "Safer at Home" phase was extended in July. In September, a new framework to open the state of Colorado was introduced on a more regional basis. In contrast, the state of New Jersey was the last state to lift the stay-at-home order on June 9. Nevertheless, temporal evolution of the Covid-19 cases at the country scale, taking into account all regions, so far manifests two peaks occurring in April and July, respectively, and starts trending toward its third peak since September (Fig. 4B). Such temporal evolution patterns support the multiphase population release mechanism manipulated by levels of travel restrictions and stay-at-home phases.

By introducing a multiphase population release mechanism, the kernel-modulated SIR model reproduces the overall trend of the Covid-19 cases evolving with time in all US states (Fig. 4 and *SI Appendix*, Fig. S10). This representation accommodates for mobility across counties through ϕ , the percentage of infections migrating across counties, the disease epidemiological properties, and the population response to them through a reproduction number R_b (that differs from R_0 or R_e) and inverse recovery times γ_R , spreading kernel shape $p(x, y)$ in space, and variations in susceptible population flow from the general population to the susceptible class in response to county-level lockdowns or stay-at-home orders. Thus, the evolution pattern of Covid-19 cases can be captured through an appropriate calibration of R_b reflecting the effectiveness of the precautionary behavior at the individual scale and the susceptible population flow (i.e., F_S) (*Materials and Methods*). Sensitivity tests further lend credence to the peak numbers and the period of their occurrence after the initial outbreak of Covid-19. It shows that the evolution pattern of Covid-19 cases is primarily driven by the population flow (i.e., F_S), although other factors (e.g., ϕ and R_b) alter the magnitude of the peaks and shift the period of their occurrence (*SI Appendix*, Fig. S11). In particular, sensitivity tests of R_b values for the period when the stay-at-home order was relaxed show reducing R_b is unlikely to prevent a subsequent outbreak of the disease due to relatively large releases of the "susceptible" population back into the "SIR" pool (*SI Appendix*, Fig. S12). A critical control of the population flow immediately after the disease outbreak mitigated the early-stage peak of the infections (e.g., California). Sustained tight control on the population flow mitigated the second outbreak of the infections (e.g., New Jersey). A multistage reopening policy resulted and caused the susceptible population release to perturb infections even when R_b was reduced below its unrestricted value (e.g., Colorado) (*SI Appendix*, Fig. S11). Model results demonstrate links between the multiphase population release and Covid-19 pandemic. In particular, the formulation of the population release plays a primary role in the control of infections; appropriate control on the population release, corresponding to practical regulation and policy of travel restriction, can effectively mitigate the subsequent wave of the disease. Sensitivity tests reveal that other measures such as self-quarantine (i.e., reducing ϕ) and masks and social distancing (i.e., R_b) likely mitigate the infections but seems to play a lesser role in comparison to the regulation of the population flow/release.

Discussion

The spatial analysis of the daily Covid-19 cases here revealed that the spread of the disease follows a dynamic multifractal scaling across the continental United States spanning from 10 to 2,600 km, consistently trending toward that of the susceptible population. Multifractal scaling properties, first introduced in turbulence research, have been widely observed in geophysical fields that arise as a result of nonlinear processes acting over a wide range of scales. Examples are plentiful and include rainfall and cloud fields, temperature fields, roughness of the ocean

surface, and the intrinsic permeability of aquifers (33), among others. In addition, phenomena such as earthquakes and sea ice spread as well as human speech have been analyzed and shown to be multifractal (33). The spread of Covid-19 also obeys a well-defined multifractal distribution delineated by $\alpha = 1.78 \pm 0.10$ in the UM model. The α is invariant with respect to time and set by the multifractal distributed susceptible human population in the United States. This indicates the intrinsic statistics for the spatial spread of the disease, and the spatial distribution of the population is the same, suggesting population agglomeration could be a harbinger of the spatial intermittency of Covid-19 when near-stationary conditions are attained. Daily multifractal results indicate that for the Covid-19 cases, the codimension parameter $c1$ decreases over time from 0.42 on March 22 to 0.18 on October 14, 2020, which is higher than its population counterpart ($c1 = 0.15$). Higher codimension values for the Covid-19 cases, compared to that of the population, indicate stronger spatial intermittency of the disease occurring at relatively smaller scales.

