communications physics

Article

https://doi.org/10.1038/s42005-024-01578-w

Validity of Markovian modeling for transient memory-dependent epidemic dynamics

Check for updates

Mi Feng [1,2], Liang Tian [1,3,4,5] ✉, Ying-Cheng Lai [6,7] & Changsong Zhou [1,2,3] ✉

The initial transient phase of an emerging epidemic is of critical importance for data-driven model building, model-based prediction of the epidemic trend, and articulation of control/prevention strategies. Quantitative models for real-world epidemics need to be memory-dependent or non-Markovian, but this presents difficulties for data collection, parameter estimation, computation, and analyses. In contrast, such difficulties do not arise in the traditional Markovian models. To uncover the conditions under which Markovian and non-Markovian models are equivalent, we develop a comprehensive computational and analytic framework. We show that the transient-state equivalence holds when the average generation time matches the average removal time, resulting in minimal Markovian estimation errors in the basic reproduction number, epidemic forecasting, and evaluation of control strategy. The errors depend primarily on the generation-to-removal time ratio, while rarely on the specific values and distributions of these times. Overall, our study provides a general criterion for modeling memory-dependent processes using Markovian frameworks.

When an epidemic emerges, the initial transient phase of the disease spreading dynamics before a steady state is reached is of paramount importance, for two reasons[1–11]. First, estimating key indicators or parameters, e.g., the generation time, serial intervals, and basic reproduction number, is crucial for predicting and formulating control strategies when the underlying dynamical process has not reached a steady state. Second, it is during the transient phase that control and mitigation strategies can be effectively applied to prevent a large-scale outbreak. Prediction and control depend, of course, on a quantitative model of the epidemic process, which can be constructed based on the key parameters estimated from data collected during the transient phase. In principle, since the dynamical processes underlying real-world epidemics are generally memory-dependent in the sense that the state evolution depends on the history, a rigorous modeling framework needs to be of the non-Markovian type, but this presents great challenges in terms of data collection, parameter estimation, computation and analyses[12–14]. The difficulties can be alleviated by adopting the traditional simplified memoryless Markovian framework, which, however, may potentially result in deviations[15–26]. An outstanding question is, are there specific conditions under which a Markovian epidemic outbreak could be equivalent to the non-Markovian counterpart during the transient phase, enabling an accurate description of memory-dependent transmission within the Markovian framework? Additionally, another important issue remains, how are the errors of Markovian estimation determined? The purpose of this paper is to provide a comprehensive answer to these questions.

The COVID-19 pandemic has highlighted the need and importance of understanding disease spreading and transmission to accurately predict, control, and manage future outbreaks through non-pharmacological interventions and vaccine allocation strategies[1–10]. To accomplish these goals, accurate mathematical modeling of the disease-spreading dynamics is key. In a general population, epidemic transmission occurs via some kind of point process, where individuals become infected at different points in time. It has been known that point processes in the real world are typically non-Markovian with a memory effect in which the distribution of the interevent times is not exponential[2,23,27–41]. For example, the interevent time distribution arising from the virus transmission with COVID-19 is not of the

[1]Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. [2]Centre for Nonlinear Studies and Beijing-Hong Kong-Singapore Joint Centre for Nonlinear and Complex Systems (Hong Kong), Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. [3]Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. [4]Institute of Systems Medicine and Health Sciences, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. [5]State Key Laboratory of Environmental and Biological Analysis, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China. [6]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA. [7]Department of Physics, Arizona State University, Tempe, AZ 85287, USA. ✉e-mail: liangtian@hkbu.edu.hk; cszhou@hkbu.edu.hk

memoryless exponential type but typically exhibits memory-dependent features characterized by the Weibull distribution[2]. Strictly speaking, from a modeling perspective, disease spreading should be described by a non-Markovian process. A non-Markovian approach takes into account historical memory of disease progression, mathematically resulting in a complex set of integro-differential equations in the form of convolution. (Note that besides the transmission process, where the infection capacity is highly dependent on the elapsed time from the infection, there are other types of memory-dependent factors. For instance, in certain meta-population network models, the infection capacity of a node is influenced by the transmission path leading to that node, rather than solely relying on the time of infection[42]. Thus, it is important to mention that we only take the elapsed time into consideration in this work.)

There are difficulties with non-Markovian modeling of memory-dependent disease spreading. The foremost is data availability. In particular, while standard epidemic spreading models are available, the model parameters need to be estimated through data. A non-Markovian model often requires detailed and granular data that can be difficult to get, especially during the early stage of the epidemic where accurate modeling is most needed[12]. From a theoretical point of view, it is desired to obtain certain closed-form solutions for key quantities such as the onset and size of the epidemic outbreak, but this is generally impossible for non-Markovian models[15,43]. Computationally, accommodating memory effects in principle makes the underlying dynamical system infinitely dimensional, practically requiring solving an unusually large number of dynamical variables through a large number of complex integro-differential equations[13,14]. In contrast, in an idealized Markovian point process, events occur at a fixed rate, leading to an exponential distribution for the interevent time intervals and consequently a memoryless process. If the spreading dynamics were of the Markovian type, the aforementioned difficulties associated with non-Markovian dynamics no longer exist. In particular, a Markovian spreading process can be described by a small number of ordinary differential equations with a few parameters that can be estimated even from sparse data, and the numerical simulations can be carried out in a computationally extremely efficient manner[15–26]. For these reasons, many recent studies of the COVID-19 pandemic assumed Markovian behaviors to avoid the difficulties associated with non-Markovian modeling[3–10]. The issue is whether such a simplified approach can be justified. Addressing this issue requires a comprehensive understanding of the extent to which the Markovian approach represents a good approximation to model non-Markovian type of memory-dependent spreading dynamics, and specifically of the conditions under which the Markovian theory can produce accurate results that match those from the non-Markovian model. This highlights the importance of studying the equivalence between non-Markovian and Markovian dynamics.

There were previous studies of the so-called steady-state equivalence between Markovian and non-Markovian modeling for epidemic spreading. In particular, when the system has reached a steady state, such an equivalence can be established through a modified definition of the effective infection rate[18,21,22]. From a realistic point of view, the equivalence limited only to steady states may not be critical as the transient phase of the spreading process before any steady state is reached is more relevant and important. For example, when an epidemic occurs, it is of fundamental interest to estimate the key indicators such as the generation time (the time interval between the infections of the infector and infectee in a transmission chain), the serial interval (the time from illness onset in the primary case to illness onset in the secondary case), and the basic reproduction number (the average number of secondary transmissions from one infected person), but they are often needed to be estimated when the dynamics have not reached a steady state[2–11]. It is the equivalence in the transient dynamics rather than the steady state that determines whether the transmission features in the early stages of a memory-dependent disease outbreak can be properly measured through Markovian modeling. Moreover, it is only during the transient phase that control and mitigation strategies can be effective in preventing a large-scale outbreak. Discovering when and how a non-Markovian process can be approximated by a Markovian process during the transient state is thus of paramount importance. To our knowledge, such a "transient-state equivalence", where the Markovian and non-Markovian transmission models produce similar behaviors over the entire transient transmission period, has not been established. In fact, the conditions under which the transient equivalence may hold are completely unknown at present.

In this paper, we present results from a comprehensive study of how memory effects impact the Markovian estimations in terms of the errors that arise from the Markovian hypothesis. We consider both the steady-state and transient-state equivalences between non-Markovian and Markovian models. We first rigorously show that, in the steady state, a memory-dependent non-Markovian spreading process is always equivalent to certain Markovian (memoryless) ones. We then turn to the transient states and find that an approximate equivalence can still be achieved but only if the average generation time matches the average removal time in the memory-dependent non-Markovian spreading dynamics. Qualitatively, the equality of the two times gives rise to a memoryless correlation between the infection and removal processes, thereby minimizing the impact of any memory effects. We establish that equality gives the condition under which Markovian theory accurately describes memory-dependent transmission.

One fundamental quantity underlying an epidemic process is the basic reproduction number R_0. Our theoretical analysis indicates that, when the average generation and removal times are equal, the transient-state equivalence between memory-dependent and memoryless transmissions will minimize the error of the Markovian approach in estimating R_0 and lead to its accurate epidemic forecasting and prevention evaluation. Another finding is that the generation-to-removal time ratio plays a decisive role in the accuracy of the Markovian approximation. Specifically, if the average generation time is smaller (greater) than the average removal time, the Markovian approximate will lead to an overestimation (underestimation) of R_0 and epidemic forecasting as well as the errors of the prevention evaluation, which can also be verified based on readily accessible clinical data of four types of real diseases.

The estimation accuracy is largely determined by the time ratio but rarely depends on the particular forms of time distributions or the specific values of the average generation and removal times. This property is of great practical significance because it is in general challenging to obtain the detailed distributions of the generation and removal times in the early stages of the epidemic[44], but their average values can be reliably estimated even during the transient phase[45–49]. Moreover, based on this property, we have developed a semi-empirical mathematical relationship that connects the errors in estimating R_0 with the generation-to-removal time ratio. This relationship holds practical value as it can be utilized to rectify errors in real-world scenarios. The rectification of R_0 and epidemic forecasting can be accomplished through our web-based application[50].

Overall, our study establishes a general criterion for modeling memory-dependent processes within the context of Markovian frameworks. Once the condition for the existence of a transient-state equivalence between Markovian and non-Markovian dynamics is fulfilled, epidemic forecasting and prevention evaluation can be carried out using the Markovian model, again based solely on the data collected from the transient phase.

Results

The overall structure of this work is depicted in Fig. 1. The section titled "Model" presents the Model building of the age-stratified Susceptible-Infected-Removed (SIR) spreading (Fig. 1a), highlighting the difference between the Markovian (memoryless) and non-Markovian (memory-dependent) dynamics (Fig. 1b). In the "Dynamical equivalence" section, we demonstrate the equivalence between Markovian (memoryless) and non-Markovian (memory-dependent) dynamics for steady state and transient dynamics, which will further lead to the accurate description of memory-dependent dynamics by the Markovian theory (Fig. 1c). The section "Markovian approximation of memory-dependent spreading dynamics" analyzes the errors of the Markovian approach in estimating R_0, epidemic forecasting, and prevention evaluation (Fig. 1d). (Note that this study

involves numerous parameters and variables, all of which have been comprehensively detailed and listed for reference in Supplementary Table S1 of Supplementary Note 1.)

Model

We articulate an age-stratified SIR spreading dynamics model, in which the entire population is partitioned into various age groups with intricate age-specific contact rates among them. The distribution of population across different age groups is represented by an age distribution vector (**p**), with the age-dependent contact matrix (**A**) quantifying the transmission rates between different age groups. Both **p** and **A** can be constructed from empirical data[51,52]. For convenience, to distinguish between the actual dynamical process and its theoretical treatment, throughout this paper we use the terms "memory-dependent" and "memoryless" to describe actual spreading processes and Monte Carlo simulations, while in various theoretical analyses, the corresponding terms are "non-Markovian" and "Markovian."

The mechanism of disease transmission across different age groups and the recovery or death of infected individuals can be described by the SIR compartmental model, as illustrated in Fig. 1a, where the individuals possess three types of states: susceptible (S), infected (I), and removed (R). Susceptible individuals (S) have not contracted the disease and are at risk of being infected. Infected individuals (I) have contracted the disease and can infect others. Removed individuals (R) who have recovered or died from the disease. There are two dynamical processes: (1) infection during which susceptible individuals become infected by others and transition to the I state so as to become capable of infecting others, as shown in Fig. 1a(i–iii), and (2) removal during which infected individuals recover or die from the disease transmission and transition to the R state, as shown in Fig. 1a(iv). The ability to infect others of an infected individual can be characterized by the infection time distribution, $\psi_{\rm inf}(\tau)$ where τ denotes the time elapsed between the time the individual is infected and the current time, and the probability of the infection process occurring during the time interval $[\tau,\tau+d\tau)$ is given by $\psi_{\rm inf}(\tau)d\tau$, as shown in Fig. 1b(i). Likewise, the removal process is described by the removal time distribution, $\psi_{\rm rem}(\tau)$, where the probability of a removal occurring within the time interval $[\tau,\tau+d\tau)$ is given by $\psi_{\rm rem}(\tau)d\tau$, as shown in Fig. 1b(ii). The time distributions of the infection and removal processes with memory effects are general, with the exponential distributions associated with the memoryless process being a special case of the memory-dependent process.

