

RESEARCH ARTICLE

Early-warning signals for infectious diseases with a social-media compartment

Francisca Olajide¹, Frithjof Lutscher², Stacey R. Smith?^{3*}

1 Department of Mathematics and Statistics, The University of Ottawa, Ottawa, Canada, **2** Department of Mathematics and Statistics, and Department of Biology, The University of Ottawa, Ottawa, Canada, **3** Department of Mathematics and Faculty of Medicine, The University of Ottawa, Ottawa, Canada

* stacey.smith@uottawa.ca



OPEN ACCESS

Citation: Olajide F, Lutscher F, Smith? SR (2026) Early-warning signals for infectious diseases with a social-media compartment. PLoS One 21(7): e0354091. <https://doi.org/10.1371/journal.pone.0354091>

Editor: Md. Kamrujjaman, University of Dhaka, BANGLADESH

Received: July 17, 2025

Accepted: June 30, 2026

Published: July 28, 2026

Copyright: © 2026 Olajide et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Our manuscript does not use data.

Funding: National Sciences and Research Engineering Council of Canada.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Early-warning signals (EWSs) are crucial tools for anticipating disease emergence and guiding public-health responses, but uncertainties in transmission and incomplete data can limit their reliability. Additionally, the performance of EWSs is rarely evaluated when disease emergence is delayed, and their use in the context of interactions between disease transmission and communication through social-media platforms has largely not been considered. We evaluate the relative performance of EWSs in predicting disease emergence under varying noise conditions and explore the use of EWSs with social-media dynamics to predict disease emergence. We develop a mechanistic model coupling infectious disease and social-media dynamics, introduce stochasticity and generate simulated time series. We detect changepoints, quantify delays relative to the bifurcation point and assess the performance of EWSs across different segments of the time series. The “reporting” infected compartment proves most reliable, and variance outperforms autocorrelation in high-noise scenarios. However, in the social-media compartment, variance and autocorrelation have weak predictive power. Our work provides a framework to advance understanding of how EWSs can be applied to forecasting disease emergence, contributing to improved disease preparedness and response.

Introduction

Early detection of infectious diseases can provide precious time to implement effective interventions [1,2], which can mitigate the often-disastrous consequences of such diseases on various aspects of human existence [3,4]. The emergence of the SARS-CoV-2 pathogen exemplifies this impact, resulting in millions of deaths, strain on healthcare resources and far-reaching economic and social consequences, including business closures, job losses and significant disruptions to daily life [3,5]. Our work proposes and evaluates the relative efficiency of some early detection methods.

Three factors make the early detection of an emerging disease particularly difficult. (1) As with any real-world process, noise is inherent in the process of disease emergence and can hide the signal of early disease transmission when case numbers are low. (2) Data on disease emergence are often incomplete, since many infectious individuals are either unaware of their infection or do not report it. (3) Mathematical models often show a delay between the theoretical prediction and the actual observation of an emerging disease. Our work helps to quantify this delay in view of different forms of noise and reporting patterns.

The emergence of an infectious disease is often viewed as a critical transition from the absence of the disease to its presence. In mathematical models, formulated as deterministic dynamical systems, such transitions correspond to bifurcations [6,7]. Typically, the stable disease-free equilibrium becomes unstable and a stable endemic equilibrium emerges in a transcritical bifurcation, which occurs when the basic reproduction number is equal to one [4,8]. Such phenomena are widely seen in numerous models in mathematical biology [9]. The basic reproduction number (R_0) is the average number of secondary infections caused by a single infectious individual in a completely susceptible population [10].

In stochastic dynamical systems, the theory of critical slowing can identify approaching critical transitions [7,11–13]. It suggests that a system gradually loses its ability to recover from perturbations as it approaches a critical transition [12,14]. Simultaneously, statistical patterns of solutions can change [15]. For example, critical slowing down leads to a decrease in the rate of change of the system, causing the current state of the system to become more correlated with its past state, thereby increasing lag-1 autocorrelation [2,7,11,16]. It also leads to increasing variance in solutions [11,13,15–17]. These and other statistical measures used to anticipate critical transitions are generally referred to as early warning signals (EWSs); they are generically applicable for identifying critical transitions [2,11,13,15–19]. Previous studies have shown that EWSs are sometimes useful in providing early indicators of disease emergence, re-emergence and elimination [2,18,20–23].

In systems where parameter values change over time, it is tempting to think that a critical transition occurs when the parameter(s) pass through the bifurcation point, but this is not necessarily true. Instead, critical transitions are typically delayed [24,25]. In the context of disease emergence, when R_0 gradually increases and crosses the critical threshold of one, the disease will not emerge immediately but rather remain nearly undetectable for some time before undergoing a delayed but quick transition to the endemic state [21,26,27]. The phenomenon that system solutions stay near an unstable state before the transition occurs has been termed “tracking unstable states” [28]. Since critical transitions can be delayed and decoupled from bifurcation points, basing their predictions on bifurcations can lead to inaccurate outbreak risk assessment and poorly timed interventions.

Critical transitions manifest in time-series data as abrupt shifts, such as a sudden rise (emergence) or a sudden decline (elimination) in the number of infected individuals. They can be found using changepoint detection methods, which indicate whether and when significant shifts in the statistical properties of the data (e.g., mean, trend

or variance) occur [29]. EWSs have been used to predict bifurcation points in time-series data [7,11,15]. However, when transitions are delayed, it is not clear whether EWSs can be used to predict changepoints [30]. Our work helps assess the performance of EWSs in detecting changepoints and provides a better understanding of how EWSs can be applied to epidemic forecasting.

Noise is an inherent feature of real-world processes that makes detection of critical transitions difficult. It can result from intrinsic fluctuations (demographic stochasticity) or from extrinsic forcing (environmental stochasticity) and can affect the stability of a system and the timing of critical transitions. Several studies have found that noise may increase the bifurcation delay [11,31], whereas others suggest that it may induce a transition before the bifurcation point [31,32]. Consequently, the performance of EWSs in predicting transitions can vary considerably depending on the nature and intensity of stochastic fluctuations. O'Regan and Burton [33] found that some EWSs, such as lag-1 autocorrelation and coefficient of variation, are robust to noise, whereas variance is more sensitive to the form of noise in a simple, one-dimensional system. Dutta et al. [34] found that while EWSs perform reliably in predicting saddle-node bifurcations, their performance is more sensitive to noise type when applied to other transitions.

Availability and quality of data is another factor potentially limiting the ability of EWSs to predict critical transitions. While EWSs can predict critical transitions based on disease prevalence in simple disease models [2,13,18,20,21], empirical data is often affected by inconsistent data collection, lags in case reporting, testing frequencies and underreporting [13]. EWSs generally require high-frequency data from longer time series for improved performance [35], but such high-quality data is often not available [13,32]. These limitations emphasize the need for additional data sources and proxies [32]. Previous studies have explored the use of Google searches (searches on symptoms and health-related issues) and activities (reported symptoms) from social-media platforms (Twitter, Facebook, et cetera) to enhance early warning of disease outbreaks [36]. These studies used techniques such as sentiment analysis and natural language processing. They found that Google searches and social-media data strongly correlate with disease incidence and, in some cases, precede public-health reports [36–40]. This raises the prospect of using EWSs to assess the potential of social-media dynamics in predicting disease emergence.

Here, we develop a framework that integrates delays in disease emergence, incomplete reporting and random noise, together with alternative data sources such as social media dynamics, to evaluate EWSs of infectious diseases. The objectives of our study are:

- (i) Exploring the effectiveness of EWSs in predicting disease emergence in the presence of interactions between delays, noise and incomplete information.
- (ii) Assessing the potential of using EWSs with partial information about infections and with social-media dynamics to predict disease emergence under different noise scenarios.

Model derivation

The deterministic SWIRM model

We extend the classic SIR model by splitting infected individuals into two compartments and by introducing a social-media compartment (see Fig 1). We assume that not all infected individuals enter the healthcare system, and only the information from those who do is available to decision-makers. Based on this assumption, we divide infected individuals into two classes: W represents infected individuals who do not report to the doctor (for example because they experience only mild symptoms, although this class may include individuals with severe symptoms who do not report), and I represents infected individuals who do report to the doctor (for example because they have severe symptoms). Furthermore, S represents susceptible individuals, R represents recovered individuals and M represents social-media posts about the disease (see below).

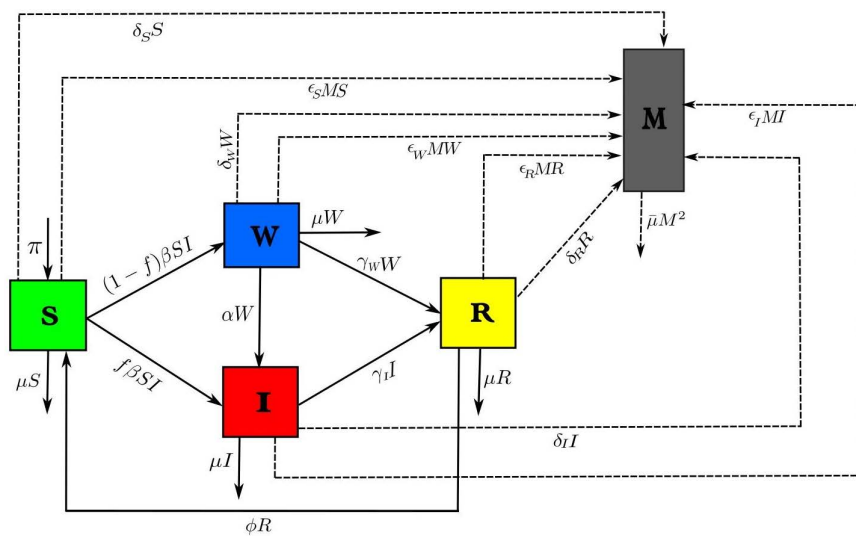


Fig 1. Flow diagram of the SWIRM model. Solid arrows represent transitions within the human population, while dashed arrows represent interactions with the media compartment.