The spatial epidemic modeling, relying on the SIR formalism, has been widely employed to represent the spread and control of infectious disease. Diffusion systems have been extensively used to describe movements of population and spatial structures in epidemic transmission (35, 36). The findings here suggest that employing the diffusion-modulated representation resulting in traveling waves (37, 38) to simulate rapid "invasion fronts" of Covid-19 infections might be problematic due to the fact that diffusion promotes space filling of the whole two-dimensional space. Such behavior likely removes over time the observed intermittency of Covid-19 cases at the finest scales (county to county).

The results here demonstrate that the kernel-modulated SIR model integrated within a multifractal population framework can reasonably reproduce the dynamics of the spatial correlation of the Covid-19 cases over at least two decades of spatial scales in comparison to the diffusion-based and immobile SIR models. A kernel-based approach that incorporates long-distance human mobility relevant for epidemic spatial spread into the SIR model incorporates key aspects of the transport network (near and far) between counties. In recent years, network models have been extensively developed to delineate interactions among individuals for better characterizing the epidemic spreading (39, 40). Overwhelming evidences have proved the emergence of complex and heterogeneous connectivity patterns over a wide range of scales during the epidemic spreading (15). Therefore, a network-based epidemic model (e.g., agent based) might encounter challenge for gathering a sufficient amount of data on social temporal networks and delivering a timely response or guidance to the pandemic. In addition, network theories likely focus on the long-range transport, but might remain "silent" on the long-range correlation of a susceptible population in space. In contrast, using the SIR with a kernel is a compromise between the agent-based and diffusion-based models. The former requires confidential data or data that might not be readily available, while the latter does not reproduce the observed spatial correlation structure. The kernel-modulated approach proposed here likely accommodates the merits of network-modulated models that support long-range transport while maintaining a local SIR dynamic at finer spatial scales. In addition, the multifractal population framework generated herein using the UM model captures the underlying statistics of the spread of the disease. It allows for a dynamic evolution of the spatial correlation of the pandemic to that of the population as observed from the daily Covid-19 data analysis.

The reproduction number is of major significance in epidemiology to assess the potential for epidemic spread in a susceptible population. Common reproduction numbers in use are R_0 , often assumed a fixed quantity, and R_e , which is pragmatically updated by fitting to the observed data of cases during the epidemic. We used a transient reproduction number R_b to reflect

exposure pathways, potency of the disease (infectivity) (41), and physical constraints to its transmission. Compared to the R_c , the value of R_b adopted herein represents the effectiveness of precautionary behavior at the individual scale. Individuals taking preventive measures such as social distancing and wearing masks will likely reduce the R_b value. The susceptible population flow $F_s(t)$ reflects collective isolation action (endogenous or exogenous through local or state regulations) to minimize exposure to the disease. For example, due to a relatively large population release by relaxing the stay-at-home order, the number of infected cases could still dramatically rise. Thus, R_c continues to increase despite a drastic reduction in R_b . Using this approach, we elucidated that preventive measures at the individual level (through wearing masks and social distancing) is not sufficient for reducing the overall spread of the disease and that government policy would need to be factored in terms of opening, which would increase the flux of susceptibles captured by F_s .

The simulations here reveal that the multiphase population release mechanism, reflecting differing regulation and policy of reopening from US states, plays a primary role in explaining the outbreak of the Covid-19 pandemic (February 15 to October 14 2020). In particular, managing release periods and the rate of population release to the susceptible category can effectively hinder Covid-19 infections. Overall, this study provides a perspective for interpreting the complex tempo-spatial evolution of Covid-19 and to predict it at numerous spatial scales based on nonpharmaceutical intervention decisions and without using data requiring tracking individuals.