The generic memory-dependent SIR spreading dynamics can be described by a set of deterministic integro-differential equations:

$$\frac{ds_l(t)}{dt} = -s_l(t)k\sum_{m=1}^{n} A_{lm}p_m \int_0^t \omega_{\rm inf}(t-t')\Psi_{\rm rem}(t-t')dc_m(t'), \quad (1)$$

$$i_l(t) = \int_0^t \Psi_{\rm rem}(t-t')dc_l(t'), \quad (2)$$

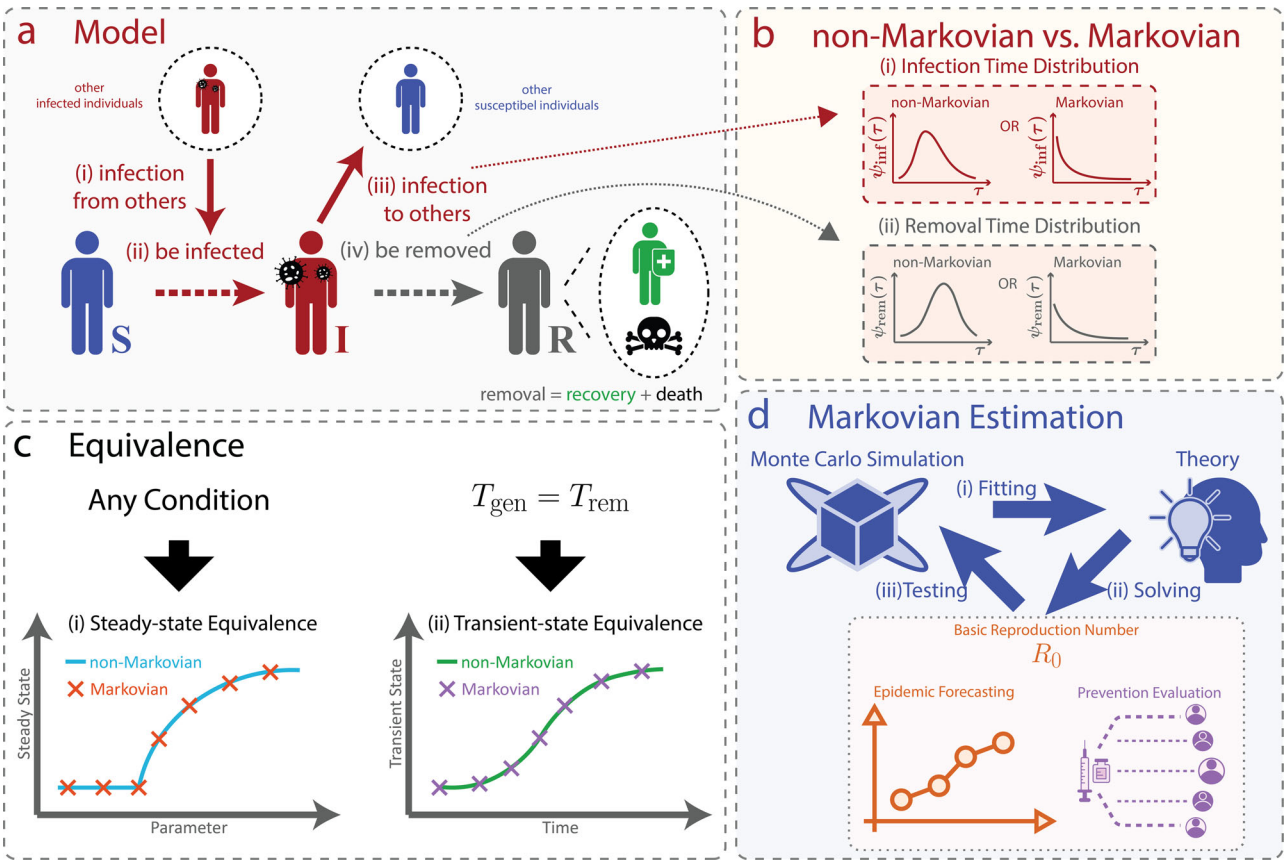

Fig. 1 | Overall structure of this work. a SIR Model. Each individual belongs to one of the three states: susceptible (S), infected (I), or removed (R). When infected (i), a susceptible individual will switch into the I state (ii) and gain the ability to infect others (iii). An infected individual is removed (through recovery or death) with a probability (iv). **b** Non-Markovian versus Markovian process. The infection capacity of an infected individual is characterized by the infection time distribution $\psi_{\rm inf}(\tau)$ and its removal can be described by the removal time distribution $\psi_{\rm rem}(\tau)$. For the non-Markovian process, the distributions can assume quite general forms, while the distributions are exponential for a Markovian process. **c** Equivalence between non-Markovian and Markovian processes: (i) steady-state equivalence holds under all conditions; (ii) transient-state equivalence only holds when $T_{\rm gen}$ is equal to $T_{\rm rem}$. **d** Markovian estimation of memory-dependent process. (i) The initial phase of the Monte Carlo simulation is used to fit the parameters according to the Markovian theory. (ii) Important issues such as the estimation of R_0, epidemic forecasting, and the evaluation of vaccination strategies can be addressed by the theory. (iii) The remaining data generated by the Monte Carlo simulation is used to test the accuracy of the estimated R_0, epidemic forecasting, and prevention evaluation.

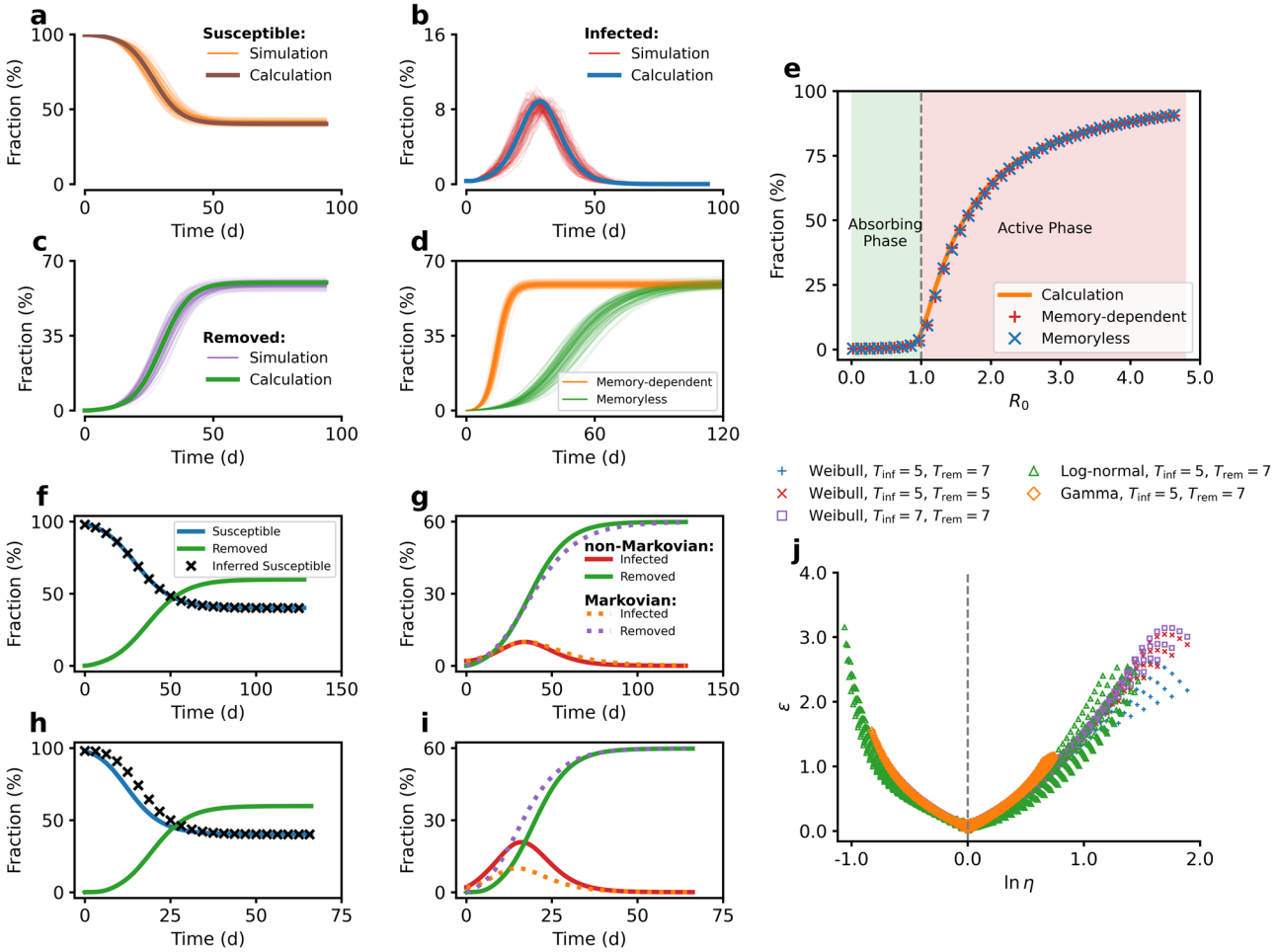

Fig. 2 | Steady-state and transient-state equivalence between Markovian and non-Markovian dynamics. a–c The solid brown, blue, and green curves represent the theoretical results of the susceptible, infected, and removed fractions, while the solid orange, red, and purple curves show the corresponding results of 100 independent Monte Carlo simulations ($\alpha_{inf}=3$, $\beta_{inf}=4$, $\alpha_{rem}=2.5$, $\beta_{rem}=3.6$, where α and β respectively represent the shape and scale parameters of time distributions of Weibull types, and subscripts "inf" and "rem" indicate the time distributions of infection and removal, respectively; $R_0=1.86$). **d** The orange and green curves, respectively, depict the removed fractions from the memory-dependent and memoryless Monte Carlo simulations of 100 independent realizations with steady-state equivalence ($\alpha_{inf}=1.5$, $\beta_{inf}=2.63$, $\alpha_{rem}=2$, $\beta_{rem}=2.4$ for memory-dependent simulations; $\gamma=0.11$ and $\mu=0.14$ for memoryless simulations; identical basic reproduction number as $R_0=1.86$ for the two types simulations). **e** Red + and blue × markers, respectively, represent the steady-state removed fractions of memory-dependent and memoryless Monte Carlo simulations for different values of R_0, where each marker is the result of averaging 100 independent simulations. The orange curve is the numerical calculations from Eq. (7), and the vertical dashed line denotes the critical point $R_0=1$. **f–g** For $T_{gen}=T_{rem}$ in the non-Markovian theory ($\alpha_{inf}=1.57$, $\beta_{inf}=5.57$, $\alpha_{rem}=1.57$, $\beta_{rem}=7.79$), the blue and green curves in (**f**) denote the susceptible and removed fractions, while the black × markers represent the inferred susceptible fractions calculated by substituting removed fractions in Eq. (11), which agrees with the susceptible curve calculated from Eq. (1). The red and green curves in (**g**) denote the non-Markovian infected and removed fractions, while the orange and purple dashed curves are the corresponding curves of the Markovian transmission ($\gamma=0.20$, $\mu=0.12$) obtained from Eqs. (13, 14), which agree with the non-Markovian results. They all have the same basic reproduction number, i.e., $R_0=1.87$. (The Euler-Lotka equation assumes exponential growth of a disease outbreak during the initial stage. As a result, the Markovian curves in (**g**) slightly deviate from the non-Markovian ones as the cumulative infections increase.) **h–i** For $T_{gen}\neq T_{rem}$ in the non-Markovian theory ($\alpha_{inf}=0.74$, $\beta_{inf}=4.16$, $\alpha_{rem}=3.32$, $\beta_{rem}=7.80$), the inferred susceptible curve in (**h**) does not match the numerical result, and the infected and removed curves of the Markovian transmission in (**i**) ($\gamma=0.41$, $\mu=0.28$) obtained from Eqs. (13, 14) do not match the corresponding non-Markovian results. They all have the same basic reproduction number, i.e., $R_0=1.87$. **j** Five scenarios for the non-Markovian time-distribution setting (within each scenario, T_{inf} and T_{rem} are fixed): Weibull, $T_{inf}=5$, $T_{rem}=7$ (blue+); Weibull, $T_{inf}=5$, $T_{rem}=5$ (red ×); Weibull, $T_{inf}=7$, $T_{rem}=7$ (purple □); log-normal, $T_{inf}=5$, $T_{rem}=7$ (green Δ); gamma, $T_{inf}=5$, $T_{rem}=7$ (orange ◇). The value of T_{gen} is modified to adjust $\ln\eta$ for better visualization.

$$r_l(t)=\acute{r}_l+\int_0^t[1-\Psi_{rem}(t-t')]dc_l(t'), \tag{3}$$

where $s_l(t)$, $i_l(t)$, and $r_l(t)$, respectively, denote the fractions of the susceptible, infected, and removed individuals in age group l. $\acute{r}_l=r_l(0)$ represents the initial fraction of the removed individuals in age group l. The term $c_l(t)=1-s_l(t)=i_l(t)+r_l(t)$ represents the fraction of cumulative infections (including both infections and removals; note that the physical meaning of $dc_l(t)$ is the new infected fraction in age group l at time t, which can further spread the infection to others, so the value of $dc_l(0)$ is considered as the fraction of infection seeds, where the initial removed fraction is not taken into account) in age group l, while p_m denotes the m-th element of the vector **p**, A_{lm} denotes the lm-th element of matrix **A**, k is a parameter to adjust the overall contacts and n is the total number of age groups. The quantity $\omega_{inf}(\tau)$ represents the hazard function of $\psi_{inf}(\tau)$, meaning the rate at which infection happens at τ, given that the infection has not occurred before τ. $\Psi_{rem}(\tau)$ is the survival function of $\psi_{rem}(\tau)$, meaning the probability that the removal has not occurred by τ (see Method for detailed calculations for hazard and

survival functions). When the infection and removal time distributions are known, Eqs. ((1)–(3)) provide an accurate description of generic (including memory-dependent and memoryless) SIR spreading Monte Carlo simulations in an age-stratified population-based system. As shown in Fig. 2a–c, the theory is validated by the agreement between the numerical solutions of Eqs. ((1)–(3)) and the results from direct Monte Carlo simulations (see Method for a detailed description of the Monte Carlo simulation procedure).