<https://doi.org/10.1371/journal.pone.0354091.g001>

Individuals are recruited into the susceptible compartment at a constant rate π ; recovered individuals may lose their immunity at rate ϕ ; susceptible individuals may become infected with transmission rate β (this rate will later depend on time). Unreported infected individuals may progress to the reported infected class at rate α ; infected individuals in the W and I compartments recover at rates γ_W and γ_I , respectively. The natural death rate μ is assumed to be the same for all the sub-populations.

Within the context of our study, the social-media compartment M is modelled as an observable proxy for disease dynamics rather than a mechanistic driver of behaviour or transmission. We assume that individuals in the S , W , I and R compartments generate posts at rates δ_S , δ_W , δ_I and δ_R , and interact with existing content at rates ϵ_S , ϵ_W , ϵ_I and ϵ_R through engagement metrics such as likes, shares, retweets, etc. We note that this assumption is a simplification and does not fully account for the variability observed in real-world social networks. We denote the rate at which social-media posts become less visible to users as $\bar{\mu}$. Parameter descriptions, units and values are given in [Table 1](#).

The dynamics of the SWIRM model are then given by the system of equations

$$\begin{aligned}\frac{dS}{dt} &= \pi - \beta SI - \mu S + \phi R, \\ \frac{dW}{dt} &= (1-f)\beta SI - \gamma_W W - \alpha W - \mu W, \\ \frac{dI}{dt} &= f\beta SI + \alpha W - \gamma_I I - \mu I, \\ \frac{dR}{dt} &= \gamma_W W + \gamma_I I - \mu R - \phi R, \\ \frac{dM}{dt} &= \delta_S S + \delta_W W + \delta_I I + \delta_R R + \epsilon_S MS + \epsilon_W MW + \epsilon_I MI + \epsilon_R MR - \bar{\mu} M^2.\end{aligned}\tag{1}$$

The disease-free equilibrium of (1) is given by

Table 1. Parameters, description, units and values. $\hat{I}$ represents individuals, $\hat{T}$ represents time (days), and $\hat{P}$ represents posts. $\delta_S, \delta_W, \delta_I, \delta_R$ are proportions of δ and $\epsilon_S, \epsilon_W, \epsilon_I, \epsilon_R$ are proportions of ϵ . The symbol (*) means the parameters were adjusted within a range of 30–86% relative to values used in previous studies. The numerical parameter r was calculated such that $R_0(t)$ reaches the threshold value of 1 at a reasonable time point (around 14 days); π was calculated so that there are 1000 people at the disease-free equilibrium; ϵ was set to 50% of δ to reflect the tendency of users to passively consume health-related content. We chose a relatively high fraction for infected individuals who report; while this may be lower for diseases like COVID-19, it is plausible for mandatory reportable infections such as measles.

	Parameters	Description	Units	Values	Source
Demographic parameters	π	Birth rate	$\hat{I}\hat{T}^{-1}$	14.28	Calculated
	μ	Natural death rate	$\hat{T}^{-1}$	0.01428	[41]*
Disease-related parameters	f	Fraction of infected individuals who report	unitless	0.8	Assumed
	$\beta = \beta(t)$	Transmission rate	$(\hat{I}\hat{T})^{-1}$	$\beta_0 + rt$	
	γ_W	Recovery rate of unreported infected individuals	$\hat{T}^{-1}$	0.07	[42]*
	γ_I	Recovery rate of infected individuals who report	$\hat{T}^{-1}$	0.024	Assumed
	ϕ	Rate of waning immunity	$\hat{T}^{-1}$	0.01	[43]
	α	Transition rate from W to I	$\hat{T}^{-1}$	0.033	[44]*
Media parameters	ϵ	Rate at which individuals engage or interact with already posted social media content (likes, shares, retweets, etc.)	$(\hat{I}\hat{T})^{-1}$	0.25	Calculated
	δ	Rate at which individuals post about the disease and self-report symptoms	$\hat{P}(\hat{I}\hat{T})^{-1}$	0.5	[44]
	$\bar{\mu}$	Fickle rate of social-media posts	$(\hat{P}\hat{T})^{-1}$	0.1428	[44]*
Numerical parameters	β_0	Initial transmission rate	$(\hat{I}\hat{T})^{-1}$	0.000005	[44]*
	r	Rate of change of $\beta(t)$		0.000002739	Calculated

<https://doi.org/10.1371/journal.pone.0354091.t001>

$$(\bar{S}, \bar{W}, \bar{I}, \bar{R}, \bar{M}) = \left(\frac{\pi}{\mu}, 0, 0, 0, \frac{\epsilon_S \frac{\pi}{\mu} + \sqrt{\left(\epsilon_S \frac{\pi}{\mu}\right)^2 + \frac{4\bar{\mu}\delta_S\pi}{\mu}}}{2\bar{\mu}} \right).$$

We observe that even though $\bar{W} = \bar{I} = 0$, we have $\bar{M} \neq 0$, which means that there are social-media posts about the disease even when there is no disease. We consider this relevant in the context where the disease is already in a different population than the one we are considering, such as in a different city or country.

We linearize the system around the disease-free equilibrium to analyze its local stability and determine the basic reproduction number; for details, see the Appendix. The basic reproduction number for our model is given by

$$R_0 = \frac{\alpha(1-f)\beta\frac{\pi}{\mu} + f\beta\frac{\pi}{\mu}(\gamma_W + \alpha + \mu)}{(\gamma_W + \alpha + \mu)(\gamma_I + \mu)}. \quad (2)$$

Although a social-media compartment is included in the SWIRM model, it does not appear in the expression for the basic reproduction number, because it is a higher-order term and hence not present in the linearisation used to derive R_0 . It thus serves as an observable variable.

At $R_0 = 1$, system (1) undergoes a transcritical bifurcation, and the critical transmission rate is

$$\beta = \frac{(\gamma_W + \alpha + \mu)(\gamma_I + \mu)}{\frac{\pi}{\mu}(\alpha(1-f) + f(\gamma_W + \alpha + \mu))}. \quad (3)$$

In real systems, parameters such as transmission rate change slowly over time [24,25,30]. To allow the transmission rate to change over time, we write β as $\beta(t)$. Substituting this function into Eq (2), we define a time-dependent reproduction number $R_0(t)$. We then obtain a bifurcation point in time by calculating t such that $R_0(t)=1$.

While the function $\beta(t)$ may take different functional forms, we adopt a linearly increasing function ($\beta(t) = \beta_0 + rt$) as a simple and interpretable forcing scenario that allows the relationship between threshold crossing and the observed transition dynamics to be examined precisely. Here, $\beta_0 > 0$ is the initial transmission rate, and r ($0 < r \ll 1$) is the rate of change in β_0 . With this choice and the parameter values in Table 1, the bifurcation point is $t = 14.5$ days.

Models with noise

We use model (1) as the deterministic skeleton to which we add three different forms of stochasticity: additive white noise, multiplicative white noise, and demographic stochasticity. From each of these models, we then generate time-series data that we analyze for early warning signals.

Additive white noise represents random fluctuations that do not mechanistically influence the deterministic components [45]; they are independent of population size. To include additive white noise, we add the increment of a Wiener process (denoted by $\mathcal{W}(t)$) with a certain intensity to each of the equations in (1). For simplicity, we choose the same intensity (denoted by σ) for all equations.

Unlike additive white noise, which is independent of population size, multiplicative white noise models environmental fluctuations whose magnitude scales with population size [16,46]. As above, we add the increment of a Wiener process, but this time, we choose the intensity to be proportional to the population size in the respective compartments. For simplicity, we assume that the proportionality constant is the same for all compartments (again denoted by σ).

Demographic stochasticity refers to the random variation in population size that results from births, deaths, immigration and other population processes. We follow the approach by Greenwood and Gordillo [47,48]. Unlike the other forms of stochasticity, the intensity of demographic stochasticity is determined by the rates of the deterministic model (1).

We note that while there is a “free” parameter that controls the intensity of additive and multiplicative noise, no such parameter exists in the case of demographic stochasticity. Hence, we can (and will) explore the effect of noise intensity on early-warning signals in the first two cases but not in the third.

From each of the three stochastic models described here, we generate 300 time series, using the Euler–Maruyama method [49]. This method is a simple variant of the classical forward Euler method in which the increment of a Wiener process is drawn from a standard normal distribution and multiplied by the size of the timestep of the method. We use a small timestep ($dt=0.005$) to reduce discretization error and ensure numerical stability. To maintain biological feasibility, we correct any non-physical values arising from the discretization by reverting to the previous state. For completeness, we present all the equations for all three forms of stochasticity explicitly in the Appendix.

Methods

Next, we describe the methods used to generate and analyze time-series data for changepoints and early-warning signals, as well as the visualization techniques employed to present the results.

Changepoint analysis

We analyze each of the time series using changepoint analysis. Specifically, we search for changepoints in the time-series data by using the mean changepoint detection (`cpt.mean`) function of the “changepoint” package in R [29,50]. Using the At

Most One Changepoint (AMOC) method, the null hypothesis (H_0) is that there are no changepoints in the data, while the alternative hypothesis (H_1) is that there is exactly one changepoint in the data. The single changepoint is estimated using the “Asymptotic” penalty with `pen.value=0.05`, which corresponds to a theoretical type I error of 0.05. We tested some of our results with different values of parameter `pen.value` and found no difference in outcome.

The `cpt.mean` function relies on assumptions of independence of observations within each segment, normally distributed errors and abrupt, rather than gradual, changes in the mean. In the simulated time-series data, these assumptions are generally satisfied, with the exception of independence within segments, which is violated due to autocorrelation. Based on the objectives of our study, identifying changepoints using `cpt.mean` represents one component of the overall pipeline, rather than the primary focus. Alternative packages, such as “`envcpt`” in R, account for autocorrelation and may provide more reliable changepoint detection in time-series data where the assumption of independence is violated. However, changepoint detection with “`envcpt`” is sensitive to the choice of minimum segment length, which can affect both the number and the location of detected changepoints. Some limitations of the changepoint package are that it may fail to detect changepoints when the underlying process changes gradually, and high noise levels can lead to spurious detections or obscure true changepoints.