Materials and Methods

Multifractals. A geophysical field, G , exhibits scaling if its power spectrum, E_G , has a form

$$E_G \sim k^{-\kappa}, \quad [1]$$

where the sign “ \sim ” implies proportionality, k is the wave number (or inverse scale) in Fourier space, and κ is the spectral “slope” (i.e., when plotted in log–log space). Scaling properties of the field G at the support scale λ can be characterized using the statistical moments of the field:

$$\langle G^q \rangle_\lambda \sim \lambda^{-K(q)}, \quad [2]$$

where $\langle G^q \rangle_\lambda$ represents the q^{th} order moment, and brackets imply its ensemble averaging, and the $K(q)$ represents the moment scaling function, reflecting scaling characteristics of the field. For a multifractal field, $K(q)$ behaves as a nonlinear convex (i.e., upward looking). We use herein the UM model (34), which allows for non-Gaussian statistics. In this model, $K(q)$ is universally described as

$$K(q) \sim \frac{c1}{\alpha-1}(q^\alpha - q), \quad [3]$$

where multifractality parameter α denotes the underlying statistics of $\ln(G)$ fields ($0 < \alpha \leq 2$); for $\alpha=2$, $\ln(G)$ is Gaussian (or G is log normal), and values of α less than 2.0 result in the α -stable probability distributions (42). A decrease in α causes the generated G field to become more “spotty” (or disordered) in space, where the large values clump together in the field. The parameter $c1$ denotes the mean field’s codimension and represents the spread from the mean, and it is equal to half of the variance when the $\ln(G)$ field is Gaussian.

In geophysical application, G , is commonly formulated as the following (34):

$$\langle G^q \rangle_\lambda = \lambda^{K(q)} \lambda^{-qH} = \lambda^{K(q)-qH}, \quad [4]$$

where the geophysical fields are derived by a fractional integration of the fields delineated in Eq. 2 at order of H . The parameter H , sometimes known as the Hurst coefficient, represents the redistribution of singularities with an average shift of $-H$. In such a field, the spectral slope κ is expressed as follows:

$$\kappa = 2H + [1 - K(2)]. \quad [5]$$

The values of H have been found to be $1/3$ for homogeneous isotropic turbulence flow fields (43, 44) and becomes an empirical coefficient for other geophysical fields such as permeability (45) and rain (46). A field with $H =$

0 (i.e., Eq. 2) is labeled a “conservative” field, attributed to an invariant mean with respect to spatial scale (47).

The spatial analysis was performed for daily Covid-19 cases reported in the US counties during March 22 to October 14, 2020. Our analysis focused on the eastern part of the United States (east of 100° W) at a 10-km scale. The moments and power spectra of daily Covid-19 cases as well as the total population in the corresponding counties were calculated to explore the evolution of their spatial correlations with time (Fig. 1). We acquired daily Covid-19 cases from Johns Hopkins Coronavirus Resources Center (<https://coronavirus.jhu.edu>), last interrogated on October 14, 2020. The corresponding county population was obtained from the US Census Bureau (<https://www.census.gov/data.html>). Multifractal fields using “continuous cascades” (33, 34) were generated based on the estimated UM parameters for the population and applied for epidemic SIR modeling (48, 49). A smaller data set (March 22 to May 7, 2020) was analyzed in a previous work (50).

SIR Model. The SIR model adopted herein assumes that the individual population in the county compartment can be categorized into three classes: susceptible (S), infectious (I), and recovered or removed (R). The dynamical SIR system subjected to local (i.e., county scale) mass action law is expressed as follows (16):

$$\frac{dS}{dt} = -\beta \frac{I}{S_0} S + F_s, \quad [6a]$$

$$\frac{dI}{dt} = + \left[\left(\beta \frac{1}{S_0} \right) S - \gamma_R \right] \left[(1 - \phi) I + \phi \iint p(x - x', y - y') I(x', y') dx dy \right], \quad [6b]$$

