

Building test data from real outbreaks for evaluating detection algorithms

Gaëtan Texier, Michael Jackson, Leonel Siwe, Jean-Baptiste Meynard, Xavier Deparis, Herve Chaudet

▶ **To cite this version:**

Gaëtan Texier, Michael Jackson, Leonel Siwe, Jean-Baptiste Meynard, Xavier Deparis, et al.. Building test data from real outbreaks for evaluating detection algorithms. PLoS ONE, 2017, 12 (9), pp.e0183992. 10.1371/journal.pone.0183992 . inserm-01986527

HAL Id: inserm-01986527

https://inserm.hal.science/inserm-01986527

Submitted on 18 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS ONE

RESEARCH ARTICLE

Building test data from real outbreaks for evaluating detection algorithms

Gaetan Texier[1,2]*, Michael L. Jackson[3], Leonel Siwe[4], Jean-Baptiste Meynard[5], Xavier Deparis[5], Herve Chaudet[2]

1 Pasteur Center in Cameroun, Yaoundé, Cameroun, **2** UMR 912 / SESSTIM - INSERM/IRD/Aix-Marseille University / Faculty of Medicine - 27, Bd Jean Moulin, Marseille, France, **3** Group Health Research Institute, Seattle, United States of America, **4** Sub-Regional Institute of Statistics and Applied Economics (ISSEA), Yaoundé, Cameroun, **5** French Armed Forces Center for Epidemiology and Public Health (CESPA), Camp de Sainte Marthe, Marseille, France

* gaetex1@gmail.com

Check for updates

OPEN ACCESS

Citation: Texier G, Jackson ML, Siwe L, Meynard J-B, Deparis X, Chaudet H (2017) Building test data from real outbreaks for evaluating detection algorithms. PLoS ONE 12(9): e0183992. https://doi.org/10.1371/journal.pone.0183992

Editor: Paul Robert Cleary, Health Protection Agency, UNITED KINGDOM

Received: June 1, 2016

Accepted: August 16, 2017

Published: September 1, 2017

Copyright: © 2017 Texier et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

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

Abstract

Benchmarking surveillance systems requires realistic simulations of disease outbreaks. However, obtaining these data in sufficient quantity, with a realistic shape and covering a sufficient range of agents, size and duration, is known to be very difficult. The dataset of outbreak signals generated should reflect the likely distribution of authentic situations faced by the surveillance system, including very unlikely outbreak signals. We propose and evaluate a new approach based on the use of historical outbreak data to simulate tailored outbreak signals. The method relies on a homothetic transformation of the historical distribution followed by resampling processes (Binomial, Inverse Transform Sampling Method—ITSM, Metropolis-Hasting Random Walk, Metropolis-Hasting Independent, Gibbs Sampler, Hybrid Gibbs Sampler). We carried out an analysis to identify the most important input parameters for simulation quality and to evaluate performance for each of the resampling algorithms. Our analysis confirms the influence of the type of algorithm used and simulation parameters (i.e. days, number of cases, outbreak shape, overall scale factor) on the results. We show that, regardless of the outbreaks, algorithms and metrics chosen for the evaluation, simulation quality decreased with the increase in the number of days simulated and increased with the number of cases simulated. Simulating outbreaks with fewer cases than days of duration (i.e. overall scale factor less than 1) resulted in an important loss of information during the simulation. We found that Gibbs sampling with a shrinkage procedure provides a good balance between accuracy and data dependency. If dependency is of little importance, binomial and ITSM methods are accurate. Given the constraint of keeping the simulation within a range of plausible epidemiological curves faced by the surveillance system, our study confirms that our approach can be used to generate a large spectrum of outbreak signals.

1 Introduction

Many outbreak detection algorithms (ODA) for disease surveillance have been proposed in the literature (see for example Chen's review [1]), and designers of epidemiological and laboratory surveillance systems need to opt for the most suitable.

The choice of an algorithm should be based on reproducible evaluations carried out under conditions as close to real life as possible [2]. Several authors including Buckeridge [3] and Jackson [4] have shown that algorithm evaluation depends on the characteristics of the outbreak curve (shape, duration and size) and the baseline (mean and variance). The evaluation process thus depends heavily on the quality of test data and the control of characteristics known to influence these results [5].

To evaluate an ODA, specific datasets composed of outbreak curves over a baseline are usually needed. These two components can be obtained from real data, synthetic data or from a combination of the two (semi-synthetic dataset), as described in Mandl [6] and Buckeridge [5].

The baseline is often obtained from the real, daily, routine data of disease surveillance systems, which are specific to the population under surveillance [3, 7]. Useful simulation tools also exist for simulating a baseline (such as Project Mimic, http://www.projectmimic.com/).