Equations (1–3) provide a general framework encompassing both non-Markovian and Markovian descriptions. If the infection and removal time distributions are exponential: $\psi_{\inf}(\tau) = \gamma e^{-\gamma\tau}$ and $\psi_{\rm rem}(\tau) = \mu e^{-\mu\tau}$, Eqs. (1–3) can be reformulated into a Markovian theory and further simplified into a set of ordinary differential equations with the constant infection and removal rates γ and μ (a detailed derivation of these equations is presented in Supplementary Note 2) :

$$\frac{ds_l(t)}{dt} = -s_l(t)k\gamma\sum_{m=1}^{n} A_{lm}p_m i_m(t), \tag{4}$$

$$\frac{di_l(t)}{dt} = s_l(t)k\gamma\sum_{m=1}^{n} A_{lm}p_m i_m(t) - \mu i_l(t), \tag{5}$$

$$\frac{dr_l(t)}{dt} = \mu i_l(t). \tag{6}$$

While the non-Markovian and Markovian theories [Eqs. (1–3) and Eqs. (4–6), respectively] accurately describe the memory-dependent and memoryless Monte Carlo simulations, we focus on whether the Markovian theory can accurately capture memory-dependent dynamics and how memory effects influence its accuracy. For this purpose, we seek to establish the equivalence between Markovian and non-Markovian approaches for describing spreading dynamics.

Dynamical equivalence

Equations (1–6) provide a base to study the steady-state equivalence and transient-state equivalence between non-Markovian and Markovian theories, where a steady state characterizes the long-term dynamics of disease spreading and a transient state is referred to as the short-term behavior prior to system's having reached the steady state. As illustrated in Fig. 1c, note that steady-state equivalence means that the two types of spreading dynamics attain identical steady states[18,21,22], whereas transient-state equivalence implies that two types of dynamics are consistent throughout the entire transmission period. Transient-state equivalence thus implies steady-state equivalence, but not vice versa. Since Eqs. (1–6) also provides a numerical framework for Monte Carlo simulations, the terms "steady-state equivalence" and "transient-state equivalence" not only describe the connection between non-Markovian and Markovian theories, but also illustrate the relationship between memory-dependent and memoryless processes. Therefore, the equivalence between the two theories implies the equivalence between the two corresponding processes, and vice versa.

Steady-state equivalence. Equations (1)–(3) give the following transcendental equation for determining the steady state (see Supplementary Note 3 for detailed derivation):

$$\tilde{s}_l = \acute{s}_l e^{-\frac{R_0}{\Lambda_{\max}}\sum_{m=1}^{n} kA_{lm}p_m(\tilde{r}_m - \acute{r}_m)}, \tag{7}$$

where $\tilde{s}_l = \lim_{t\to+\infty} s_l(t)$ and $\tilde{r}_l = \lim_{t\to+\infty} r_l(t)$ denote the fractions of the susceptible and removed individuals in age group l at the steady state (note that $\tilde{s}_l = 1 - \tilde{r}_l$, because at steady state, no infection exists), while $\acute{s}_l = s_l(0)$ and $\acute{r}_l = r_l(0)$ represent the initial fractions of the susceptible and removed individuals in this age group. For non-Markovian dynamics, basic reproduction number R_0 can be determined by:

$$R_0 = \Lambda_{\max}\int_0^{+\infty} \omega_{\inf}(\tau)\Psi_{\rm rem}(\tau)d\tau. \tag{8}$$

For Markovian dynamics, R_0 is given by:

$$R_0 = \frac{\gamma\Lambda_{\max}}{\mu}. \tag{9}$$

where $\Lambda_{\max}$ is the maximum eigenvalue of the matrix $k\mathbf{A}\circ\mathbf{p}$, and $\circ$ denotes a row-wise Hadamard product between a matrix and a vector (see Supplementary Note 3 for a detailed description). Since Eq. (7) applies to both non-Markovian and Markovian dynamics, an identical R_0 value in the two cases will result in equivalent steady states from the same initial conditions. Consequently, for a given non-Markovian spreading process, there exists an infinite number of Markovian models with the same steady state, as the R_0 value is only determined by the ratio of γ to μ, but not by either value.

As shown in Fig. 2d, memory-dependent and memoryless spreading dynamics that reach the same steady state with the identical R_0 value confirm the steady-state equivalence. Figure 2e demonstrates that, even for R_0 ranging from 0.023 to 4.63, the equivalent memory-dependent and memoryless spreading dynamics still produce highly consistent steady states that can be calculated from Eq. (7), which share the same critical point of phase transition at $R_0 = 1$.

Transient-state equivalence. From the preceding section, we used the basic reproduction number R_0, a fundamental metric quantifying the number of secondary infections generated by a single individual, to characterize the steady-state equivalence. Here, we propose to quantify the transient-state equivalence through the average generation time $T_{\rm gen}$ that measures the "velocity" at which secondary infections occur. This time can be calculated as[2,53]:

$$T_{\rm gen} = \int_0^{+\infty} \tau\psi_{\rm gen}(\tau)d\tau,$$

where

$$\psi_{\rm gen}(\tau) = \frac{\omega_{\inf}(\tau)\Psi_{\rm rem}(\tau)}{\int_0^{+\infty}\omega_{\inf}(\tau')\Psi_{\rm rem}(\tau')d\tau'} \tag{10}$$

is the generation time distribution. Effectively, $T_{\rm gen}$ measures the average duration of disease transmission from an infected individual to the next generation of individuals. Likewise, the average infection time $T_{\inf}$ and the average removal time $T_{\rm rem}$ are defined as the mean values of the infection and removal time distributions:

$$T_{\inf} = \int_0^{+\infty} \tau\psi_{\inf}(\tau)d\tau,$$
$$T_{\rm rem} = \int_0^{+\infty} \tau\psi_{\rm rem}(\tau)d\tau.$$

In calculating $T_{\rm gen}$, the individual's removal is taken into account, while $T_{\inf}$ measures the average time of the first disease transmission of an infectious individual without factoring in removal. In the classical memoryless transmission with exponential distributions $\psi_{\inf}(\tau)$ and $\psi_{\rm rem}(\tau)$, the equality $T_{\rm gen} = T_{\rm rem}$ holds. However, for memory-dependent spreading, the possible scenarios are: $T_{\rm gen} = T_{\rm rem}$, $T_{\rm gen} < T_{\rm rem}$, or $T_{\rm gen} > T_{\rm rem}$. Specifically, because $T_{\rm gen}$, $T_{\inf}$ and $T_{\rm rem}$ all represent the mean values of distributions, it is possible for $T_{\rm rem}$ to be shorter than $T_{\rm gen}$ or $T_{\inf}$ in some situations. And our web-based application demonstrates the impact of parameters on the time distributions (infection, removal, and generation) as well as their average times[50].

For a non-Markovian spreading process, if the equality $T_{gen} = T_{rem}$ holds, it can be approximately equivalent to a Markovian one in the transient state, because it satisfies the following equation:

$$s_l(t) \simeq \acute{s}_l e^{-\frac{R_0}{\Lambda_{max}}\sum_{m=1}^{n} kA_{lm}p_m[r_m(t)-\acute{r}_m]}, \quad (11)$$

which exhibits a memoryless transmission pattern similar to Markovian dynamics (see Supplementary Note 4 for a detailed analysis). Intuitively, the equality $T_{gen} = T_{rem}$ signifies that the infection and removal processes occur concurrently, which in turn leads to a memoryless relationship between the two processes, thereby minimizing the memory effects. Furthermore, to determine the corresponding Markovian parameters γ and μ of the Markovian transmission which is equivalent to the non-Markovian dynamics in the transient state, we need to utilize the Euler-Lotka equation[45,53,54]:

$$1 = R_0 \int_0^{+\infty} e^{-g\tau}\psi_{gen}(\tau)d\tau, \quad (12)$$

where g denotes the growth rate of the non-Markovian dynamics and is another measure of how quickly the epidemic is spreading within a population. Therefore, we can calculate the values of the basic reproduction number, R_0, and growth rate, g, in the non-Markovian dynamic by using Eqs. (8), (10), and (12). Additionally, the Markovian form of $\psi_{gen}(\tau)$ according to Eq. (10) is $\mu e^{-\mu\tau}$, and the equivalent Markovian and non-Markovian dynamics in the transient state have the same values of R_0 and the equal values of g. By substituting $\psi_{gen}(\tau) = \mu e^{-\mu\tau}$ and the calculated R_0 and g into Eq. (12), we can determine the value of μ. Furthermore, using Eq. (9), we can find the value of γ based on μ. Hence, the Markovian parameters γ and μ are determined as follows:

$$\gamma = \frac{gR_0}{\Lambda_{max}(R_0 - 1)}, \quad (13)$$

$$\mu = \frac{g}{R_0 - 1}. \quad (14)$$

And we provide visualizations that illustrate how the values of γ and μ are influenced by the distribution parameters in our web-based application[50].

As illustrated in Fig. 2f–g, when the equality $T_{gen} = T_{rem}$ holds for the non-Markovian dynamics, Eq. (11) holds, which can be seen by comparing the susceptible curve calculated from Eq. (1) to that inferred from Eq. (11), as shown in Fig. 2f. In this case, the Markovian spreading curves deduced from Eqs. ((13) and (14)) closely align with the non-Markovian transient curves, as shown in Fig. 2g. However, as shown in Fig. 2h–i, if the equality does not hold, the equivalence in transient states breaks down. It is important to note that the Euler-Lotka equation assumes an exponential growth of a disease outbreak and is only reasonable at the initial stage. Consequently, as the cumulative infections increase (Fig. 2g), the Markovian curves will exhibit slight deviations from the non-Markovian counterparts. Meanwhile, because the equivalent dynamics share the same R_0, they will ultimately reach the same steady state, ensuring that deviations will diminish while they approach the steady state.

To evaluate, under different values of the generation-to-removal time ratio $\eta \equiv T_{gen}/T_{rem}$ for non-Markovian dynamics we introduce a metric, ε, to quantify the difference from the corresponding Markovian results calculated from Eqs. ((13) and (14)) (see "Method" for detailed definition of ε). Figure 2j shows, for non-Markovian numerical calculations, five scenarios under various forms of time distributions $\psi_{inf}(\tau)$ and $\psi_{rem}(\tau)$ constrained by certain average infection and removal times: Weibull, $T_{inf} = 5$, $T_{rem} = 7$; Weibull, $T_{inf} = 7$, $T_{rem} = 7$; Weibull, $T_{inf} = 5$, $T_{rem} = 5$; log-normal, $T_{inf} = 5$, $T_{rem} = 7$; and gamma, $T_{inf} = 5$, $T_{rem} = 7$ (see "Method" for detailed definitions of Weibull, log-normal, and gamma distributions). The T_{gen} value is adjusted to obtain different values of $\ln\eta$. For $\ln\eta = 0$, i.e., $T_{gen} = T_{rem}$, the "distance" ε between the transient states of non-Markovian and Markovian dynamics with parameters determined from Eqs. ((13)–(14)) is minimal. Otherwise, ε increases as $\ln\eta$ deviates from zero. Meanwhile, Fig. 2j shows that ε depends primarily on the ratio of T_{gen} to T_{rem}, but rarely on the values of T_{gen}, T_{inf}, or T_{rem}. Meanwhile, the specific form of the time distributions also has limited influence on it.

Furthermore, it is important to note that the condition where $T_{gen} = T_{rem}$ in a non-Markovian dynamic ensures transient-state equivalence between this non-Markovian transmission and a Markovian one, but according to Eqs. (13) and (14), it does not imply that the average generation and removal times of the non-Markovian dynamic must be equal to those of the equivalent Markovian one. For instance, if a non-Markovian dynamic satisfies the condition of transient-state equivalence and we keep its average generation and removal times fixed, altering the shape of the corresponding time distributions will change the transmission speed[53]. This change, in turn, affects the infection and removal rates of the equivalent Markovian dynamic, leading to different average generation and removal times for the Markovian equivalent dynamic (see Supplementary Fig. S1 and Supplementary Notes 5 and 6 for a detailed analysis).

Markovian approximation of memory-dependent spreading dynamics

As illustrated in Fig. 1d, testing the applicability of Markovian theory for memory-dependent spreading dynamics requires three steps. The first step is fitting, where the memory-dependent Monte Carlo simulation data are divided into two parts: (a) a short early stage used as the training data for fitting the Markovian parameters in Eqs. ((4)–(6)), and (b) the remaining testing data for evaluating the performance of the Markovian model (see "Method" for details of the fitting procedure). The second step is to employ the Markovian model, equipped with fitted Markovian parameters, to accomplish various tasks, such as estimating R_0, predicting outbreaks and assessing the prevention effects of different vaccination strategies. The third step is testing, i.e., evaluating the accuracy of the Markovian model, e.g., by comparing the estimated and actual R_0 values, disease outbreaks and prevention effects. As real-world disease spreading is subject to environmental, social and political disturbing factors, for the fitting and testing steps, we conduct Monte Carlo simulations of stochastic memory-dependent disease outbreaks to generate the training and testing data.