, and

$$\frac{dR}{dt} = +\gamma_R I, \quad [6c]$$

where F_s represents the influx rate of susceptible (i.e., population flow) into the pool from within the county-level population of size N . Thus, $S + I + R \leq N$ at the county scale. The term S_0 denotes the total number of infected because of the first wave of the Covid-19 infections. Note that instead of the first wave of the Covid-19 infections, the total susceptible population can be introduced in the denominator as common in mass action mixing models. However, in that case, as the susceptible population flux (i.e., F_s) is introduced into the model, the denominator must be updated at each time step according to the new population. This reduces the model stability and lends the model to be less practical. Therefore, the total population infected within the first epidemic wave (i.e., a constant parameter defined as S_0) was initially estimated based on the reported number at the end of the first phase. Then, S_0 was calibrated along with other key parameters by fitting the model to reported daily confirmed cases. The terms β and γ_R are known as the rates of infection and recovery, respectively. The ratio β/γ_R represents the average number of secondary infections caused by a primary case, which is expressed as R_b , defined as the reproduction number based on the behavior of an individual. By contrast, the effective reproduction number, R_e , is obtained by fitting to the observed number of cases, and it ranged from 2.0 to 7.0 for Covid-19 (16)—but a global convergence to 4.5 has been recently reported when $\gamma_R = 1/14 \text{ d}^{-1}$. Note that the behavior-based reproduction number R_b defined herein is a time-varying entity, reflecting exposure pathways, potency of the disease, and physical constraints to its transmission at the individual scale through β . The term β includes the multiplicative effects of contact probability between individuals and disease transmission, leading to infections probability conditioned upon contact. Therefore, individual’s precaution behaviors such as social distancing and wearing masks will likely reduce the value of R_b . The values of R_b and F_s are related, and knowing one of them would provide the other.

The parameter ϕ represents the percentage of infections migrating across counties per unit time. The integral term is a convolution, and $p(x, y)$ is the kernel function, assumed herein to be Gaussian in space, and therefore has two parameters, L and σ . In the steady-state limit with a pointwise initial infection and for a spatially uniform susceptible population, the Gaussian kernel recovers the diffusion predicted steady-state front speed, which is why this kernel was chosen here. The parameter L denotes the length of the kernel function, indicating the maximum counties that an individual infected case can travel. The parameter σ denotes the variance of the distribution within the kernel. The convolution of a multifractal field is illustrated in Fig. 2B. The kernel function was only performed to the infectious class for simplicity. However, additional simulations were conducted by including the same kernel term for susceptible and recovered classes, and the simulated results obtained from the two models were nearly identical (*SI Appendix*,

Figs. S13 and S14). This indicates that the dynamics of the infections are not restricted by availability of susceptible cases, which is due to the larger susceptible population compared to the infectious population. An immobile SIR model, assuming isolated counties with no migration, was also run for comparison (label 3, Fig. 2A). Note that the original SIR model is recovered when setting F_s and ϕ to zero. We are assuming herein that the travel into and out of the compartment (i.e., the county scale) is small when compared to the total population in the county, and thus, the N is invariant with respect to time. It is to be noted here that when $\phi = 0$, the equilibrium state is disease free ($I = 0$), along with $R = 0$ and $S = N$, necessitating an $F_s = 0$ at equilibrium consistent with the original SIR.

The SIR simulations were run for 250 d, starting on Feb 15, 2020, to predict the temporal evolution of infected cases based on the local population (i.e., within each county-scale simulation cell). The initial value of R_b was taken as R_{c0} , estimated to be 4.5, taking into account Covid-19 data for all US states, and agrees with other estimates derived from 57 countries with initial minimal intervention (16). Herein, we assumed a relatively large $R_b \sim 5.0$ at the beginning of the pandemic and linearly decreased R_b from 5.0 to 3.0 between mid-March and mid-April, taking into account stay-at-home orders and social distancing; then, we decreased R_b from 3.0 to 2.0 between later May and later June to take into account further precautions of the disease after the conditional reopening (e.g., wearing masks are mandatory in stores in many states).