We check whether the changepoint (CP) in the time series occurs before or after the bifurcation point (BP). If the CP occurs after the BP, we define the scenario as a delay (Fig 2(a)); conversely if the CP precedes the BP, we define the scenario as an advance (Fig 2(b)).

We note that the BP is obtained analytically; therefore, any sensitivity to timestep size is expected to arise through the estimation of changepoints and the corresponding delay values. To clarify this, we carry out additional simulations for the lowest noise intensity case under additive white noise using timesteps $dt=0.0025$ and $dt=0.01$ in the Appendix.

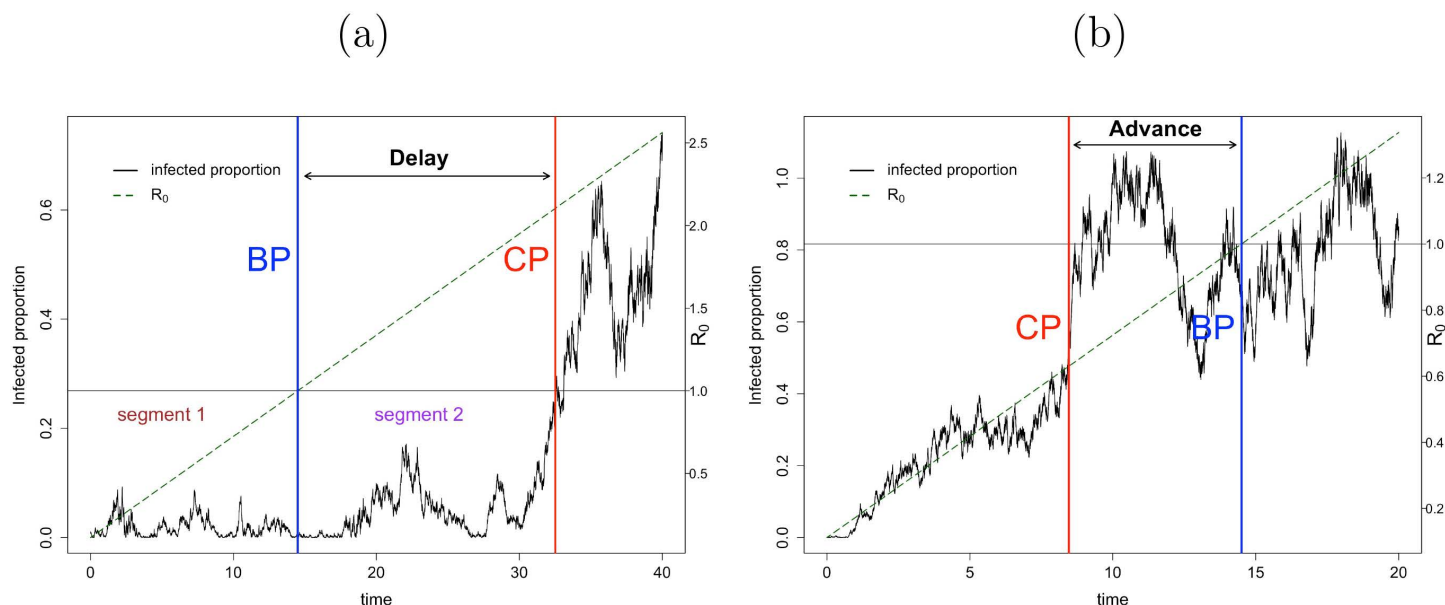


Fig 2. (a) Illustration of a “delay” based on a realization of the infected population in the model with demographic stochasticity. The transmission parameter is slowly forced through the bifurcation point $R_0 = 1$ (the blue line). The red line representing the changepoint is computed using the `cpt.mean` function in R. The delay occurs between the system crossing the bifurcation point (i.e., $R_0 = 1$) and the subsequent spike in infections. Since the transmission rate varies continuously in time, crossing the threshold does not immediately produce a sharp increase in infections. (b) Illustration of an “advance” based on a realization of the infected population.

<https://doi.org/10.1371/journal.pone.0354091.g002>

Early-warning signals

We use the “earlywarnings” package in R to calculate EWSs for a given time series. More specifically, we calculate the autocorrelation and variance of a time series and compute the Kendall- τ correlation coefficient (ranging from -1 to $+1$) to detect trends in these quantities. Positive and negative Kendall- τ values represent an increasing and decreasing trend of each signal, respectively, and increasing trends of autocorrelation and variance are indicators of an upcoming transition [7,11,12,31].

Since we want to compare the performance of EWSs regarding the changepoint (CP) and the bifurcation point (BP), we select two segments from each time series that show the expected delay (i.e., CP succeeds BP; see previous section). The two segments are of equal length: the length of the time series from the beginning to the BP (2900 data points). In line with standard practice [11,51], we apply Gaussian detrending to each segment to eliminate long-term trends that may obscure short-term variations. Then we use a rolling window of 50% of the detrended time-series data. Since the transition occurs at the CP and not at the BP, the ideal EWS is negative for the time series leading up to the BP and positive for the time series leading up to the CP. If the Kendall- τ value is positive and there is no transition (BP), we say it is a false positive. If there is a transition (CP) and the Kendall- τ value is positive, we say it is a true positive. Similarly, a negative Kendall- τ value despite a transition (CP) is a false negative while a negative Kendall- τ value with no transition (BP) is a true negative.

We use raincloud plots to visualize the distribution of Kendall- τ values for autocorrelation and variance. Raincloud plots provide a robust method for data visualization, incorporating a density plot (“the cloud”), raw data observations (“the rain”) and a box and whisker plot [52].

To compare the performance of autocorrelation and variance in predicting CPs, we use receiver-operating characteristic (ROC) curves. An ROC curve plots the true positive rate against the false positive rate at various threshold settings. We use the area under the curve (AUC) to measure the performance of autocorrelation and variance in predicting the changepoint. A higher AUC indicates a better EWS. An AUC of 1 means that the EWS perfectly identifies the positives and negatives, while an AUC of 0.5 indicates that the EWS performs no better than random guessing.

Results

We begin by reporting whether there is a delay or an advance in each compartment; i.e., whether the changepoint (CP) is after or before the bifurcation point (BP). We also compare the changepoints between the different compartments; that is, whether the transitions in one compartment occur before or after the transitions in the other compartment. These results help us identify compartments that can be used to detect transitions early or predict other compartments.

Then we present our results regarding which EWSs (autocorrelation or variance) from which compartment (I , W or M) correctly predict the changepoint or incorrectly predict the bifurcation point. We reiterate that social-media posts do not influence transmission in our model, so all effects of M are observed through the EWSs. We present one example of results on these EWSs with raincloud plots and summarize the remaining results from each compartment for the different forms and intensities of noise using ROC curves and AUC values. These results will help us determine which EWSs have the best performance (high AUC values) and which compartment gives the strongest signal.

Changepoints — Additive white noise

In the model with additive white noise (see eq. (5)), almost all time series show the expected delays for all values of σ ; i.e., the detected CPs in all time series occur significantly after the parameter passed the BP. The only exception of an advance occurs in a time series of the M compartment when $\sigma = 3$. Different compartments experience different delays.

The detected CPs for W precede those for I across the different noise intensities. This may indicate that the W compartment responds more rapidly to changes in system dynamics before these changes become pronounced in the I compartment. The occurrence of the detected CPs for M , relative to those for I , depends on the intensity of noise. As noise

intensity increases, more and more CPs for M occur later than those for I , indicating a higher sensitivity of the M compartment to noise (see Fig 3). For example, with $\sigma = 0.5$, the CPs for M precede the CPs for I in all time series. When $\sigma = 1.5$, the CPs for M occur before the CPs for I in 78% of the time series. With $\sigma = 3$, only 42% of the time series show that the CPs for M occur before the CPs for I . Time series where the CPs for M precede those for I suggest a potential connection between social-media dynamics and infection trends.

The proximity of the CPs for M and W suggests a potential association between active social-media posting and the presence of mild infectious cases, with this pattern being more apparent under lower stochastic fluctuations. As σ increases, the CPs for M occur substantially later than those for W , suggesting that the association between the two variables becomes weaker at higher noise levels.

Changepoints — Multiplicative noise

In the model with multiplicative noise, time series for I and W show the expected delay for all values of σ . Time series for M show mostly delays, with advances occurring in no more than 2% of the time series across the different noise intensities.

The CPs for W generally precede those for I (see Fig 4). However, we observe a few instances where this is not the case, particularly for higher noise intensity (here $\sigma = 0.1$). As noise intensity increases, there are more instances where the CPs for M occur before the CPs for I — a result that contrasts with the observation from the model with additive white noise. This may be due to the state-dependent nature of the multiplicative noise.

Unlike the model with additive white noise, social-media dynamics show no association with mild infectious cases (Fig 4). Overall, these results suggest that the timing of changepoints across variables depends on the form and intensity of noise in the system.

Changepoints — Demographic Stochasticity

In the model with demographic stochasticity, time series show more advances than delays. Time series for I show 30% delays and 70% advances; time series for M show 28% delays and 72% advances; and time series for W show 32% delays and 68% advances. These advances may be due to the fact that demographic noise can lead to the amplification of early infections.