Here, we first analyze the influence of η on the estimation of R_0 using the Markovian theory, and design two tasks to evaluate the applicability of the theory in epidemic forecasting and prevention evaluation of memory-dependent spreading. For comparison, we also generate the corresponding results from the non-Markovian theory in the two tasks.

Estimation of basic reproduction number. Estimating basic reproduction number R_0 is crucial for determining the ultimate prevalence of disease spreading and for assessing the effectiveness of various disease containment measures[45,53,54]. When using the Markovian theory to fit the early-stage transmission of a memory-dependent process, a key parameter that can affect the estimation of R_0 is the ratio η. To develop an analysis, recall the basic principle for estimating R_0: disease spreading dynamics can be viewed as a combination of two parallel processes: infection and removal. In particular, the infection process is the reproduction of the disease within each generation, where each infected individual generates an average of R_0 newly infected individuals in the subsequent generation after a mean time period T_{gen}. In the removal process, infected individuals are removed from the spreading chain, where each generation takes an average time T_{rem} to be removed. For a Markovian type of dynamics with constant γ and μ, the equality $T_{gen} = T_{rem}$ holds. Consequently, during the Markovian fitting step, the average number of new infections upon the removal of a single infected individual is taken as the value of R_0. For memory-dependent spreading, if the equality $T_{gen} = T_{rem}$ holds, the memory-dependent spreading curves will possess an approximate memoryless feature so that R_0 can be still be estimated by counting the number of new infections at the time when the current generation of infections is removed, as shown in Fig. 3a. However, for $T_{gen} < T_{rem}$, more than one generation is produced while

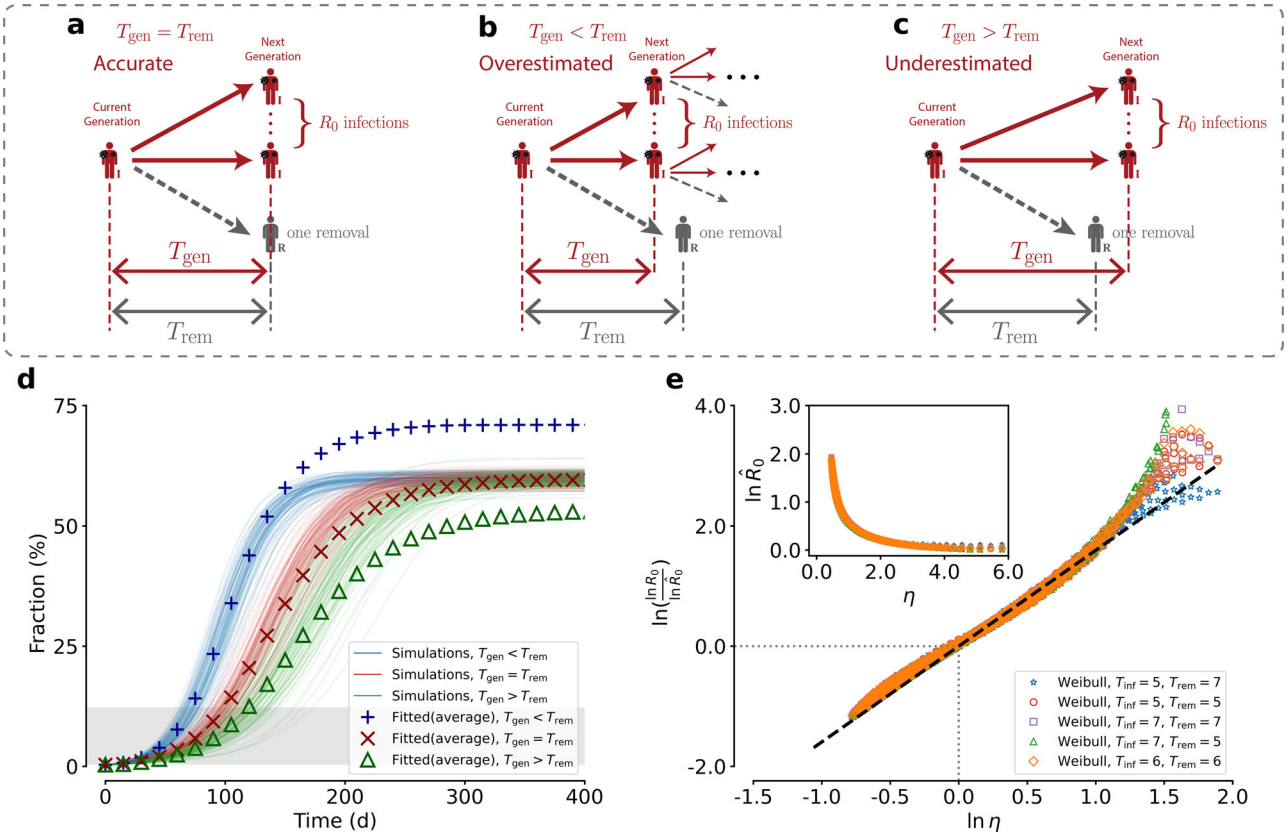

Fig. 3 | Estimation of R_0. a–c Mechanism of the R_0 estimation. The red arrows represent the infection process of the next generation by the current generation, while the dashed arrows denote the removal of the current generation. The relationship between T_{gen} and T_{rem} influences the number of new infections when the current generation of infections is removed. For $T_{gen} = T_{rem}$, the number of new infections is exactly R_0. For $T_{gen} < T_{rem}$, the number of new infections is greater than R_0. For $T_{gen} > T_{rem}$, the number of new infections is smaller than R_0. **d** There are three distinct categories of memory-dependent disease spreading, all sharing the same value of the basic reproduction number R_0 as 1.86: $T_{gen} < T_{rem}$ (blue curves; $\alpha_{inf} = 1.5, \beta_{inf} = 4.64, \alpha_{rem} = 2.5, \beta_{rem} = 5$), $T_{gen} = T_{rem}$ (red curves; $\alpha_{inf} = 2.5, \beta_{inf} = 5, \alpha_{rem} = 2.5, \beta_{rem} = 5$), and $T_{gen} > T_{rem}$ (green curves; $\alpha_{inf} = 3.5, \beta_{inf} = 5.32, \alpha_{rem} = 2.5, \beta_{rem} = 5$), where the fractions of cumulative infected individuals (i.e., sum of infected and removed fractions) are simulated using 100 independent realizations. The predicted future evolution of the spreading dynamics by the Markovian theory with the fitted parameters are also shown: $T_{gen} < T_{rem}$ (blue + symbols), $T_{gen} = T_{rem}$ (red × symbols), and $T_{gen} > T_{rem}$ (green Δ symbols). The gray area marks the training data (see Method for detailed approach of training data selection). **e** The relationship between $\ln(\ln R_0 / \ln \hat{R}_0)$ (R_0 represents the real basic reproduction number, while $\hat{R}_0$ denoted the estimated one) and $\ln \eta$. The horizontal and vertical dotted lines show that the equality between T_{gen} and T_{rem} results in an accurate estimation of R_0 and the dashed line represents a linear fitting with the slope 1.59. Inset: the relation between $\ln \hat{R}_0$ and η with the asymptotic behaviors: for $\eta \to 0$, $\hat{R}_0 \to +\infty$ (i.e., $\ln \hat{R}_0 \to +\infty$), and for $\eta \to +\infty$, $\hat{R}_0 \to 1$ (i.e., $\ln \hat{R}_0 \to 0$).

the current generation is removed, R_0 estimated by the Markovian theory will represent an overestimate, as shown in Fig. 3b. For $T_{gen} > T_{rem}$, less than one generation is created during T_{rem}, the Markovian theory will give an underestimate of R_0, as shown in Fig. 3c.

Figure 3d shows the Markovian fitting for memory-dependent spreading simulation curves, which exhibit identical values of R_0 under $T_{gen} = T_{rem}$ (red curves), $T_{gen} < T_{rem}$ (blue curves), and $T_{gen} > T_{rem}$ (green curves). When the equality $T_{gen} = T_{rem}$ holds, the Markovian theory with fitted parameters generates accurate predictions of the future evolution (red × symbols). For $T_{gen} < T_{rem}$, the outbreak in the initial stage is accelerated, resulting in an overestimation by the Markovian theory (blue + symbols). For $T_{gen} > T_{rem}$, the initial outbreaks are decelerated, leading to an underestimation by the Markovian approach (green Δ symbols).

The above qualitative insights lead to a semi-empirical relationship between the Markovian-estimated basic reproduction number $\hat{R}_0$ and its actual value R_0 as:

$$\hat{R}_0 = (R_0)^{\eta^{-a}}, \tag{15}$$

where a is a positive coefficient (see "Method" for a detailed derivation). The value of a is a crucial and constant parameter in Eq. (15), and it needs to be determined by fitting it to the data. Once this constant a is obtained, the actual value of R_0 can be derived by adjusting the estimated $\hat{R}_0$ based on Eq. (15), and more accurate steady state can be calculated by using Eq. (7).

Equation (15) implies the relationship $\ln(\ln R_0 / \ln \hat{R}_0) = a \ln \eta$. We use Weibull time distributions $\psi_{inf}(\tau)$ and $\psi_{rem}(\tau)$ constrained by certain average infection and removal times: $T_{inf} = 5$, $T_{rem} = 7$; $T_{inf} = 7$, $T_{rem} = 7$; $T_{inf} = 5$, $T_{rem} = 5$; $T_{inf} = 7$, $T_{rem} = 5$; and $T_{inf} = 6$, $T_{rem} = 6$ for memory-dependent Monte Carlo simulations. Figure 3e shows the linear relationship between $\ln(\ln R_0 / \ln \hat{R}_0)$ and $\ln \eta$, providing support for our qualitative analysis of the Markovian estimation. The estimation of R_0 also depends on the ratio η and is relatively insensitive to the specific values of T_{gen}, T_{inf}, or T_{rem} for the same form of time distribution. The results in the inset of Fig. 3e further confirm that the estimated $\hat{R}_0$ approaches 1 when T_{gen} is much larger than T_{rem} and tends to $+\infty$ when T_{gen} is much smaller than T_{rem}. By fitting the available data, we have determined the value of a to be 1.59. After obtaining the value of a, we can develop our web-based application for rectifying R_0 and epidemic forecasting[50]. We also plot additional experimental results for log-normal and gamma distributions (log-normal, $T_{inf} = 5$, $T_{rem} = 7$; and gamma, $T_{inf} = 5$, $T_{rem} = 7$) in Supplementary Note 7. Referring to Supplementary Fig. S2, various forms of time distributions may affect the Markovian estimation of R_0. However, their influence appears less compared to the impact of the value of η. Hence, it can be concluded that the distribution form seldom affects the estimation of R_0.