The parameter γ_R is often interpreted as the inverse of the mean recovery period. With regards to Covid-19, the time period for mild illness from the onset of symptoms to natural recovery is, on average, 2 wk; therefore, the value γ_R is estimated to be $1/14 \text{ d}^{-1}$ (16). Note that the product of R_b and γ_R represents the transmission rate of the infections; therefore, the R_b reduction also accommodates the γ_R increase, which reflects, for example, improved hospital capacity. The initial spatial distribution of the population (i.e., N) across counties was assumed to be multifractal, consistent with the observations. The initial infected cases I_0 were randomly seeded in the domain by a uniformly distributed random number between 0 and 1, accounting for $\sim 0.4\%$ of the susceptible cases. Notice that the I_0 selected herein causes the model to have a better match to the observed data. A different percentage of I_0 was tested in the model; it turns out that the different percentage of I_0 only translates the SIR curves in time, but the overall temporal shape of the curves is not appreciably altered.

Traditionally, the contact and dispersal of a contagious disease has been modeled using diffusion equations in which the spreading of the disease relies on the density gradient of infections between two adjacent compartments (label 1, Fig. 2A), expressed as follows:

$$\frac{dI}{dt} = + \left[\left(\beta \frac{1}{N} \right) S - \gamma_R \right] I + \nabla(D \nabla I), \quad [7]$$

where the term $\nabla(D \nabla I)$ denotes density-driven migration of infections, and D denotes the diffusion coefficient of the disease. Diffusion models may underestimate the extent of epidemic spreading, as they represent "spillage" to neighboring communities but do not account for long-distance travel across more than two counties and unavoidable travel to severe epidemic compartments (i.e., regions). One solution to this problem is to simulate the

advection of people (i.e., travel) using agent-based models, discussed earlier. An alternate approach is to use the kernel approach (Eq. 6).

Multiphase Population Release Mechanism. With the Covid-19 pandemic in near constant fluctuation in the United States, the US states adapted rapidly with rules and regulations. Between March 21 to April 6, all the US states issued stay-at-home orders to mitigate the spread of Covid-19. Such travel restriction has been identified to dramatically decrease the number of travelers nationally and internationally; reports from the Transportation Security Administration (TSA) indicate that at the end of March after the Covid-19 outbreak, the numbers of air passengers dropped $\sim 96\%$. On May 20, all states began to partially lift restrictions, and on June 29, 48 US locations and eight host nations met the conditions to lift travel restrictions; such conditional reopening policies subsequently resulted in a gradual increase in TSA traveler numbers (Fig. 2C). However, due to a second wave of Covid-19 infections in the United States, many states introduced a "mitigation" plan for areas with increasing positive test rates. For example, in Oregon, in late July, because of an increase in Covid-19 cases, Umatilla County moved back to stay-at-home status, while Morrow and Malheur Counties moved from Phase 2 reopening back to Phase 1 reopening.

To take into account such policy-modulated population flow, we incorporated multiphase population release into the SIR model by allowing the number of susceptibles to increase over time rather than assuming that the whole population is susceptible at the initial outbreak. Thus, the initial susceptible population was labeled S_{in} , which is estimated based on the occurrence of the first epidemic peak (SI Appendix, Fig. S15). The calibrated values for each state are reported in SI Appendix, Table S1. Subsequently, the number of susceptibles was increased at the rate $F_s(t)$, and therefore, the susceptible population flow was $S_{in,t} = \int_0^t F_s(t) dt$.

Numerical trials were performed to calibrate the pattern of the F_s function along with the parameter S_{in} , S_0 , and ϕ by fitting the model to daily confirmed cases for each US state. Note that developing a systematic optimization algorithm seems unlikely. This is because the fitting is based on the duration and shape of epidemic curves, which diverged across US states according to their orders and rules. Therefore, calibration has to be performed state by state to reflect these differences in orders and rules. The simulations started with an initial approximation of S_{in} , S_0 , and ϕ along with a multifractal population field, and then numerical trials were performed to calibrate values of these parameters and temporal patterns of piecewise linear function F_s until a match between the model and observation was achieved. The scheme of calibration is illustrated in SI Appendix, Fig. S15. The calibrated F_s were reported in Fig. 4 and SI Appendix, Fig. S10, and the calibrated values of S_{in} , S_0 , and ϕ were reported in SI Appendix, Table S1.