The occurrence of the CPs for the compartments varies. In some cases, the detected CPs for W and M occur close to each other while those for I occur much later, whereas in other instances, the CPs for I align with those for W or M (see

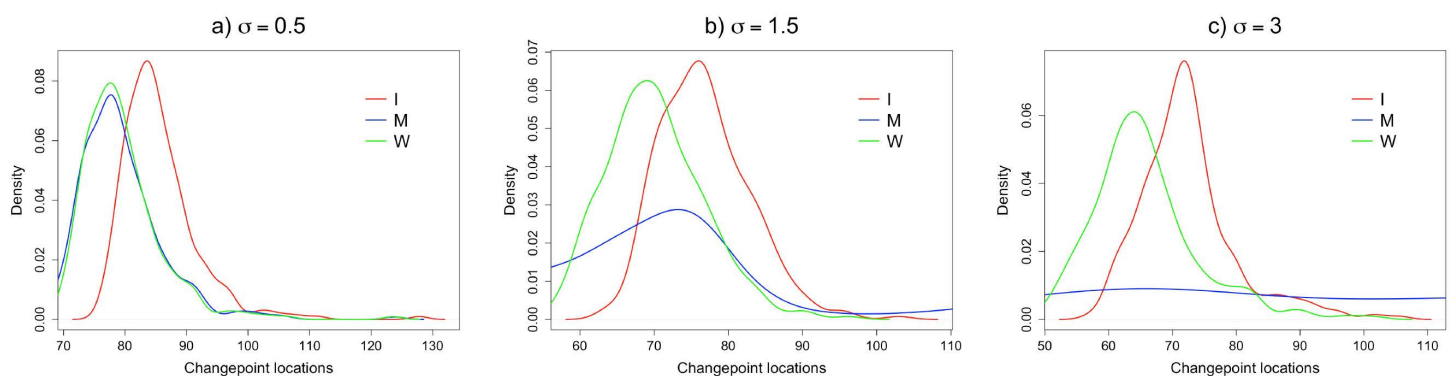


Fig 3. Density of changepoint locations detected for 300 time series generated from the model with additive white noise when (a) $\sigma = 0.5$, (b) $\sigma = 1.5$, (c) $\sigma = 3$. Other parameters are as in Table 1.

<https://doi.org/10.1371/journal.pone.0354091.g003>

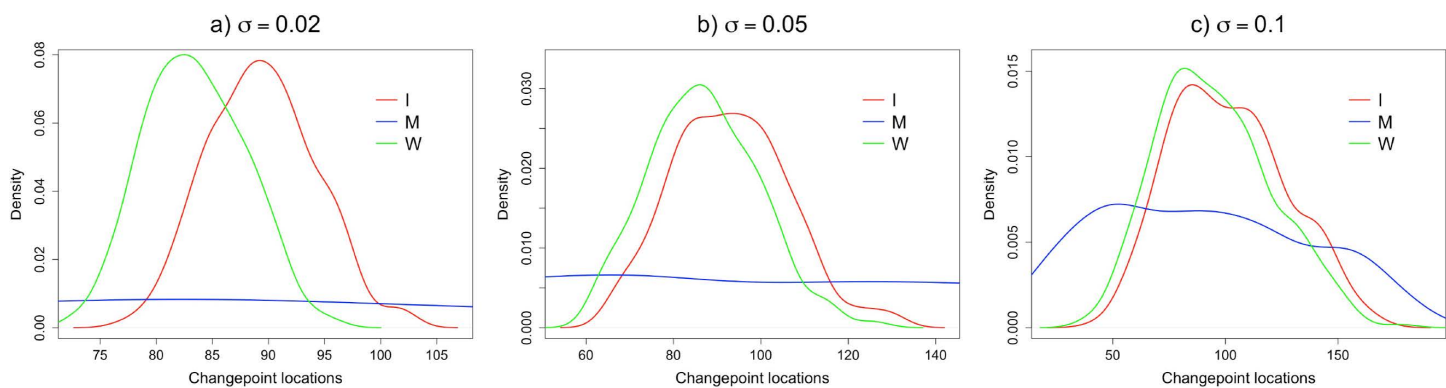


Fig 4. Density of changepoint locations detected for 300 time series generated from the model with multiplicative noise when (a) $\sigma = 0.02$, (b) $\sigma = 0.05$, (c) $\sigma = 0.1$.

<https://doi.org/10.1371/journal.pone.0354091.g004>

Fig 5). This is not unexpected, because the time series for the different compartments show advances and delays at different times.

EWSs — Raincloud plots

In the lowest noise-intensity case of the multiplicative noise, autocorrelation (lag-1) shows 48% and 62% positives for time series of I leading up to the bifurcation point (BP) and changepoint (CP), respectively. Hence, autocorrelation correctly predicts CP in 62% of cases (true positives) but fails to predict in 38% of cases (false positives) (**Fig 6**, left). Variance computed for time series leading up to the BP shows 26% positives for I . Since no transition is occurring (CP occurs after BP), these are false positives. Conversely, variance computed for time series leading up to the CP shows 100% positives, correctly predicting true transitions in the time series (**Fig 6**, right).

Variance computed on time series of W shows similar result as time series of I in detecting true positives. However, this also results in a high number of false positives. With autocorrelation, not all true positives are identified, but there are fewer false positives compared to time series of I . See **Table 2** for all numerical values. Autocorrelation and variance

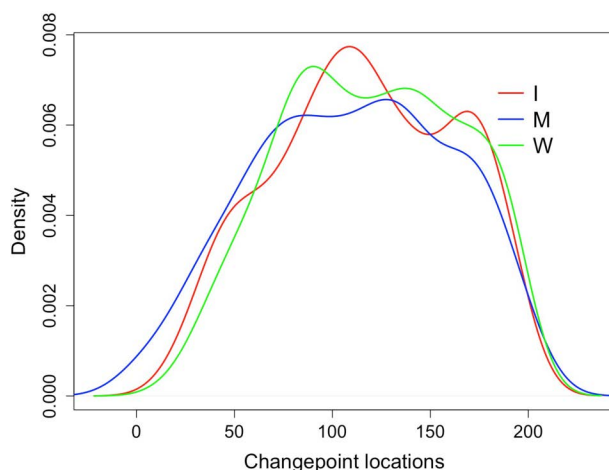


Fig 5. Density of changepoint locations detected for 300 time series generated from the model with demographic stochasticity.

<https://doi.org/10.1371/journal.pone.0354091.g005>

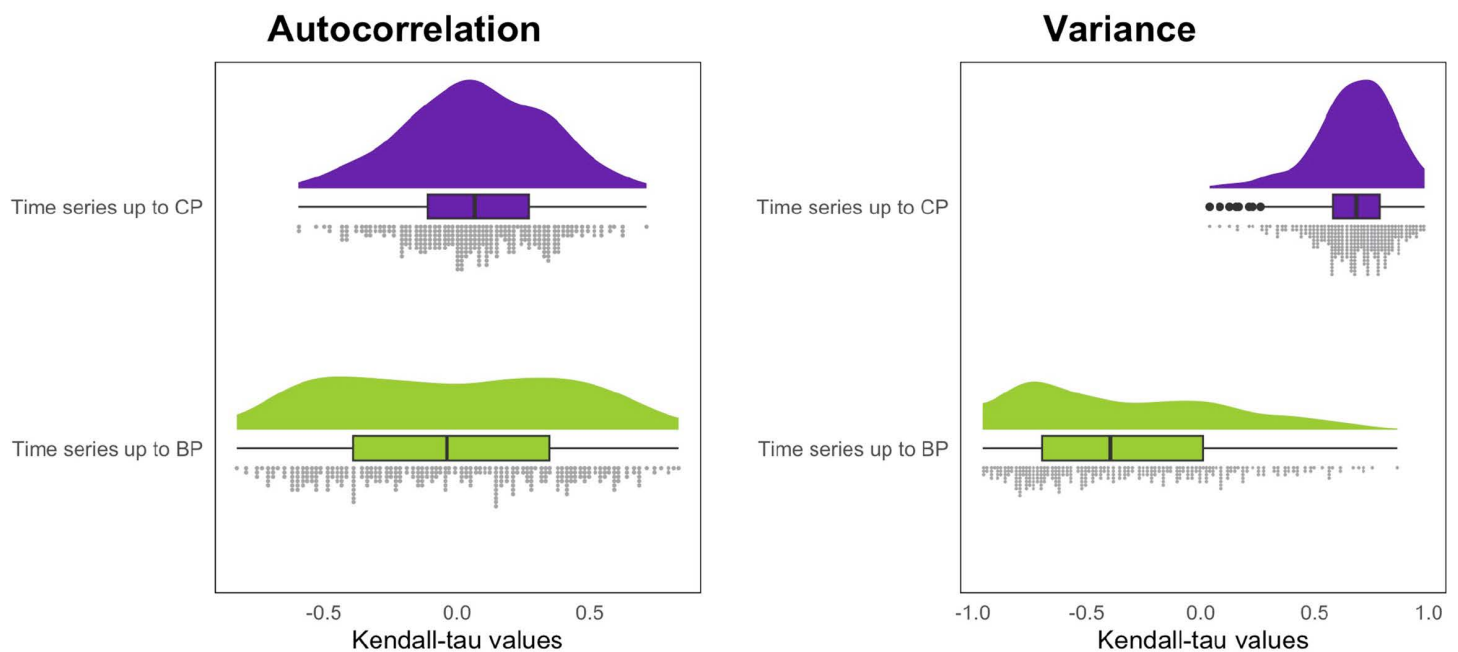


Fig 6. Raincloud plots showing Kendall- τ values for autocorrelation and variance, computed on time series of I generated from the model with multiplicative noise ($\sigma = 0.02$). The black dots in the right figure represent outliers (data points beyond the whiskers of the boxplot: $Q1 - 1.5 \times IQR$, $Q3 + 1.5 \times IQR$).

<https://doi.org/10.1371/journal.pone.0354091.g006>

computed on time series of M show similar performance: not all true positives are identified, and false positives occur in roughly half of the time series (see Table 2). The remaining noise-intensity cases of the multiplicative noise show qualitatively similar results.

Since many of the time series generated from the model with demographic stochasticity show advances, we could only use fewer time series for our summary; namely, those time series that showed delays.

EWSs — Receiver-Operating Characteristic (ROC) curves

The effectiveness of the EWSs in correctly identifying transitions in time series of I varies depending on the form of noise and its intensity. Variance has better predictive performance than autocorrelation, with almost perfect area under the curve (AUC) values for the model with multiplicative noise (Figs 7(a)–(b)). In the lowest noise-intensity case of the model with additive white noise, variance demonstrates a moderate predictive performance ($AUC \approx 0.71$), outperforming autocorrelation, whose performance is closer to random guessing ($AUC \approx 0.57$); see Fig 7(c). Variance and autocorrelation show similar performance for the remaining noise-intensity cases of the model with additive white noise (Figs 7(d)–(e)). For the models with additive or multiplicative white noise, the AUC for variance decreases as noise intensity increases (Figs 7(a)–(e)). Variance outperforms autocorrelation in the model with demographic stochasticity (Fig 7(f)), but note that these results are obtained from fewer time series than in the other two cases because the model with demographic stochasticity produced more time series with advances.