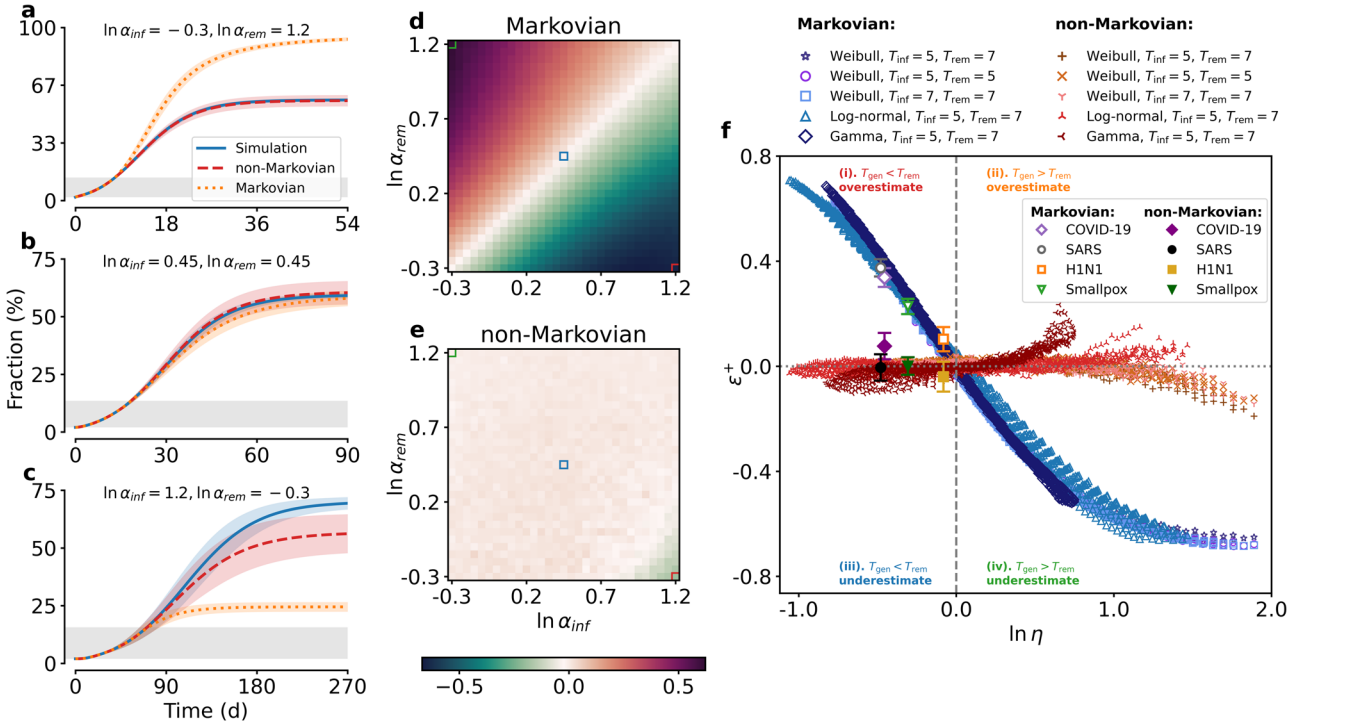

Fig. 4 | Epidemic forecasting. a–c Predicted evolution of the cumulative infected fraction (i.e., sum of infected and removed fractions) by the Markovian (orange dotted curves) and non-Markovian (red dashed curves) theories, in comparison with the Monte Carlo simulations with Weibull time distributions (blue solid curves), for three sets of simulation parameters, respectively. The results are the averages of 100 independent realizations with the standard deviations indicated by the shaded regions. And all the memory-dependent Monte Carlo simulations exhibit a similar R_0 value of 1.83. The gray area marks the training data (see "Method" for detailed approach of training data selection). **d, e** The forecasting errors ε^+ of Markovian and non-Markovian theories with respect to the memory-dependent Monte Carlo simulations in the parameter plane of $\ln \alpha_{inf}$ and $\ln \alpha_{rem}$ in the range $[-0.3, 1.2]$. The green, blue and red squares mark the parameters of Monte Carlo simulations in (**a**–**c**), respectively. **f** The forecasting errors ε^+ from the Markovian and non-Markovian theories for different values of $\ln \eta$ under five scenarios of time distribution setting for Monte Carlo simulations. The corresponding estimations for a number of real-world diseases (COVID-19, SARS, H1N1 and Smallpox) are also included.

Epidemic forecasting. As suggested in Fig. 1d, we evaluate the efficacy of Markovian theory for epidemic forecasting. We use the early stage of Monte Carlo simulation data to fit parameters under both Markovian and non-Markovian hypotheses and then to predict future disease outbreaks. The remaining simulation data are leveraged to evaluate the accuracy of the Markovian and non-Markovian forecasting results. Regardless of the forms of time distributions in the memory-dependent Monte Carlo simulations (Weibull, log-normal, or gamma), the non-Markovian model fits the training data in a consistent manner, i.e., by assuming the actual time distributions as unknown and treating them as Weibull time distributions.

Figure 4a–c shows the evolution of the spreading dynamics from three types of memory-dependent Monte Carlo simulations with Weibull infection and removal distributions, where the shape parameters α_{inf} and α_{rem} are selected according to $\ln \alpha_{inf} = -0.3, \ln \alpha_{rem} = 1.2$ (Fig. 4a), $\ln \alpha_{inf} = 0.45, \ln \alpha_{rem} = 0.45$ (Fig. 4b), and $\ln \alpha_{inf} = 1.2, \ln \alpha_{rem} = -0.3$ (Fig. 4c), for $T_{inf} = 5$ and $T_{rem} = 7$. For the Weibull distributions, we have $\alpha_{inf} < \alpha_{rem}$, $\alpha_{inf} = \alpha_{rem}$ and $\alpha_{inf} > \alpha_{rem}$, corresponding to $T_{gen} < T_{rem}$, $T_{gen} = T_{rem}$, and $T_{gen} > T_{rem}$, respectively. We compare the simulated cumulative infected fractions to those predicted by the Markovian and non-Markovian theories. In general, the non-Markovian theory provides more accurate predictions than the Markovian theory. For the specific parameter setting $\ln \alpha_{inf} = 0.45, \ln \alpha_{rem} = 0.45$ (i.e., $T_{gen} = T_{rem}$), both theories yield a high accuracy.

The accuracy can be assessed through the forecasting error ε^+ that evaluates whether a theory overestimates or underestimates the steady-state cumulative infection, i.e., quantifying the extent of deviation between the results obtained from Markovian or non-Markovian theories and those derived from Monte Carlo simulations (see "Method" for detailed definition of ε^+). A plus value of ε^+ means overestimation while minus value indicates underestimation. We evaluate the accuracy measure ε^+ in the parameter plane of $\ln \alpha_{inf}$ and $\ln \alpha_{rem}$, ranging from -0.3 to 1.2. Figure 4d, e shows that the Markovian accuracy is sensitive to parameter changes: underestimated if α_{inf} is greater than α_{rem} ($T_{gen} > T_{rem}$), overestimated when α_{inf} is smaller than α_{rem} ($T_{gen} < T_{rem}$), and a high forecasting accuracy is achieved only for $\alpha_{inf} = \alpha_{rem}$ ($T_{gen} = T_{rem}$). In contrast, the non-Markovian theory yields highly accurate results in the whole parameter plane, with only a slight underestimation for $\alpha_{inf} \gg \alpha_{rem}$. This can be primarily attributed to the growing challenge of fitting simulation data, where the absolute derivatives of R_0 with respect to parameters become extremely high, making the fitting process much sensitive to the parameters[55,56]. This occurs when α_{inf} is greater than α_{rem} or T_{gen} is much larger than T_{rem}, causing the Monte Carlo curves to become more sensitive to the parameters and potentially deviate from the accurate theoretical curves (see Supplementary Fig. S3 and Supplementary Note 8 for details).

Using the five scenarios specified in Fig. 2j for memory-dependent Monte Carlo simulations, we obtain the relationship between ε^+ and $\ln \eta$, as shown in Fig. 4f. It can be seen that, in the Markovian framework, an overestimation arises for $T_{gen} < T_{rem}$, and an underestimation occurs for $T_{gen} > T_{rem}$. Only when $T_{gen} = T_{rem}$ is an accurate estimate achieved. In general, the non-Markovian theory provides much more accurate forecasting than the Markovian theory, especially when T_{gen} and T_{rem} are not equal. The results further illustrate that the specific values of T_{gen}, T_{inf}, or T_{rem} have little impact on forecasting accuracy, and the impact of the time distribution forms is much lower compared to the influence exerted by the η value.

To establish the relevance of these results to real-world diseases, we obtain the distributions of $\psi_{inf}(\tau)$ and $\psi_{rem}(\tau)$ for four known infectious diseases, including COVID-19, SARS, H1N1 influenza, and smallpox, using

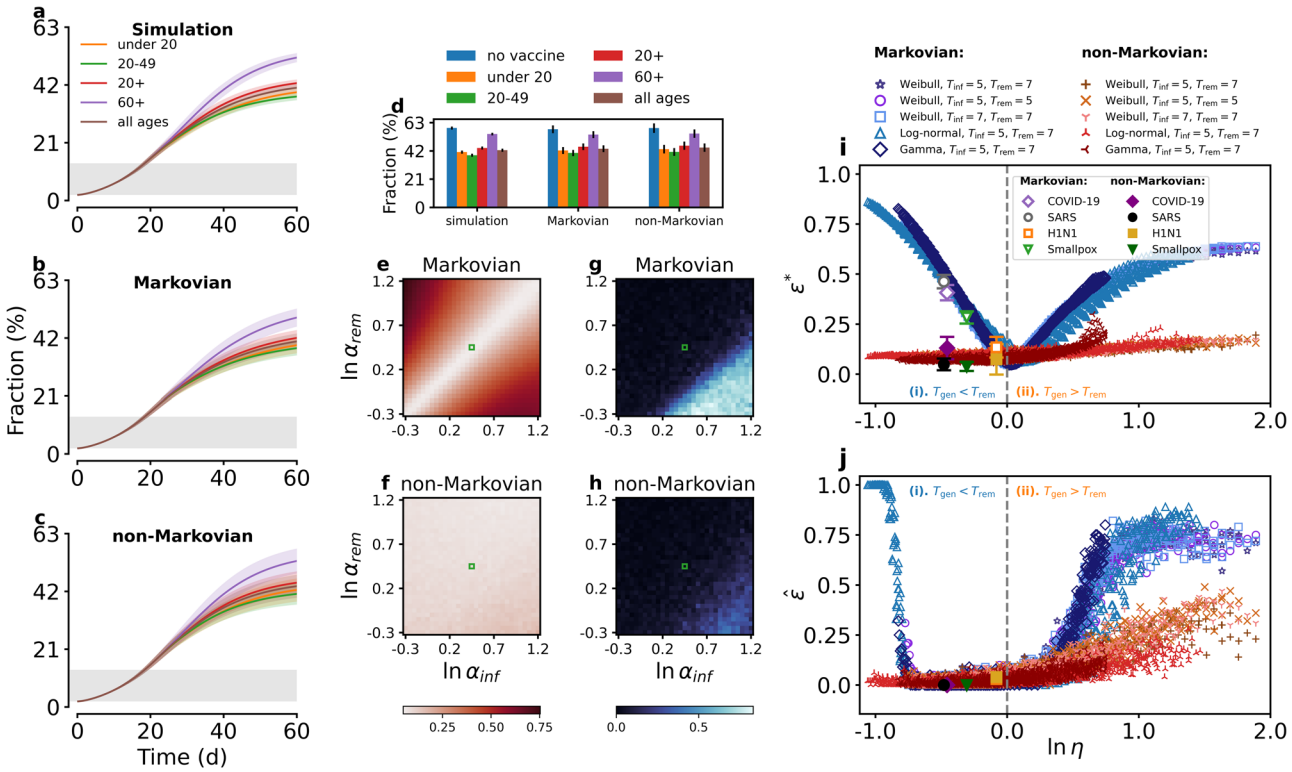

Fig. 5 | Evaluation of vaccination strategies. a–c For simulation parameters chosen according to $\ln \alpha_{inf} = 0.45$ and $\ln \alpha_{rem} = 0.45$, the cumulative infected fraction (i.e., sum of infected and removed fractions) curves from Monte Carlo simulations and the corresponding Markovian and non-Markovian theories with fitting parameters for five vaccination strategies. The average results are obtained from 100 independent realizations with the shaded regions representing the standard deviations. And all the memory-dependent Monte Carlo simulations exhibit a similar R_0 value of 1.83. The gray area marks the training data (see "Method" for detailed approach of training data selection). **d** Vector δ calculated from the results in Fig. 5a–c. **e**, **f** The prevention evaluation errors ε^* of Markovian and non-Markovian theories for evaluating the effects of vaccination prevention in the parameter plane ($\ln \alpha_{inf}$, $\ln \alpha_{rem}$). The green squares mark the selected parameters in (**a**–**c**). **g**, **h** The optimization failure probabilities $\hat{\varepsilon}$ arising from the Markovian and non-Markovian within the parameter plane ($\ln \alpha_{inf}$, $\ln \alpha_{rem}$). The green squares mark the selected parameters in (**a**–**c**). **i** The prevention evaluation errors ε^* from the Markovian and non-Markovian theories versus $\ln \eta$ under five scenarios of time distribution setting for Monte Carlo simulations. The estimated errors for four real diseases (COVID-19, SARS, H1N1, and Smallpox) are also shown. **j** The optimization failure probabilities $\hat{\varepsilon}$ from the Markovian and non-Markovian theories against $\ln \eta$ in five different time distribution scenarios for Monte Carlo simulations. The optimization failure probabilities for four real diseases (COVID-19, SARS, H1N1, and Smallpox) are also presented. The color and symbol theme is the same as (**i**).

the information in refs. [2,33–40]. We then calculate the corresponding values of ε^+ and $\ln \eta$ based on the Markovian and non-Markovian approaches. As demonstrated in Fig. 4f, the positions of the four diseases in the ($\ln \eta$, ε^+) plane are consistent with the results of our estimations. Because the data were from the reports of laboratory-confirmed cases incorporating the effects of the quarantine and distancing from susceptible individuals after the confirmation of the diagnosis, T_{gen} of the four diseases are all smaller than the corresponding values of T_{rem}, leading to some overestimation for the Markovian forecasting results.

Evaluation of vaccination strategies. In the development and application of a theory for disease spreading, assessing the effects of different vaccination strategies is an important task. Here we consider five prioritization strategies for vaccine distribution[5]: individuals under 20 years (denoted as $m = 1$), adults between 20 and 49 years ($m = 2$), adults above 20 years ($m = 3$), adults above 60 years ($m = 4$), and all age groups ($m = 5$), and implement these strategies in Monte Carlo simulations (see "Method" for the detailed procedure of vaccination in Monte Carlo simulations). Figure 5a shows the results of epidemic evolution with these vaccination strategies, where the shape parameters are chosen according to $\ln \alpha_{inf} = 0.45$ and $\ln \alpha_{rem} = 0.45$ ($T_{gen} = T_{rem}$). Figure 5b–c shows the results from the Markovian and non-Markovian theories, respectively, with the corresponding fitted parameters for the vaccination strategies (see "Method" for the detailed procedure of vaccination in theoretical calculations). These results indicate that the Markovian and non-Markovian theories yield the correct epidemic evolution and future outbreaks under different vaccination scenarios, when $T_{gen} = T_{rem}$.