Data Availability. All study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. This work was funded by a Rapid Response Research grant from the US NSF (Chemical, Bioengineering, Environmental and Transport Systems [CBET] 2028271). However, it does not necessarily reflect the views of the funding agency, and no official endorsement should be inferred.

1. J. H. University, *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE)*, (2021).
2. P. Sharkey, The US has a collective action problem that's larger than the coronavirus crisis: Data show one of the strongest predictors of social distancing behavior is attitudes toward climate change. *Vox* (2020). <https://www.vox.com/2020/4/10/21216216/coronavirus-social-distancing-texas-unacast-climate-change>. Accessed 14 October 2020.
3. F. Sun, S. A. Matthews, T.-C. Yang, M.-H. Hu, A spatial analysis of the COVID-19 period prevalence in U.S. Counties through June 28, 2020: Where geography matters? *Ann. Epidemiol.* **52**, 54–59.e1 (2020).
4. J. Cordes, M. C. Castro, Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spat. Spatio-Temporal Epidemiol.* **34**, 100355 (2020).
5. A. Mollalo, B. Vahedi, K. M. Rivera, GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Sci. Total Environ.* **728**, 138884 (2020).
6. S. Gao, J. Rao, Y. Kang, Y. Liang, J. Kruse, Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSPATIAL Spec.* **12**, 16–26 (2020).
7. C. H. Zhang, G. G. Schwartz, Spatial disparities in coronavirus incidence and mortality in the United States: An ecological analysis as of May 2020. *J. Rural Health* **36**, 433–445 (2020).
8. K. Dietz, J. A. Heesterbeek, Daniel Bernoulli's epidemiological model revisited. *Math. Biosci.* **180**, 1–21 (2020).
9. H. W. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
10. W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **138**, 55–83 (1932).
11. W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics. III.—Further studies of the problem of endemicity. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **141**, 94–122 (1933).
12. M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2011).
13. J. Satsuma, R. Willox, A. Ramani, B. Grammaticos, A. Carstea, Extending the SIR epidemic model. *Physica A* **336**, 369–375 (2004).
14. V. Colizza, M. Barthélemy, A. Barrat, A. Vespignani, Epidemic modeling in complex realities. *C. R. Biol.* **330**, 364–374 (2007).
15. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
16. G. G. Katul, A. Mrad, S. Bonetti, G. Manoli, A. J. Parolari, Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics. *PLoS One* **15**, e0239800 (2020).
17. D. Gunzler, A. R. Sehgal, Time-varying COVID-19 reproduction number in the United States. *medRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.04.10.20060863>.
18. C. You et al., Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *Int. J. Hyg. Environ. Health* **228**, 113555 (2020).
19. Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021 (2020).