The high sensitivity of the social-media compartment to noise diminishes the detectability of the EWSs. For the models with additive or multiplicative white noise, variance and autocorrelation show similar performance, with AUC slightly above 0.5. Regardless of the noise intensity, variance and autocorrelation perform no better than random guessing in correctly predicting the transitions for the media compartment (Figs 8(a)–(e)). In the model with demographic stochasticity,

Table 2. Positive Kendall- τ values (%) for autocorrelation and variance computed prior to BP and prior to CP for the different stochastic models. AWN stands for additive white noise, MN stands for multiplicative noise and DS stands for demographic stochasticity. The best outcome in each circumstance is bolded.

Noise type	Variable	Noise intensity	Autocorrelation		Variance	
			BP	CP	BP	CP
AWN	<i>I</i>	low	45	59	42	100
		medium	49	78	51	93
		high	58	71	58	81
	<i>M</i>	low	53	65	56	68
		medium	55	45	56	46
		high	55	49	55	47
	<i>W</i>	low	53	72	55	85
		medium	58	51	61	55
		high	55	54	58	55
MN	<i>I</i>	low	48	62	26	100
		medium	51	51	33	99
		high	51	49	28	95
	<i>M</i>	low	50	45	49	45
		medium	46	47	47	50
		high	50	44	48	48
	<i>W</i>	low	13	55	96	100
		medium	28	49	100	98
		high	37	41	99	96
DS	<i>I</i>		46	55	48	69
	<i>M</i>		49	39	53	42
	<i>W</i>		45	55	43	73

<https://doi.org/10.1371/journal.pone.0354091.t002>

autocorrelation demonstrates a slightly better predictive performance ($AUC \approx 0.62$) compared to variance ($AUC \approx 0.56$), although both reflect limited predictive ability (see Fig 8(f)).

To further evaluate the sensitivity of our conclusions regarding the performance of the early warning signals (EWSs) on social-media time-series data, we repeated the simulations with the post-generation rates δ_S , δ_W , and δ_R set to zero. This restriction limited postings to only the “reporting” infected population. The resulting receiver-operating characteristic (ROC) curves are shown in Fig 9, providing a comparative assessment of EWSs’ performance under modified conditions. A comparable outcome was observed in the performance of EWSs when posts were generated exclusively by the “reporting” infected population *I*.

The performance of the EWSs in identifying transitions in the *W* compartment varies depending on the form of noise and noise intensity. In the lowest noise-intensity scenario of the model with multiplicative noise, variance and autocorrelation demonstrate better predictive performance, and variance outperformed autocorrelation with AUCs of 0.94 and 0.79, respectively (Fig 10(a)). As noise intensity increases, the AUC values for variance and autocorrelation decrease, indicating a decline in their predictive performance (Fig 10(a)–(b)). For higher noise intensities of the model with multiplicative noise, variance correctly identifies the true positives but also generates many false positives for the *W* compartment. Variance and autocorrelation demonstrate similar performance for the model with additive white noise, and their AUC values decrease as noise intensity increases (Figs 10(c)–(e)). In the model with demographic stochasticity (Fig 10(f)), variance demonstrates a better predictive performance ($AUC \approx 0.7$) than autocorrelation ($AUC \approx 0.55$).

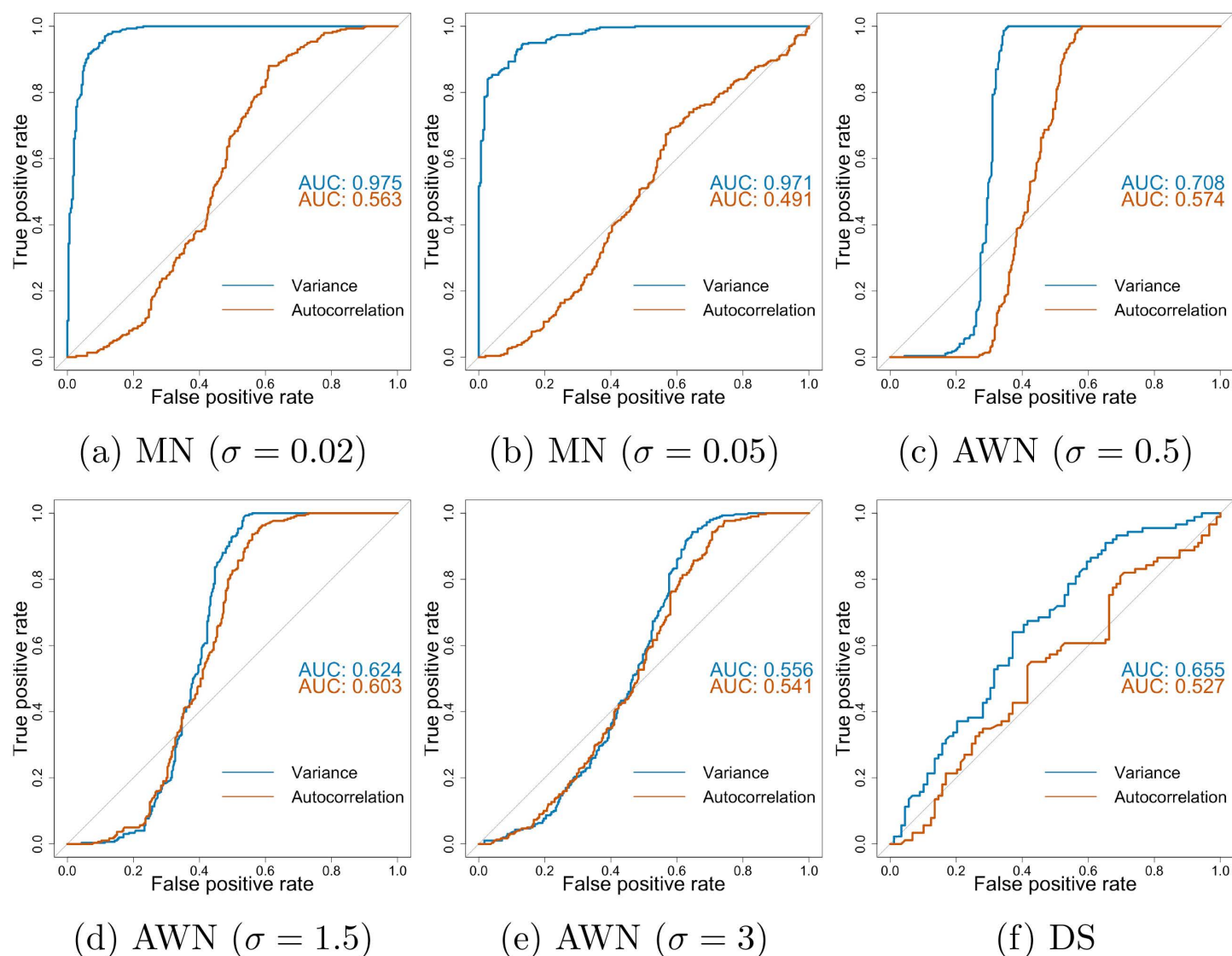


Fig 7. ROC curves comparing the performance of variance (blue) and autocorrelation (orange) in predicting the changepoints for I . The grey diagonal line represents a random classifier (AUC=0.5). MN stands for multiplicative noise, AWN stands for additive white noise and DS stands for demographic stochasticity.

<https://doi.org/10.1371/journal.pone.0354091.g007>

Discussion

Summary

Anticipating disease emergence using early warning signals (EWSs) is crucial for public-health preparedness. However, several factors can substantially affect the precision and reliability of EWSs. Our goal was to assess different components of disease dynamics while including a social-media compartment as a proxy, to determine whether they can be used to forecast disease emergence under various scenarios using EWSs and identify which were the most reliable. We used a minimal mechanistic model to study the general pattern of disease emergence, rather than focusing on a specific disease. We implemented different forms and intensities of noise to understand the effect of noise on EWSs. We also determined