To characterize the effectiveness of different vaccination strategies in blocking disease transmission, we introduce a vaccination effectiveness vector, δ, whose m-th element quantifies the cumulative infected fraction with the m-th vaccination strategy in the steady state: $\delta_m = \tilde{c}_m$, for $m = 0, \ldots, 5$ (subscript $m = 0$ indicates the results without vaccination). Figure 5d shows that the δ vectors from the Monte Carlo simulation, Markovian and non-Markovian theories from Fig. 5a–c, respectively.

We further introduce a metric, the so-called prevention evaluation error ε^*, that gauges the ability of the Markovian and non-Markovian theories to estimate the total effectiveness of vaccination, i.e., measuring the disparity between the results calculated by the Markovian or non-Markovian theories and those obtained through Monte Carlo simulations considering various vaccination strategies (see "Method" for the detailed definition of ε^*). Figure 5e, f shows the average values of ε^* of the two theories in the simulation parameter plane using 100 independent realizations, which are similar to those in Fig. 4d–e, indicating that the error mainly comes from the R_0 estimation. In general, the Markovian theory performs well only in the diagonal area of the parameter plane where $\alpha_{inf} = \alpha_{rem}$, as shown in Fig. 5e, and the non-Markovian theory outperforms the Markovian counterpart in most cases, as shown in Fig. 5f.

Meanwhile, we assess the ability of both the Markovian and non-Markovian theories to detect the optimal vaccination strategy. We also define a quantity, optimization failure probability $\hat{\varepsilon}$, to quantify the probability of a theory failing to identify the optimal strategy, i.e., that leads to the lowest cumulative infection among these strategies (see "Method" for the detailed definition of $\hat{\varepsilon}$). Figure 5g–h illustrates the results of $\hat{\varepsilon}$ for the two theories within the parameter plane ($\ln \alpha_{\text{inf}}, \ln \alpha_{\text{rem}}$). While the non-Markovian theory still demonstrates superior performance, the Markovian approach proves capable of identifying the optimal strategy across a larger parameter space compared to the Markovian results depicted in Fig. 5e.

We obtain the relationships between ε^* and $\ln \eta$, as well as between $\hat{\varepsilon}$ and $\ln \eta$, as shown in Fig. 5i–j with the same five time-distribution scenarios as in Fig. 2j. In all cases of Fig. 5i, ε^* reaches a minimum for $\ln \eta = 0$ and increases as $\ln \eta$ deviates from zero. The results from the five scenarios further illustrates that the specific values of T_{gen}, T_{inf}, or T_{rem} play little role in the errors in vaccination evaluation, and the effect of time distribution forms is also limited. Figure 5i also includes the values of ε^* for the real-world infectious diseases COVID-19, SARS, H1N1 influenza, and smallpox, which are consistent with those from the non-Markovian and Markovian theories. Regarding the results depicted in Fig. 5j, it is observed that the non-Markovian theories consistently outperform the Markovian counterparts. On the other hand, within a wide range of $\ln \eta$ values around 0, the Markovian theories successfully identify the optimal vaccination strategy among various commonly employed ones. When the value of $\ln \eta$ deviates from 0, Markovian theories become ineffective in determining the optimal strategy. (Note that on the left side of Fig. 5j, we only present the failures of Markovian theories to identify the optimal strategy in the Monte Carlo simulations with log-normal distribution. This is primarily due to the fact that the parameters associated with the Weibull and gamma distributions fall outside the acceptable range when we keep T_{inf} and T_{rem} fixed to modify $\ln \eta$ to a very low value.) Furthermore, we demonstrate that even when employing Markovian approaches, the optimal vaccination strategy can still be determined among these strategies considered for the four distinct real diseases.

Conducting accurate evaluations in prevention serves as the sufficient condition of the successful identification of the optimal strategy. In comparison to the prevention evaluation errors ε^* of Markovian theories, the optimization failure probability $\hat{\varepsilon}$ exhibits a wider range of $\ln \eta$ values that result in the lowest value. The lack of mathematical continuity among these strategies is the primary reason for this. It indicates that there is no smooth transition or mathematical relationship connecting these strategies, resulting in the rank of the strategies not changing promptly when the value of $\ln \eta$ deviates from 0. Therefore, only large errors from the Markovian theories can result in the failure to detect the optimal strategy. Based on this analysis, the extent to which $\ln \eta$ deviates from 0, leading to the failure of Markovian theories, as well as whether such failure will occur, depends on the selection of the tested strategies.

Discussion

The COVID-19 pandemic has emphasized the importance of investigating disease transmission in human society through modeling. Empirical observations have consistently demonstrated strong memory effects in real-world transmission phenomena. The initial transient stage of an epidemic is critical for data collection, prediction, and articulation of control strategies, but an accurate non-Markovian model presents difficulties. In contrast, a Markovian model offers great advantages in parameter estimation, computation, and analyses. Uncovering the conditions under which Markovian modeling is suitable for transient epidemic dynamics is necessary.

We have developed a comprehensive mathematical framework for both Markovian and non-Markovian compartmentalized SIR disease transmissions in an age-stratified population, which allows us to identify two types of equivalence between Markovian and non-Markovian dynamics: in the steady state and transient phase of the epidemic. Our theoretical analysis reveals that, in the steady state, non-Markovian (memory-dependent) transmissions are always equivalent to the Markovian (memoryless) dynamics. However, transient-state equivalence is approximate and holds when the average generation and removal times match each other. In particular, when the average generation time is approximately equal to the average removal time, the disease transmission and removal of an infected individual exhibit a memoryless correlation, thereby minimizing the memory effects of the dynamic process. This results in highly accurate results from the Markovian theory that captures the characteristics of memory-dependent transmission based solely on the early epidemic curves. Our analysis also suggests that the Markovian accuracy is mainly determined by the value of generation-to-removal time ratio in disease transmission, where a larger-than-one (smaller-than-one) ratio can lead to underestimation (overestimation) of the basic reproduction number and epidemic forecasting, as well as the errors in the evaluation of control or prevention measures. The estimation accuracy primarily depends on this ratio, but not greatly affected by the specific values at the various times associated with the epidemic; although distribution forms might affect accuracy to some extent, their influence is much less compared to the impact of the ratio. This property exhibits substantial practical importance, because the average generation and removal times can be readily assessed based on sparse data collected from the transient phase of the epidemic, but to estimate their distributions with only sparse data is infeasible. These results provide deeper quantitative insights into the influence of memory effects on epidemic transmissions, leading to a better understanding of the connection and interplay between Markovian and non-Markovian dynamics.

There were previous studies of the equivalence between Markovian and non-Markovian transmission in the SIS model[18,21,22]. However, these studies addressed the steady-state equivalence rather than the transient-state equivalence. To our knowledge, our work is the first to investigate the transient-state equivalence of the SIR model. In addition, previous studies mainly examined the impact of the average generation time on the transmission dynamics, such as how the shape of the generation time distribution affects the estimation of R_0[53] or the use of serial time distributions in estimating R_0 during an epidemic[45]. There was a gap in the literature regarding how generation times affect the accuracy of different models. Our paper fills this gap by providing a criterion for using Markovian frameworks to model memory-dependent transmission based on the relationship between the average generation and removal times.

From an application perspective, our study suggests that the impact of the time distribution forms on Markovian estimation accuracy is limited, making it easier to select models between Markovian and non-Markovian dynamics in the initial outbreak of an epidemic based primarily on the generation-to-removal time ratio. This insight is especially useful since detailed time distribution forms are often harder to detect than their corresponding mean values. In addition, we note that in previous studies, it was observed that in various scenarios, serial intervals, albeit with larger variances, are anticipated to possess a consistent mean value with the average generation time and are more straightforward to measure[2,45–49]. Given the practical difficulties in observing the generation time, our finding of minimal impact from the distribution forms suggests that the average serial interval can be utilized as a substitute of the average generation time to determine the applicability of the Markovian theories for modeling purposes without compromising accuracy, although numerous studies have indicated that replacing the generation time distribution with the serial interval distribution may affect the analysis of transmission dynamics[44,45,57]. Meanwhile, based on Eq. (15), if we determine the ratio of generation-to-removal time, the estimated R_0 obtained through the Markovian approach can be adjusted to approximate the true value. And our web-based application showcases the demonstration of rectifying R_0 and epidemic forecasting[50].

Our study highlights the critical importance of accurately quantifying R_0 for achieving precise epidemic forecasting and prevention evaluation. A previous work[54] revealed that the value of R_0 depends on three key components: the duration of the infectious period (e.g., $\psi_{\text{rem}}(\tau)$), the probability of infection resulting from a single contact between an infected individual and a susceptible one (e.g., $\psi_{\text{inf}}(\tau)$), and the number of new susceptible individuals contacted per unit of time. However, given the practical limitations inherent in obtaining all three components, numerous methods

have been developed for estimating R_0. Although our work presents a specific approach, which fits the parameters of exponential or non-exponential time distribution by using the initial outbreak curves, it is not the only one available. For example, when contact patterns are unknown, R_0 can be estimated by fitting the growth rate g and the generation time distribution $\psi_{\text{gen}}(\tau)$, and then applying them in the Euler-Lotka equation[45,53,54]. However, since the focus of our work is on epidemic forecasting and evaluation of prevention measure where the data of contact patterns are given, R_0 can be directly calculated once $\psi_{\text{inf}}(\tau)$ and $\psi_{\text{rem}}(\tau)$ are fitted, without requiring the fitting of any additional quantity such as growth rate g. The estimation of R_0 can also be achieved by using data in the steady state, such as the final size of an epidemic or equilibrium conditions[54]. However, this method is not suitable for the transient phase where only early-stage curves are available. Utilizing the approach delineated in this paper is practically more appropriate for estimating R_0.

While our study focused on transmission within the SIR framework, extension to SEIR or SIS models is feasible. While we emphasized the significance of the transient-state equivalence in disease transmission, transient dynamics are more relevant or even more crucial than the steady state in nonlinear dynamical systems[58]. For example, in ecological systems, transient dynamics play a vital role in empirical observations and are therefore a key force driving natural evolution[59–64]. In neural dynamics, transient changes in neural activity can mediate synaptic plasticity, a crucial mechanism for learning and memory[65–67]. Therefore, the identification of suitable conditions for choosing between Markovian and non-Markovian dynamics may not be limited to epidemic dynamics alone and may serve as a valuable reference for other fields as well.

Taken together, our study establishes an approximate equivalence between Markovian and non-Markovian dynamics in the transient state, assuming that time distributions follow Weibull forms (see Supplementary Note 4 for details). While the applicability of our findings to most synthetic and empirical distributions has been analyzed qualitatively, a quantitative analysis requires further studies. For extreme cases with non-Weibull distributions, the transmission should be evaluated using other specific methods. While we have provided a qualitative analysis of the mechanism underlying why time distribution forms have minimal impacts on the errors of Markovian estimations, a more rigorous theoretical analysis is needed and requires further exploration. Meanwhile, in the spreading process, there are other memory effects beyond the "temporal" memory considered in our study, for example, the transmission capacity can be influenced by the previous transmission path[42], which is referred to as "spatial" memory effect. However, our study specifically focuses on analyzing the impact of "temporal" memory effects. To understand how "spatial" memory effects affect the accuracy of the Markovian approach, further research is needed. In addition, due to the complexity of the nonlinear transmission, our study has produced a semi-empirical relationship to estimate the overestimation and underestimation of Markovian methods. Further research is required to develop a rigorous formula that can accurately predict these effects. Additionally, the non-Markovian SIR model utilized in this study shares similarities with the Hawkes or HawkesN process[41,68] (see Supplementary Note 9 for details). These processes are commonly employed for the statistical modeling of events in diverse fields, where random events exhibit self-exciting behavior. We anticipate that our research will offer fresh perspectives and valuable insights to advance the understanding of related studies in this area.