20. Z. Du, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, COVID-19 serial interval estimates based on confirmed cases in public reports from 86 Chinese cities. *medRxiv* [Preprint] (2020). [10.1101/2020.04.23.20075796](https://doi.org/10.1101/2020.04.23.20075796) (14 October 2020).
21. D. D. R. Bandyopadhyay, B. C. Weimer, Pandemic dynamics of COVID-19 using epidemic stage, instantaneous reproductive number and pathogen genome identity (GENI) score: Modeling molecular epidemiology. *medRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.03.17.20037481> (14 October 2020).
22. J. Gamero, J. A. Tamayo, J. A. Martinez-Roman, Forecast of the evolution of the contagious disease caused by novel coronavirus (2019-nCoV) in China. *arXiv* [Preprint] (2020). <https://arxiv.org/abs/2002.04739> (14 October 2020).
23. A. L. Ziff, R. M. Ziff, Fractal kinetics of COVID-19 pandemic. *medRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.02.16.20023820> (14 October 2020).
24. M. Chinazzi et al., The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
25. R. Li et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).
26. Institute for Health Metrics and Evaluation (IHME), Covid-19 Projections (2020). <https://covid19.healthdata.org/united-states-of-america>. Accessed 14 October 2020.
27. M. Ienca, E. Vayena, On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat. Med.* **26**, 463–464 (2020).
28. S. Park, G. J. Choi, H. Ko, Information technology-based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *JAMA* **323**, 2129–2130 (2020).
29. D. G. Kendall, “Mathematical models of the spread of infection” in *Mathematics and Computer Science in Biology and Medicine*, Medical Research Council, Ed. (His/Her Majesty’s Stationery Office (H.M.S.O.), London, 1965), pp. 213–225.
30. J. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications* (Springer-Verlag, 2001), vol. 3.
31. S. Appleby, Multifractal characterization of the distribution pattern of the human population. *Geogr. Anal.* **28**, 147–160 (1996).
32. T. Mori, T. E. Smith, W.-T. Hsu, Common power laws for cities and spatial fractal structures. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 6469–6475 (2020).
33. S. Pecknold, S. Lovejoy, D. Schertzer, C. Hooge, J. Malouin, “The simulation of universal multifractals” in *Cellular Automata*, J. M. Perdang, A. Lejeune, Eds. (World Scientific, 1993), pp. 228–267.
34. D. Schertzer, S. Lovejoy, Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.* **92**, 9693–9714 (1987).
35. S. Anija, V. Capasso, On the stabilization of reaction–diffusion systems modeling a class of man-environment epidemics: A review. *Math. Methods Appl. Sci.* **33**, 1235–1244 (2010).
36. G. Hariharan, K. Kannan, Review of wavelet methods for the solution of reaction–diffusion problems in science and engineering. *Appl. Math. Model.* **38**, 799–813 (2014).
37. M. S. Bartlett, “Deterministic and stochastic models for recurrent epidemics” in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed. (University of California Press, Berkeley, 1956), p. 109.
38. I. Cruickshank, W. S. Gurney, A. R. Veitch, The characteristics of epidemics and invasions with thresholds. *Theor. Popul. Biol.* **56**, 279–292 (1999).
39. M. E. Newman, Spread of epidemic disease on networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 016128 (2002).
40. L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Kowronski, R. C. Brunham, Network theory and SARS: Predicting outbreak diversity. *J. Theor. Biol.* **232**, 71–81 (2005).
41. R. M. Anderson, B. Anderson, R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1992).
42. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw Hill, New York, ed. 3, 1991), p. 666.
43. G. Katul, B. Vidakovic, J. Albertson, Estimating global and local scaling exponents in turbulent flows using discrete wavelet transformations. *Phys. Fluids* **13**, 241–250 (2001).
44. C. Meneveau, K. R. Sreenivasan, The multifractal spectrum of the dissipation field in turbulent flows. *Nucl. Phys. B Proc. Suppl.* **2**, 49–76 (1987).
45. L. Tennekoon, M. C. Boufadel, D. Lavalley, J. Weaver, Multifractal anisotropic scaling of the hydraulic conductivity. *Water Resour. Res.* **39**, 1193 (2003).
46. Y. Tessier, S. Lovejoy, D. Schertzer, Universal multifractals: Theory and observations for rain and clouds. *J. Appl. Meteorol.* **32**, 223–250 (1993).
47. S. Lovejoy, D. Schertzer, Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. *J. Geophys. Res. D Atmospheres* **95**, 2021–2034 (1990).
48. O. N. Bjørnstad, B. F. Finkenstädt, B. T. Grenfell, Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* **72**, 169–184 (2002).
49. S. Funk, M. Salathé, V. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: A review. *J. R. Soc. Interface* **7**, 1247–1256 (2010).
50. X. Geng et al., Population agglomeration is a harbinger of the spatial complexity of COVID-19. *Chem. Eng. J.*, [10.1016/j.cej.2020.127702](https://doi.org/10.1016/j.cej.2020.127702) (2020).