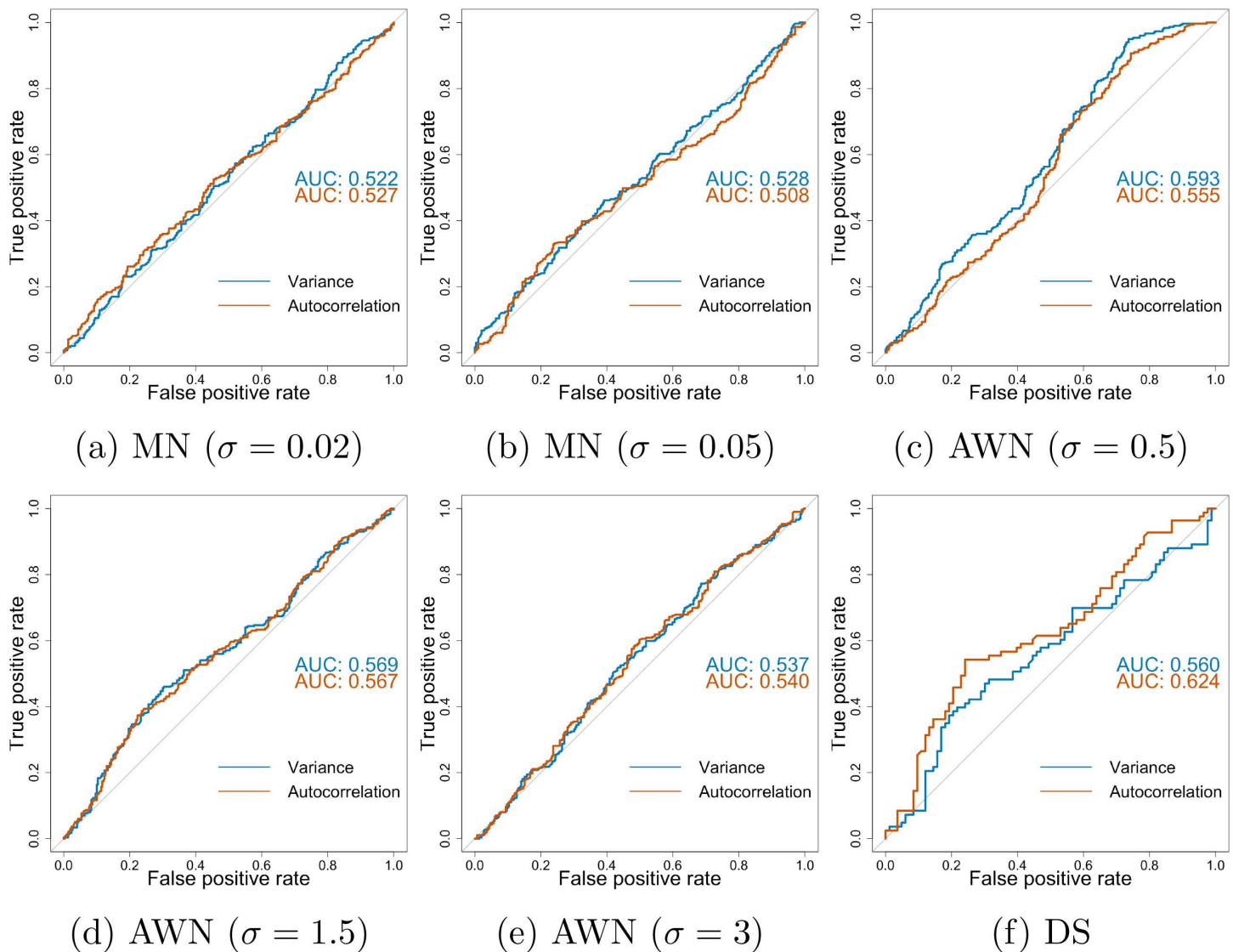


Fig 8. ROC curves comparing the performance of variance (blue) and autocorrelation (orange) in predicting the changepoint for M .

<https://doi.org/10.1371/journal.pone.0354091.g008>

which compartments gave the least delays (i.e., the earliest transitions). Finally, we analyzed the potential of social-media dynamics in predicting disease emergence. We used statistical techniques to assess and compare the performance of EWSs in these different scenarios.

Advance and delays

Many of our stochastic simulations showed the expected delay. This result is in agreement with the findings of Dibble et al. [26] and O'Regan and Drake [21], who reported a time lag between the system crossing the critical threshold and the onset of an outbreak. However, we found — particularly for the model with demographic stochasticity — that transitions can be induced before critical thresholds are reached, resulting in advances rather than delays. In the context of delays, we found that the “non-reporting” infected population (W) modelled with additive white noise or multiplicative noise

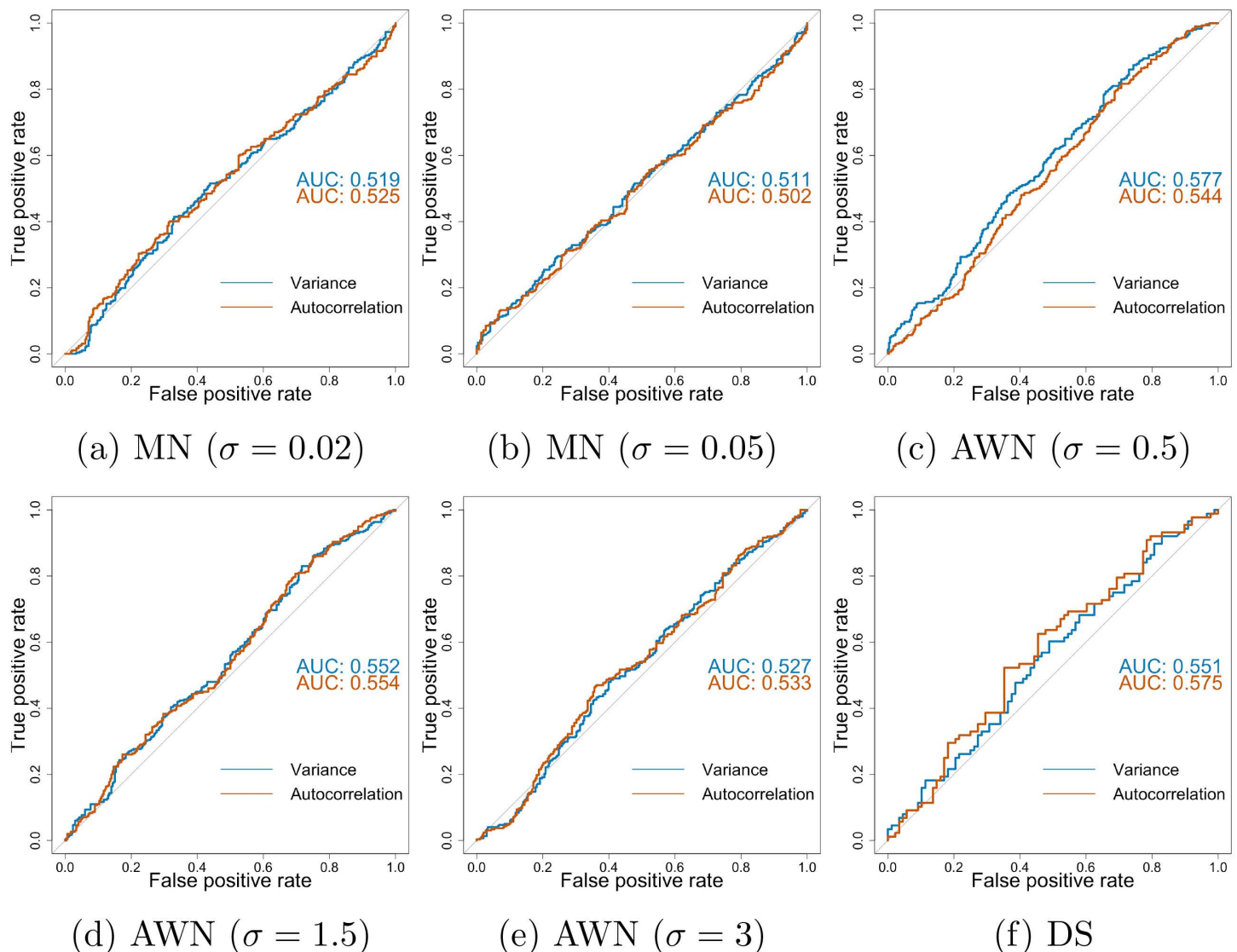


Fig 9. ROC curves comparing the performance of variance (blue) and autocorrelation (orange) in predicting the changepoint for M .

<https://doi.org/10.1371/journal.pone.0354091.g009>

exhibited earlier transitions than the “reporting” infected population (I). Since time-series data for the former is not available, we checked whether additional information can be obtained from the social-media compartment. The hypothesis that transitions in the social-media compartment may reflect transitions in the non-reporting infected population was only supported by the model with additive white noise (lowest noise-intensity case). Hence, the possibility that transitions in the infected populations can be predicted by transitions in the social-media compartment is greatly compromised by noise.

Performance of EWSs

The performances of variance and autocorrelation (lag-1) vary based on the form and intensity of noise. For the reporting infected population, variance had almost perfect area under the curve (AUC) with multiplicative noise, whereas the performance was moderate with additive white noise. However, its accuracy in distinguishing between true and false

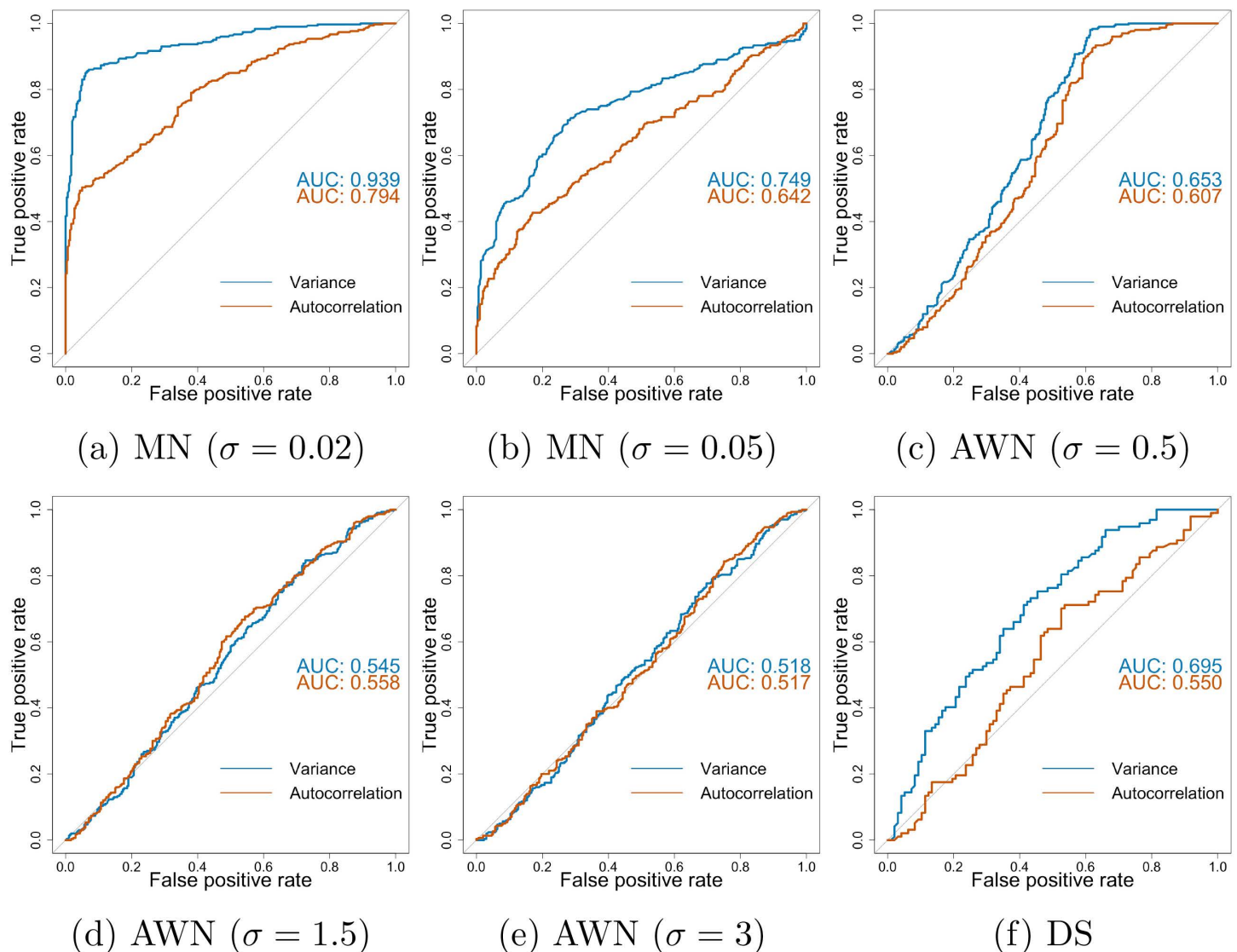


Fig 10. ROC curves comparing the performance of variance (blue) and autocorrelation (orange) in predicting the changepoint for W .