Method

Monte Carlo simulation

In the simulation, we classify N individuals into n subgroups based on the age distribution $\mathbf{p}$. The index of the subgroup to which an individual belongs is denoted by l (where $1 \le l \le n$), and the index of the individual within the subgroup is denoted by u (where $1 \le u \le p_l N$). The state of the individual u in the age group l is represented by X_{lu}, which includes the states S (susceptible), I (infected), W (recovered), and D (dead), where W and D both represent R (removed), while I and R both represent C (cumulative infected). For each individual, we also record the absolute time of infection and removal using two variables: t_{lu}^{inf} and t_{lu}^{rem}, respectively. The absolute time of the system is denoted by t, and we implement the total spreading simulation step by step using a finite time step Δt as follows:

i. Initialization: set $t = 0$, X_{lu} for every individual is set to S.
ii. Set infection seeds: choose a set of individuals as the infection seeds and the corresponding X_{lu} are set to I, the corresponding t_{lu}^{inf} are set to 0, and t_{lu}^{rem} are set to a random value following the removal time distribution $\psi_{\text{rem}}(\tau)$.
iii. Infection in one time step: calculate the infection rate, $\hat{\omega}_{lu}^{\text{inf}}(t)$, of infected individual u in age group l during the current time step by

$$\hat{\omega}_{lu}^{\text{inf}}(t) = 1 - \Psi_{\text{inf}}(t - t_{lu}^{\text{inf}} + \Delta t)/\Psi_{\text{inf}}(t - t_{lu}^{\text{inf}}).$$

The probability $\bar{\omega}_l^{\text{inf}}(t)$ of each susceptible individual in age group l being infected can be calculated by

$$\bar{\omega}_l^{\text{inf}}(t) = \sum_{m=1}^{n} p_m \left[1 - \left(1 - \frac{\sum_{v \in \mathcal{I}_m(t)} \hat{\omega}_{mv}^{\text{inf}}(t)}{p_m N}\right)^{k\mathbf{A}_{lm}}\right],$$

where $\mathcal{I}_m(t)$ is the index set of the infected individuals in age group m at time t. The number of the susceptible individuals being infected in age group l follows a binomial distribution $B(s_l^*(t)p_l N, \bar{\omega}_l^{\text{inf}}(t))$, where $s_l^*(t)$ denotes the fraction of susceptible individuals in age group l at time t. Then generate a random number $N_l(t)$ following this binomial distribution and randomly select $N_l(t)$ susceptible individuals in age group l to set them as I state. The corresponding t_{lu}^{inf} of the new infected individuals are set to the current t and t_{lu}^{rem} are set to $\tau_{\text{rem}} + t$, where the random τ_{rem} follow the removal time distribution $\psi_{\text{rem}}(\tau)$.

iv. Removal in one time step: check if t_{lu}^{rem} of each infected individual is during the current time step. If this condition is satisfied, set their state to D with the probability σ_l and to W otherwise, where σ_l is the infection fatality rate of age group l[69]. Then let $t \leftarrow t + \Delta t$.
v. Repeat the process iii) and iv), until no individual with I index exists.

Time distributions

In the numerical calculations or Monte Carlo simulations, we employ three types of time distributions, i.e., Weibull, log-normal, and gamma, to describe the non-Markovian or memory-dependent transmission process.

For Weibull time distribution, it follows:

$$\psi(\tau) = \frac{\alpha}{\beta}\left(\frac{\tau}{\beta}\right)^{\alpha-1} e^{-(\frac{\tau}{\beta})^{\alpha}},$$

where α and β denote the shape and scale parameters, respectively.

The log-normal time distribution is defined as follows:

$$\psi(\tau) = \frac{1}{\tau\beta\sqrt{2\pi}} \exp\left(-\frac{(\ln\tau - \alpha)^2}{2\beta^2}\right).$$

The gamma time distribution is expressed as follows:

$$\psi(\tau) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \tau^{\alpha-1} e^{-\frac{\tau}{\beta}},$$

where $\Gamma(\cdot)$ denotes gamma function, while α and β represent the shape and scale parameters, respectively.

Additionally, the survival function could be calculated by:

$$\Psi(\tau) = \int_{\tau}^{+\infty} \psi(\tau')d\tau',$$

while the hazard function could be derived from:

$$\omega(\tau) = \frac{\psi(\tau)}{\int_{\tau}^{+\infty} \psi(\tau')d\tau'}.$$

For time distribution, survival function, hazard function and the corresponding parameters, incorporating subscripts "inf" and "rem" signifies those related to infection and removal, respectively. For example, $\psi_{\text{inf}}(\tau)$ represents the time distribution of infection, while α_{inf} denotes its shape parameter.

Derivation of semi-empirical estimation of basic reproduction number

Intuitively, the period of T_{rem} can accommodate $T_{\text{rem}}/T_{\text{gen}} = 1/\eta$ time intervals of length T_{gen}, corresponding to the result of $1/\eta$ generations of infections. This can lead to an exponential increase in the number of infections during T_{rem}. This intuition suggests a relationship between the fitted basic reproduction number $\hat{R}_0$ and the actual R_0, which can be expressed as an exponential function:

$$\hat{R}_0 = R_0^{f(1/\eta)},$$

where $f(\cdot)$ is a monotonically increasing function that satisfies three conditions. First, $f(1) = 1$, indicating that $\hat{R}_0$ can be accurately estimated when $T_{\text{gen}} = T_{\text{rem}}$. Second, $f(0) = 0$, meaning that if T_{rem} is an extremely small fraction of T_{gen}, the transmission will take a long time to reach the steady state, causing the curve to be flat in the initial stage and potentially causing the Markovian fitting to produce the estimate $\hat{R}_0 = 1$. Third, $f(+\infty) = +\infty$, implying that if T_{rem} is extremely large compared to T_{gen}, the transmission will quickly reach the steady state, causing the Markovian fitting to give an extremely large estimate of $\hat{R}_0$. Because the actual transmission process involves many complicated nonlinear relationships, identifying the specific form of the function $f(\cdot)$ is a challenging task. We thus assume

$$f(x) = x^a,$$

where a is an unknown positive coefficient. This leads to Eq. (15).

Definition of errors

The difference ε between non-Markovian and the corresponding Markovian results calculated from Eqs. ((13), (14)) is defined as:

$$\varepsilon = \sum_{x^{\ddagger},x^{\dagger}\in\{(s^{\ddagger},s^{\dagger}),(i^{\ddagger},i^{\dagger}),(r^{\ddagger},r^{\dagger})\}} \frac{\| x^{\ddagger} - x^{\dagger}\|_{2,\mathbb{T}_{\theta}^{\ddagger}}}{\| x^{\ddagger}\|_{2,\mathbb{T}_{\theta}^{\ddagger}}},$$

where the pairs $(s^{\ddagger}, s^{\dagger})$, $(i^{\ddagger}, i^{\dagger})$ and $(r^{\ddagger}, r^{\dagger})$ correspond to the non-Markovian and Markovian susceptible, infected and removed curves, respectively. The 2-norm $\|\cdot\|_{2,\mathbb{T}_{\theta}^{\ddagger}}$ on time duration $\mathbb{T}_{\theta}^{\ddagger}$ ensures that ε measures the "distance" between non-Markovian and the Markovian transient states. It is not appropriate to set time duration as the total transmission period because the cumulative infected fraction approaches the steady-state value asymptotically, making it difficult to determine the exact time point of the steady state. To address this issue, we choose $\mathbb{T}_{\theta}^{\ddagger}$ as $[0, t_{\theta}^{\ddagger}]$, where $t_{\theta}^{\ddagger}$ is the time when the non-Markovian cumulative infected fraction $c^{\ddagger}(t)$ reaches the θ percentile point within the range that spans from its initial value to its steady-state value. For instance, let $\acute{c}^{\ddagger}$ denote the cumulative infected fraction in non-Markovian calculation at time 0, i.e., $\acute{c}^{\ddagger} = c^{\ddagger}(0)$, while $\tilde{c}^{\ddagger}$ represents this fraction at the steady state, i.e., $\tilde{c}^{\ddagger} = \lim_{t\to+\infty} c^{\ddagger}(t)$. $t_{\theta}^{\ddagger}$ is determined to satisfy the following equation:

$$c^{\ddagger}(t_{\theta}^{\ddagger}) = \acute{c}^{\ddagger} + \frac{(\tilde{c}^{\ddagger} - \acute{c}^{\ddagger})\theta}{100} = \frac{\tilde{c}^{\ddagger}\theta + (100-\theta)\acute{c}^{\ddagger}}{100}.$$

Therefore, θ determines the time period during which we measure the "distance" ε between non-Markovian and the Markovian transient states. The value of θ in Fig. 2j is selected as 50 (see Supplementary Figs. S4, S5 and Supplementary Note 10 for more selection of θ and detailed analysis).

The forecasting error ε^{+} that evaluates whether a theory overestimates or underestimates the steady-state cumulative infected fraction is defined as:

$$\varepsilon^{+} = \frac{c^{\diamond}(\tilde{t}^{\star}) - c^{\star}(\tilde{t}^{\star})}{c^{\star}(\tilde{t}^{\star})},$$

where $\tilde{t}^{\star}$ denotes the time when the Monte Carlo stochastic simulation reaches the steady state when no infection occurs in the population, $c^{\diamond}(\tilde{t}^{\star})$ and $c^{\star}(\tilde{t}^{\star})$ are the cumulative infected fractions from theory and simulation, respectively. A positive value of ε^{+} indicates overestimation, whereas a negative value indicates underestimation.

The prevention evaluation error ε^{*}, which gauges the ability of the Markovian and non-Markovian theories to estimate the total effectiveness of vaccination, is defined as:

$$\varepsilon^{*} = \frac{\| \delta^{\star} - \delta^{\diamond}\|_2}{\| \delta^{\star}\|_2},$$

where $\delta^{\star}$ is the result from Monte Carlo simulation, $\delta^{\diamond}$ represents the result from theoretical calculation and $\|\cdot\|_2$ is the 2-norm of a vector.

The optimization failure probability $\hat{\varepsilon}$, which measures the probability that a theory fails to identify the optimal vaccination strategy, is defined as:

$$\hat{\varepsilon} = \frac{\sum_{l=1}^{z} \xi_{(l)}}{z},$$

where

$$\xi_{(l)} = \begin{cases} 0 & \text{if } \operatorname{argmin} \delta_{(l)}^{\star} = \operatorname{argmin} \delta_{(l)}^{\diamond} \\ 1 & \text{otherwise} \end{cases},$$

$\delta_{(l)}^{\star}$ and $\delta_{(l)}^{\diamond}$ represent the vectors, $\delta^{\star}$ and $\delta^{\diamond}$, for the l-th experiment, respectively, and z denotes the total number of experiments (in this paper, z is set to 100). Consequently, $\hat{\varepsilon}$ quantifies the fraction of experiments in which a theory fails to identify the optimal vaccination strategy, and serves as a measure of the probability of failure in optimizing the vaccination strategy.

Selection of training data

In our study, we selected the curves of all states that occurred prior to the time point at which the cumulative infected fraction reached a specific percentile ζ situated between the initial and steady-state cumulative infected fractions, as the training data. For example, let $\acute{c}^{\star}$ denote the cumulative infected fraction at time 0, while $\tilde{c}^{\star}$ represents this fraction at the steady state. The selection of the maximum time point of the time duration $\mathbb{T}_{\zeta}^{\star}$ for training data, denoted as $t_{\zeta}^{\star}$, is determined to meet the following equation:

$$c^{\star}(t_{\zeta}^{\star}) = \acute{c}^{\star} + \frac{(\tilde{c}^{\star} - \acute{c}^{\star})\zeta}{100} = \frac{\tilde{c}^{\star}\zeta + (100-\zeta)\acute{c}^{\star}}{100}.$$

Therefore, ζ determines the time duration, i.e., $\mathbb{T}_{\zeta}^{\star} = [0, t_{\zeta}^{\star}]$, of the training data, which could vary over different Monte Carlo simulations. In our study, we consistently choose ζ to be 20. Note that choosing a specific constant time period as the training data may not be appropriate, as it can result in an overabundance of data points for fitting due to some instances of fast transmission already having reached the steady state, while some instances of slow transmission may not have spread out yet, leading to invalid training data.

Fitting method

Because the removal process is independent of the infection one, we divide the fitting method into two parts: removal parameter fitting and infection parameter fitting. Specifically, we use $c^\star(t)$ and $r^\star(t)$ to denote the cumulative infected and removed fractions of a Monte Carlo simulation at time t. These two types of data are substituted into the Eq. (3) to fit the parameters of $\psi_{\rm rem}(\tau)$. To explain further, we can define a loss function:

$$L_{\rm rem} = \int_0^{t_\zeta^\star} [\hat{r}^\star(t) - r^\star(t)]^2 dt,$$

where $t_\zeta^\star$ represents the maximum time point of the training data, $\hat{r}^\star(t)$ denotes the removed fraction calculated by cumulative infected fraction, i.e.,

$$\hat{r}^\star(t) = \acute{r}^\star + \int_0^t [1 - \Psi_{\rm rem}(t - t')] dc^\star(t').$$

In this equation, $\Psi_{\rm rem}(\tau)$ is the survival function of a specific removal time distribution $\psi_{\rm rem}(\tau)$, such as Weibull, log-normal, or gamma distributions for the non-Markovian framework, and exponential distribution for the Markovian framework. By minimizing the loss function $L_{\rm rem}$, we can determine the optimal parameters for removal time distribution $\psi_{\rm rem}(\tau)$. This can be accomplished using the L-BFGS-B optimization algorithm, which is well-suited for minimizing the loss function[70].