<https://doi.org/10.1371/journal.pone.0354091.g010>

positives, as measured by the AUC, reduces as noise intensity increases. A similar pattern in the EWSs was observed for the non-reporting infected population, but to a lesser degree. Our observation that the form of noise affects the performance of EWSs is in partial agreement with the findings of O'Regan and Burton [33]. Using a one-dimensional system, they found that trend in variance as indicator of disease elimination is sensitive to the underlying noise structure, but autocorrelation is robust to the different forms of noise. Chakraborty et al. [53] also found that autocorrelation outperformed variance in predicting disease emergence for different noise-induced models except for the SIR model with multiplicative noise. While these studies reported autocorrelation as a more reliable EWS than variance, other studies [2, 18, 54] have identified variance as the better EWS. In our case, autocorrelation performed poorly in indicating disease emergence, and variance was a more reliable EWS. Further work is necessary to evaluate when these EWSs perform well based on model complexity and dimensionality.

EWSs with social media

Although many studies have demonstrated the potential of social-media data and Google searches to predict infection trends using techniques in deep learning and natural language processing [36–38,55], the application of EWSs with social-media data in this context has not been explored. Our results showed that both variance and autocorrelation performed no better than random chance for predictions of disease emergence from the social-media compartment, regardless of the form and intensity of noise. This poor performance suggests that EWSs are more sensitive to the underlying social-media dynamics than variations in noise structure. Using a coupled behaviour-disease model with additive white noise, Pananos et al. [56] demonstrated that EWSs were effective in detecting a decline in vaccine uptake as a proxy for re-emergence of measles outbreak. To enhance the performance of EWSs in this regard, it may be necessary to account for additional critical factors, such as behavioural patterns, within the broader context of social-media activities.

Implications of delay

Delays may underestimate the risk of a disease outbreak over an extended period, thereby limiting the window for early and effective interventions. Using systems that exhibited delays, O'Regan and Drake [21] found that EWSs failed to predict disease emergence based on pre-bifurcation time-series data. This suggests the possibility of missed detections when transitions are delayed, underscoring the importance of continuous and adaptive monitoring of disease progression. While EWSs for disease models have typically been calculated close to the bifurcation point, our study shows that they are sometimes useful in anticipating transitions that happen later.

Potential validation strategy

Our study is based on simulated time-series data. However, to assess the performance of EWSs using past outbreaks, a feasible approach would be to consider high-resolution incidence data from diseases with known changes in transmission or interventions. The time-varying reproduction number R_t can be reconstructed using methods such as EpiEstim with an appropriate serial-interval distribution [53,57,58]. The same data can be analyzed for changepoints, and EWSs can be computed on segments of the time series leading up to both the R_t and the changepoint. Since we modelled the social-media compartment as a data source, EWSs could also be applied directly to tweets about the disease or its symptoms, following the same framework. (See, for example, Pananos et al. [56]).

Limitations

Despite the insights provided by our study, some limitations should be acknowledged. First, our model assumes that the entire population are social-media users and that there is uniformity in user behaviour. This does not fully reflect the heterogeneity observed in real-world social networks [37]. Incorporating structure into social-media dynamics and exploring parameter choices based on empirical data could provide a better representation for future studies on social-epidemic models and EWSs. Second, the data for our study was generated with specific parameters. These parameters were selected based on the objective of our study, which was to determine whether we can forecast disease emergence under specific yet representative conditions using EWSs. This approach allowed us to focus on the performance of EWSs, rather than on possible variations due to extensive parameter choices. While additional sensitivity analyses could be performed, they are beyond the scope of the present study and would be more appropriately undertaken in a disease-specific setting, which we will investigate in future work. Third, we used a specific length of time-series data to evaluate the performance of EWSs. Our choice was based on the exact length of time series leading up to the bifurcation point. EWSs generally require high-frequency data from longer time series for improved performance [2,34,35]. A practical limitation is that sufficiently long time-series data is often unavailable when a new disease is emerging. It would be helpful to examine how the performance of EWSs compares with shorter time-series data. Finally, we used identical noise intensity across

compartments for additive and multiplicative noise, which were chosen for comparison purposes, but which may be restrictive.

Recommendations

Our study highlights that variance as an EWS can be used to predict disease emergence from reported infected cases. We recommend that public-health authorities prioritize improving the timeliness and quality of reporting while strengthening monitoring efforts. This will increase the effectiveness of early disease prediction, enabling quicker and more targeted measures to mitigate disease outbreaks. One of the results of our study is that noise diminishes the performance of EWSs. Future studies should evaluate the performance of other existing EWSs using complex epidemiological models. Additionally, more robust EWSs are needed for real-world applications, where noise characteristics are often unknown or heterogeneous. Our findings in using EWSs to predict disease emergence from social media suggest that the reliability of this approach may be highly context-dependent. Future work should evaluate this approach across different diseases and clarify the effect of the underlying dynamics on how EWSs reflect disease trends from social-media activities. These aspects will further aid our understanding on how EWSs can be used with social-media dynamics as a complementary tool for early disease detection within public-health settings.

Supporting information

S1 Appendix. Appendix.
(PDF)

Acknowledgments

We are grateful to two anonymous reviewers for comments that helped improve the manuscript. This manuscript forms part of FO's Ph.D. thesis.

Author contributions

Conceptualization: Frithjof Lutscher.

Formal analysis: Francisca Olajide.

Investigation: Francisca Olajide, Frithjof Lutscher, Stacey R. Smith?.

Methodology: Francisca Olajide, Frithjof Lutscher.

Project administration: Frithjof Lutscher, Stacey R. Smith?.

Software: Francisca Olajide.

Supervision: Frithjof Lutscher, Stacey R. Smith?.

Validation: Francisca Olajide.

Visualization: Francisca Olajide.

Writing – original draft: Francisca Olajide.

Writing – review & editing: Frithjof Lutscher, Stacey R. Smith?.

References

1. Han BA, Drake JM. Future directions in analytics for infectious disease intelligence: Toward an integrated warning system for emerging pathogens. *EMBO Rep.* 2016;17(6):785–9. <https://doi.org/10.15252/embr.201642534> PMID: [27170620](https://pubmed.ncbi.nlm.nih.gov/27170620/)
2. Brett TS, Drake JM, Rohani P. Anticipating the emergence of infectious diseases. *J R Soc Interface.* 2017;14(132):20170115. <https://doi.org/10.1098/rsif.2017.0115> PMID: [28679666](https://pubmed.ncbi.nlm.nih.gov/28679666/)

3. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. 2022;20(4):193–205. <https://doi.org/10.1038/s41579-021-00639-z> PMID: [34646006](#)
4. Smith R. Modelling disease ecology with mathematics. Springfield: American Institute of Mathematical Sciences. 2008.
5. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int J Surg*. 2020;78:185–93. <https://doi.org/10.1016/j.ijsu.2020.04.018> PMID: [32305533](#)
6. Thompson JMT, Sieber J. Predicting Climate Tipping As A Noisy Bifurcation: A Review. *Int J Bifurcation Chaos*. 2011;21(02):399–423. <https://doi.org/10.1142/s0218127411028519>
7. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. *Nature*. 2009;461(7260):53–9. <https://doi.org/10.1038/nature08227> PMID: [19727193](#)
8. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface*. 2005;2(4):281–93. <https://doi.org/10.1098/rsif.2005.0042> PMID: [16849186](#)
9. Mohammad KM, Kamrujjaman Md. Stochastic differential equations to model influenza transmission with continuous and discrete-time Markov chains. *Alexandria Engineering Journal*. 2025;110:329–45. <https://doi.org/10.1016/j.aej.2024.10.012>
10. Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J Math Biol*. 1990;28(4):365–82. <https://doi.org/10.1007/BF00178324> PMID: [2117040](#)
11. Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One*. 2012;7(7):e41010. <https://doi.org/10.1371/journal.pone.0041010> PMID: [22815897](#)
12. Dakos V, Scheffer M, van Nes EH, Brovkin V, Petoukhov V, Held H. Slowing down as an early warning signal for abrupt climate change. *Proc Natl Acad Sci U S A*. 2008;105(38):14308–12. <https://doi.org/10.1073/pnas.0802430105> PMID: [18787119](#)
13. Southall E, Brett TS, Tildesley MJ, Dyson L. Early warning signals of infectious disease transitions: a review. *J R Soc Interface*. 2021;18(182):20210555. <https://doi.org/10.1098/rsif.2021.0555> PMID: [34583561](#)
14. van Nes EH, Scheffer M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am Nat*. 2007;169(6):738–47. <https://doi.org/10.1086/516845> PMID: [17479460](#)
15. Carpenter SR, Brock WA. Rising variance: a leading indicator of ecological transition. *Ecol Lett*. 2006;9(3):311–8. <https://doi.org/10.1111/j.1461-0248.2005.00877.x> PMID: [16958897](#)
16. Dakos V, van Nes EH, D'Odorico P, Scheffer M. Robustness of variance and autocorrelation as indicators of critical slowing down. *Ecology*. 2012;93(2):264–71. <https://doi.org/10.1890/11-0889.1> PMID: [22624308](#)
17. Carpenter SR, Brock WA. Early warnings of unknown nonlinear shifts: a nonparametric approach. *Ecology*. 2011;92(12):2196–201. <https://doi.org/10.1890/11-0716.1> PMID: [22352158](#)
18. Proverbio D, Kemp F, Magni S, Gonçalves J. Performance of early warning signals for disease re-emergence: A case study on COVID-19 data. *PLoS Comput Biol*. 2022;18(3):e1009958. <https://doi.org/10.1371/journal.pcbi.1009958> PMID: [35353809](#)
19. Scheffer M, Carpenter SR. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution*. 2003;18(12):648–56. <https://doi.org/10.1016/j.tree.2003.09.002>
20. O'Regan SM, Lillie JW, Drake JM. Leading indicators of mosquito-borne disease elimination. *Theor Ecol*. 2016;9:269–86. <https://doi.org/10.1007/s12080-015-0285-5> PMID: [27512522](#)
21. O'Regan SM, Drake JM. Theory of early warning signals of disease emergence and leading indicators of elimination. *Theor Ecol*. 2013;6(3):333–57. <https://doi.org/10.1007/s12080-013-0185-5> PMID: [32218877](#)
22. Tredennick AT, O'Dea EB, Ferrari MJ, Park AW, Rohani P, Drake JM. Anticipating infectious disease re-emergence and elimination: a test of early warning signals using empirically based models. *J R Soc Interface*. 2022;19(193):20220123. <https://doi.org/10.1098/rsif.2022.0123> PMID: [35919978](#)
23. Brett T, Ajelli M, Liu Q-H, Krauland MG, Grefenstette JJ, van Panhuis WG, et al. Detecting critical slowing down in high-dimensional epidemiological systems. *PLoS Comput Biol*. 2020;16(3):e1007679. <https://doi.org/10.1371/journal.pcbi.1007679> PMID: [32150536](#)
24. Erneux T, Mandel P. Imperfect Bifurcation with a Slowly-Varying Control Parameter. *SIAM J Appl Math*. 1986;46(1):1–15. <https://doi.org/10.1137/0146001>
25. Haberman R. Slowly Varying Jump and Transition Phenomena Associated with Algebraic Bifurcation Problems. *SIAM J Appl Math*. 1979;37(1):69–106. <https://doi.org/10.1137/0137006>
26. Dibble CJ, O'Dea EB, Park AW, Drake JM. Waiting time to infectious disease emergence. *J R Soc Interface*. 2016;13(123):20160540. <https://doi.org/10.1098/rsif.2016.0540> PMID: [27798277](#)
27. Pan Q, Huang J, Wang H. An SIRS model with nonmonotone incidence and saturated treatment in a changing environment. *J Math Biol*. 2022;85(3):23. <https://doi.org/10.1007/s00285-022-01787-3> PMID: [35986794](#)
28. Arumugam R, Lutscher F, Guichard F. Tracking unstable states: ecosystem dynamics in a changing world. *Oikos*. 2021;130(4):525–40. <https://doi.org/10.1111/oik.08051>
29. Eckley IA, Fearnhead P, Killick R. Analysis of changepoint models. *Bayesian Time Series Models*. Cambridge University Press. 2011. p. 205–24. <https://doi.org/10.1017/cbo9780511984679.011>

30. Arumugam R, Guichard F, Lutscher F. Early warning indicators capture catastrophic transitions driven by explicit rates of environmental change. *Ecology*. 2024;105(4):e4240. <https://doi.org/10.1002/ecy.4240> PMID: 38400588
31. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, et al. Anticipating critical transitions. *Science*. 2012;338(6105):344–8. <https://doi.org/10.1126/science.1225244> PMID: 23087241
32. Wang R, Dearing JA, Langdon PG, Zhang E, Yang X, Dakos V, et al. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature*. 2012;492(7429):419–22. <https://doi.org/10.1038/nature11655> PMID: 23160492
33. O'Regan SM, Burton DL. How Stochasticity Influences Leading Indicators of Critical Transitions. *Bull Math Biol*. 2018;80(6):1630–54. <https://doi.org/10.1007/s11538-018-0429-z> PMID: 29713924
34. Dutta PS, Sharma Y, Abbott KC. Robustness of early warning signals for catastrophic and non-catastrophic transitions. *Oikos*. 2018;127(9):1251–63. <https://doi.org/10.1111/oik.05172>
35. Bury TM, Sujith RI, Pavithran I, Scheffer M, Lenton TM, Anand M, et al. Deep learning for early warning signals of tipping points. *Proc Natl Acad Sci U S A*. 2021;118(39):e2106140118. <https://doi.org/10.1073/pnas.2106140118> PMID: 34544867
36. Yousefinaghani S, Dara R, Mubareka S, Sharif S. Prediction of COVID-19 Waves Using Social Media and Google Search: A Case Study of the US and Canada. *Front Public Health*. 2021;9:656635. <https://doi.org/10.3389/fpubh.2021.656635> PMID: 33937179
37. Wang A, Dara R, Yousefinaghani S, Maier E, Sharif S. A Review of Social Media Data Utilization for the Prediction of Disease Outbreaks and Understanding Public Perception. *BDCC*. 2023;7(2):72. <https://doi.org/10.3390/bdcc7020072>
38. Cheng IK, Heyl J, Lad N, Facini G, Grout Z. Evaluation of Twitter data for an emerging crisis: an application to the first wave of COVID-19 in the UK. *Sci Rep*. 2021;11(1):19009. <https://doi.org/10.1038/s41598-021-98396-9> PMID: 34561514
39. Loprete M, Panzarasa P, Puliga M, Riccaboni M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci Rep*. 2021;11(1):2147. <https://doi.org/10.1038/s41598-021-81333-1> PMID: 33495534
40. Sookninan J, Mays N. Harnessing Social Media in the Modelling of Pandemics-Challenges and Opportunities. *Bull Math Biol*. 2021;83(5):57. <https://doi.org/10.1007/s11538-021-00895-3> PMID: 33835296
41. Tchuente JM, Dube N, Bhunu CP, Smith RJ, Bauch CT. The impact of media coverage on the transmission dynamics of human influenza. *BMC Public Health*. 2011;11 Suppl 1(Suppl 1):S5. <https://doi.org/10.1186/1471-2458-11-S1-S5> PMID: 21356134
42. Casagrandi R, Bolzoni L, Levin SA, Andreasen V. The SIRC model and influenza A. *Math Biosci*. 2006;200(2):152–69. <https://doi.org/10.1016/j.mbs.2005.12.029> PMID: 16504214
43. Cui J, Sun Y, Zhu H. The Impact of Media on the Control of Infectious Diseases. *J Dyn Differ Equ*. 2008;20(1):31–53. <https://doi.org/10.1007/s10884-007-9075-0> PMID: 32214759
44. Pawelek KA, Oeldorf-Hirsch A, Rong L. Modeling the impact of twitter on influenza epidemics. *Math Biosci Eng*. 2014;11(6):1337–56. <https://doi.org/10.3934/mbe.2014.11.1337> PMID: 25365604
45. Allen LJ. An introduction to stochastic processes with applications to biology. CRC Press. 2010.
46. May RM. Stability and complexity in model ecosystems. Princeton University Press. 2019.
47. Greenwood PE, Gordillo LF. Stochastic Epidemic Modeling. *Mathematical and Statistical Estimation Approaches in Epidemiology*. Springer Netherlands. 2009. p. 31–52. https://doi.org/10.1007/978-90-481-2313-1_2
48. Kuske R, Gordillo LF, Greenwood P. Sustained oscillations via coherence resonance in SIR. *J Theor Biol*. 2007;245(3):459–69. <https://doi.org/10.1016/j.jtbi.2006.10.029> PMID: 17173935
49. Sauer T. Computational solution of stochastic differential equations. *WIREs Computational Stats*. 2013;5(5):362–71. <https://doi.org/10.1002/wics.1272>
50. Killick R, Eckley I. Changepoint: An R package for changepoint analysis. *Journal of Statistical Software*. 2014;58(3):1–19.
51. Dakos V, Lahti L. Earlywarnings: R Package Manual. 2022. <https://CRAN.R-project.org/package=earlywarnings>
52. Allen M, Poggiali D, Whitaker K, Marshall TR, van Langen J, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res*. 2021;4:63. <https://doi.org/10.12688/wellcomeopenres.15191.2> PMID: 31069261
53. Chakraborty AK, Gao S, Miry R, Ramazi P, Greiner R, Lewis MA, et al. An early warning indicator trained on stochastic disease-spreading models with different noises. *J R Soc Interface*. 2024;21(217):20240199. <https://doi.org/10.1098/rsif.2024.0199> PMID: 39118548
54. O'Brien DA, Clements CF. Early warning signal reliability varies with COVID-19 waves. *Biol Lett*. 2021;17(12):20210487. <https://doi.org/10.1098/rsbl.2021.0487> PMID: 34875183
55. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. *PLoS Negl Trop Dis*. 2017;11(1):e0005295. <https://doi.org/10.1371/journal.pntd.0005295> PMID: 28085877
56. Pananos AD, Bury TM, Wang C, Schonfeld J, Mohanty SP, Nyhan B, et al. Critical dynamics in population vaccinating behavior. *Proc Natl Acad Sci U S A*. 2017;114(52):13762–7. <https://doi.org/10.1073/pnas.1704093114> PMID: 29229821
57. Cori A, Cauchemez S, Ferguson NM, Fraser C, Dahlqvist E, Demarsh PA. Package 'epiestim'. Vienna, Austria: CRAN. 2020.
58. Ogi-Gittins I, Hart WS, Song J, Nash RK, Polonsky J, Cori A, et al. A simulation-based approach for estimating the time-dependent reproduction number from temporally aggregated disease incidence time series data. *Epidemics*. 2024;47:100773. <https://doi.org/10.1016/j.epidem.2024.100773> PMID: 38781911