Likewise, we use $c_l^\star(t)$ to denotes the cumulative infected fraction of age group l of a Monte Carlo simulation at time t and $s_l^\star(t) = 1 - c_l^\star(t)$ represents the corresponding susceptible fraction (note that cumulative infected fraction is precisely defined as the sum of the infected and removed fractions, which is accurately expressed as $c_l^\star(t) = i_l^\star(t) + r_l^\star(t) = 1 - s_l^\star(t)$, leading to $s_l^\star(t) = 1 - c_l^\star(t)$). After obtaining the removal parameters, these two types of data are put into Eq. (1) to fit the infection time distribution parameters. In details, we define a loss function $L_{\rm inf}$, which is given by:

$$L_{\rm inf} = \int_0^{t_\zeta^\star} \left\{ \sum_{l=1}^{n} p_l [\hat{c}_l^\star(t) - c_l^\star(t)] \right\}^2 dt,$$

where $\hat{c}_l^\star(t)$ represents the cumulative infected fraction of age group l at time t calculated by the cumulative infected fractions of simulation data by using the equation:

$$\hat{c}_l^\star(t) = \acute{c}_l^\star + \int_0^t k[1 - c_l^\star(t')] \sum_{m=1}^{n} A_{lm} p_m \int_0^{t'} \omega_{\rm inf}(t' - t'') \Psi_{\rm rem}(t' - t'') dc_m^\star(t'') dt'.$$

In this equation, $\omega_{\rm inf}(\tau)$ is the hazard function of a specific infection time distribution $\psi_{\rm inf}(\tau)$, such as Weibull, log-normal, or gamma distributions for the non-Markovian framework, and exponential distribution for the Markovian framework. And $\Psi_{\rm rem}(\tau)$ is the survival function of the removal time distribution $\psi_{\rm rem}(\tau)$, which is already obtained by fitting. To determine the optimal parameters for the infection time distribution $\psi_{\rm inf}(\tau)$, we minimize the loss function $L_{\rm inf}$ using the L-BFGS-B optimization algorithm[70].

Vaccination method

We assume that the individuals will build enough immune protection from the disease κ days after vaccination with the probability ρ, where ρ is the vaccine efficacy. In Monte Carlo simulations, if a susceptible individual gets vaccinated at time $t_{\rm vac}$, he/she will be marked with the probability ρ. When the absolute time reaches $t_{\rm vac} + \kappa$, if this individual has not been infected, he/she will be set to a state called protected state, indicating that this individual is protected from the disease and will never be infected.

In theoretical calculation, when a fraction of individuals in age group l get vaccinated with the detailed vaccination fraction υ_l, vaccination time $t_{\rm vac}$ and the fraction of susceptible individuals, which have not been vaccinated by time t, is recorded (denoted as $s_{l,*}(t_{\rm vac})$). When the absolute time reaches $t_{\rm vac} + \kappa$, the corresponding value of $s_l(t_{\rm vac} + \kappa)$ will be set as $s_l(t_{\rm vac} + \kappa)[1 - \frac{\rho \upsilon_l}{s_{l,*}(t_{\rm vac})}]$.

Data availability

All relevant data are available at https://github.com/fengmi9312/Validity-of-Markovian-for-Memory/tree/main/FigureData.

Code availability

The web-based application can be visited at https://cns.hkbu.edu.hk/toolbox/Validity-of-Markovian-for-Memory/main.html. The GitHub repository, which includes the source code for all the figure results, the web-based application, and an additional Python application, can be accessed at https://github.com/fengmi9312/Validity-of-Markovian-for-Memory.git.

Received: 4 October 2023; Accepted: 23 February 2024;
Published online: 08 March 2024

References

1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
2. Ferretti, L. et al. Quantifying SARS-COV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
3. Giordano, G. et al. Modeling vaccination rollouts, SARS-COV-2 variants and the requirement for non-pharmaceutical interventions in Italy. *Nat. Med.* **27**, 993–998 (2021).
4. Jentsch, P. C., Anand, M. & Bauch, C. T. Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *Lancet Infect. Dis.* **21**, 1097–1106 (2021).
5. Bubar, K. M. et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science* **371**, 916–921 (2021).
6. Buckner, J. H., Chowell, G. & Springborn, M. R. Dynamic prioritization of COVID-19 vaccines when social distancing is limited for essential workers. *Proc. Natl Acad. Sci.* **118**, e2025786118 (2021).
7. Goldstein, J. R., Cassidy, T. & Wachter, K. W. Vaccinating the oldest against COVID-19 saves both the most lives and most years of life. *Proc. Natl Acad. Sci.* **118**, e2026322118 (2021).
8. Viana, J. et al. Controlling the pandemic during the sars-cov-2 vaccination rollout. *Nat. Commun.* **12**, 1–15 (2021).
9. Matrajt, L., Eaton, J., Leung, T. & Brown, E. R. Vaccine optimization for COVID-19: who to vaccinate first? *Sci. Adv.* **7**, eabf1374 (2021).
10. Matrajt, L. et al. Optimizing vaccine allocation for COVID-19 vaccines shows the potential role of single-dose vaccination. *Nat. Commun.* **12**, 1–18 (2021).
11. Zhao, S. et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-NCOV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* **92**, 214–217 (2020).
12. Metcalf, C. J. E. & Lessler, J. Opportunities and challenges in modeling emerging infectious diseases. *Science* **357**, 149–152 (2017).
13. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–1301 (2007).
14. Levin, S. A., Grenfell, B., Hastings, A. & Perelson, A. S. Mathematical and computational challenges in population biology and ecosystems science. *Science* **275**, 334–343 (1997).
15. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
16. Lambiotte, R., Tabourier, L. & Delvenne, J.-C. Burstiness and spreading on temporal networks. *Eur. Phys. J. B* **86**, 1–4 (2013).

17. Lin, Z.-H. et al. Non-markovian recovery makes complex networks more resilient against large-scale failures. *Nat. Commun.* **11**, 1–10 (2020).
18. Cator, E., Van de Bovenkamp, R. & Van Mieghem, P. Susceptible-infected-susceptible epidemics on networks with general infection and cure times. *Phys. Rev. E* **87**, 062816 (2013).
19. Boguná, M., Lafuerza, L. F., Toral, R. & Serrano, M. Á. Simulating non-markovian stochastic processes. *Phys. Rev. E* **90**, 042108 (2014).
20. Van Mieghem, P. & Van de Bovenkamp, R. Non-markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Phys. Rev. Lett.* **110**, 108701 (2013).
21. Starnini, M., Gleeson, J. P. & Boguñá, M. Equivalence between non-markovian and Markovian dynamics in epidemic spreading processes. *Phys. Rev. Lett.* **118**, 128301 (2017).
22. Feng, M., Cai, S.-M., Tang, M. & Lai, Y.-C. Equivalence and its invalidation between non-markovian and Markovian spreading dynamics on complex networks. *Nat. Commun.* **10**, 1–10 (2019).
23. Min, B., Goh, K.-I. & Kim, I.-M. Suppression of epidemic outbreaks with heavy-tailed contact dynamics. *EPL* **103**, 50002 (2013).
24. Karrer, B. & Newman, M. E. Message passing approach for general epidemic models. *Phys. Rev. E* **82**, 016101 (2010).
25. Wang, Y., Chakrabarti, D., Wang, C. & Faloutsos, C. Epidemic spreading in real networks: an eigenvalue viewpoint. In *Proc. 22nd International Symposium on Reliable Distributed Systems,* 25–34 (IEEE, 2003).
26. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J. & Faloutsos, C. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.* **10**, 1–26 (2008).
27. Keeling, M. J. & Grenfell, B. T. Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65–67 (1997).
28. Barabasi, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
29. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 1–8 (2015).
30. Yan, X.-Y., Wang, W.-X., Gao, Z.-Y. & Lai, Y.-C. Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* **8**, 1–9 (2017).
31. Bratsun, D., Volfson, D., Tsimring, L. S. & Hasty, J. Delay-induced stochastic oscillations in gene regulation. *Proc. Natl Acad. Sci.* **102**, 14593–14598 (2005).
32. Vazquez, A., Racz, B., Lukacs, A. & Barabasi, A.-L. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.* **98**, 158702 (2007).
33. Moghadas, S. M. et al. The implications of silent transmission for the control of COVID-19 outbreaks. *Proc. Natl Acad. Sci.* **117**, 17513–17515 (2020).
34. Lauer, S. A. et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).
35. Gatto, M. et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci.* **117**, 10484–10491 (2020).
36. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-COV-2). *Science* **368**, 489–493 (2020).
37. Buitrago-Garcia, D. et al. Occurrence and transmission potential of asymptomatic and presymptomatic sars-cov-2 infections: a living systematic review and meta-analysis. *PLoS Med.* **17**, e1003346 (2020).
38. World Health Organization, et al. Consensus document on the epidemiology of severe acute respiratory syndrome (sars). Technical Report, (World Health Organization, 2003).
39. Chowell, G. et al. Model parameters and outbreak control for sars. *Emerg. Infect. Dis.* **10**, 1258 (2004).
40. Eichner, M. & Dietz, K. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemiol.* **158**, 110–117 (2003).
41. Rizoiu, M.-A., Mishra, S., Kong, Q., Carman, M. & Xie, L. Sir-Hawkes: linking epidemic models and Hawkes processes to model diffusions in finite populations. In *Proc. World Wide Web Conference*, 419–428 (2018).
42. Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D. & Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.* **5**, 4630 (2014).
43. Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* **6**, 477–487 (2008).
44. Ganyani, T. et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, march 2020. *Eurosurveillance* **25**, 2000257 (2020).
45. Park, S. W. et al. Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011548118 (2021).
46. Svensson, Å A note on generation times in epidemic models. *Math. Biosci.* **208**, 300–311 (2007).
47. Klinkenberg, D. & Nishiura, H. The correlation between infectivity and incubation period of measles, estimated from households with two cases. *J. Theor. Biol.* **284**, 52–60 (2011).
48. te Beest, D. E., Wallinga, J., Donker, T. & van Boven, M. Estimating the generation interval of influenza a (h1n1) in a range of social settings. *Epidemiology* **24**, 244–250 (2013).
49. Champredon, D., Dushoff, J. & Earn, D. J. Equivalence of the erlang-distributed seir epidemic model and the renewal equation. *SIAM J. Appl. Math.* **78**, 3258–3278 (2018).
50. Feng, M., Tian, L., Lai, Y.-C. & Zhou, C. Web-based application for generation time distribution and error rectification of Markovian modeling. https://cns.hkbu.edu.hk/toolbox/Validity-of-Markovian-for-Memory/main.html (2023).
51. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019 https://population.un.org/wpp/. Accessed on 29 June 2021 (2021).
52. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**, e1005697 (2017).
53. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604 (2007).
54. Dietz, K. The estimation of the basic reproduction number for infectious diseases. *Stat. Methods Med. Res.* **2**, 23–41 (1993).
55. Castro, M., Ares, S., Cuesta, J. A. & Manrubia, S. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proc. Natl Acad. Sci.* **117**, 26190–26196 (2020).
56. Case, B., Young, J.-G. & Hébert-Dufresne, L. Accurately summarizing an outbreak using epidemiological models takes time. *arXiv preprint arXiv:2301.08799* (2023).
57. Britton, T. & Scalia Tomba, G. Estimation in emerging epidemics: biases and remedies. *J. R.Soc. Interface* **16**, 20180670 (2019).
58. Lai, Y.-C. & Tél, T.Transient Chaos—Complex Dynamics on Finite Time Scales (Springer, New York, 2011).
59. Hastings, A. & Higgins, K. Persistence of transients in spatially structured ecological models. *Science* **263**, 1133–1136 (1994).
60. Hastings, A. Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* **19**, 39–45 (2004).
61. Hastings, A. Timescales and the management of ecological systems. *Proc. Natl Acad. Sci.* **113**, 14568–14573 (2016).
62. Hastings, A. et al. Transient phenomena in ecology. *Science* **361**, eaat6412 (2018).
63. Morozov, A. et al. Long transients in ecology: theory and applications. *Phys. Life Rev.* **32**, 1–40 (2020).
64. Meng, Y., Lai, Y.-C. & Grebogi, C. Tipping point and noise-induced transients in ecological networks. *J. R. Soc. Interface* **17**, 20200645 (2020).

65. Sjöström, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
66. Martin, S. J., Grimwood, P. D. & Morris, R. G. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
67. Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
68. Hawkes, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90 (1971).
69. Centro de coordinación de alertas y emergencias sanitarias. Actualización nº 103. Enfermedad por el coronavirus (COVID-19). 12.05.2020 (datos consolidados a las 21:00 horas del 11.05.2020) SITUACIÓN EN ESPAÑA. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_103_COVID-19.pdf. (Accessed on 24 Dec 2023).
70. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound-constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190–1208 (1995).

Acknowledgements

This work was partially supported by the Hong Kong Baptist University (HKBU) Strategic Development Fund; the Research Grants Council of Hong Kong (Grant No. C2005-22Y), the National Natural Science Foundation of China (Grant No. 12275229), and the Hong Kong Chinese Medicine Development Fund (Grant No. 22B2/049A) to L.T.; the Research Grants Council of Hong Kong (Grant No. GRF12201421) to C.-S.Z. This research was conducted using the resources of the High-Performance Computing Cluster Centre at HKBU, which receives funding from the Hong Kong Research Grant Council and the HKBU. Y.-C.L. was supported by the Office of Naval Research through Grant No. N00014-21-1-2323.

Author contributions

M.F., L.T. and C.-S.Z. designed research; M.F. performed research; L.T. and C.-S.Z. contributed analytic tools; M.F., L.T. and C.-S.Z. analyzed data; M.F., L.T., Y.-C.L. and C.-S.Z. discussed the results and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42005-024-01578-w.

Correspondence and requests for materials should be addressed to Liang Tian or Changsong Zhou.

Peer review information *Communications Physics* thanks David Soriano-Paños and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.