However, obtaining a sufficient quantity of outbreak curves with a realistic shape and covering a sufficient range of agents, size and duration, is known to be very difficult [3–5, 8]. With regard to disease outbreak curves, several authors have pointed out [3, 5, 7, 9] that the use of real data for outbreak signals is hindered by the shortage of real datasets that include enough clearly-defined outbreak signals. Natural outbreak signals cannot enable the control of features required for evaluating ODA (such as outbreak size, duration and shape). Buckeridge et al. [5] show that among 22 studies using natural outbreak signals, only five had enough outbreaks to evaluate and report quantitative measures such as sensitivity and timeliness. They conclude that evaluations based on natural signals should be reserved for qualitative evaluation. Similarly, Wagner [9] believes that a minimum of roughly 10 outbreaks is required to build a quantitative evaluation.

In contrast, synthetic outbreak curves have the potential to express the diversity of biological threats that a surveillance system could encounter and must detect. Synthetic datasets allow precise specification of the outbreak curve and make it possible to build large test sets. They are more convenient for sensitivity analyses and for exploring the effects of curve changes. Most evaluation studies in the last decade have used semi-synthetic datasets that combine real baseline data with synthetic outbreaks [5].

Several methods can be used to build synthetic outbreak curves. Mechanistic disease models, such as the Susceptible-Infected-Recovered (SIR) model, are commonly used [10, 11]. Agent-based models (ABM) such as EpiSims [12] build a network-based model of disease spread instead of a compartment-based one [9].

However, the purpose of most studies that simulate curves based on these approaches is not to generate outbreak curves to evaluate ODA, but is usually to predict an overall outbreak trend (until real-time quality forecasting is achieved and to identify realistic averaged parameters for describing outbreaks. For example, Nsoesie et al. [13] used an agent-based model, which requires the building offive preliminary models. They worked mainly on influenza outbreak forecasting and conceded the need for each surveillance team to build a social network model adapted to the population under surveillance.

One limitation of this type of approach is that a specific model must be constructed for each type of disease to be simulated [6, 13]. These models can be computationally expensive and data intensive, as well as difficult to set up for a non-specialist user. Building a realistic

model can be a particularly challenging task due to the lack of knowledge and real data available for setting up model parameters and for the validation process [14]. Held [15] considers such mechanistic modeling as too ambitious to be used routinely in a disease surveillance system, pointing out that the non-availability of certain information (number of susceptible people, model parameters, etc.) renders the process unfeasible. In addition, these models are usually based on multiple assumptions that may be difficult to verify [9]. Due to the high level of statistical and mathematical modeling needed and according to our experience of disease surveillance system deployment, we are not convinced that the majority of users in charge of disease surveillance systems are able to generate curves with this approach and wish to do it.

A simpler approach is to build the synthetic outbreak curve by using a simple mathematical distribution that describes its expected shape. Uniform, linear, exponential, sigmoid, lognormal or Dirac distributions are often used [6, 7, 16–19] (S1 Fig). However, estimating the behavior of an ODA for real outbreaks from simple synthetic signals is not straightforward. The principal limitation of evaluations using semi-synthetic datasets is that they only find the smallest peak that could be detected, leaving the critical question of detectability unanswered: "What is the smallest outbreak that can be detected?". We have known since Wagner[9], Buckeridge [3] and Jackson [20] that the answer is influenced by the outbreak shape for the selected size and duration, leading us to believe that ODA evaluations based solely on the injection of geometric shape can only be considered as partial evaluations.

Bollen [21] considers that a curve reflects latent determinants, which are hidden. Baker [22] and Tucker [23] also consider that observations of repeated measures over time are the result of latent factors and characteristics that can only be indirectly observed using real curves. Similarly, we may consider that a curve shape depicts a type of disease outbreak determined by a transmissible agent, a population, an environment and a collecting structure, which are the latent determinants. To improve the realism of produced curves, some authors propose that the modeling process be extended, by taking into consideration the effects of disease surveillance systems. In the case of the statistical distribution simulation approach, Lotze and coll. [7] proposed that system effect (day of the week, seasonal and holiday effect) be added to initial outbreak lognormal signatures.

Despite this improvement, the traditional methods (mechanistic or statistical distribution-based methods) for obtaining synthetic epidemic curves remain, for the most part, sophisticated or poorly realistic.

Finally, the main way to generate an outbreak curve is by using an explicit propagation model (ABM, SEIR, distribution modeling, etc.). Modeling this process and quantifying the relative weight of each known factor are valuable for understanding the mechanisms influencing the outbreak dynamics. But all this modeling process is then required even if the final goal of this simulation is not to understand the epidemiological mechanisms producing the curve or to implement an explicative/forecasting model of the outbreak. Because our main goal here is to produce a realistic yet customized curve for a pathogen in a large range of surveillance situations, the relevance of model-based simulation for this specific purpose must be questioned. Indeed, there is a risk of generating only generalist or averaged curves that can be non-representative of the wide variety of surveillance curves, and the simulation process may suffer from a lack of realism. For example, the "classic" norovirus epicurve used in our study (Fig 1) [24] should be simulated by a Gaussian model, but is not able to represent the "atypical" epicurve observed in a specific population during a cruise [25]. Outbreak modeling is known to be highly complex [26], covering highly variable outbreak expressions and requiring numerous hypotheses and parameters. Taking an extreme position, some authors consider that carrying out realistic outbreak modeling is risky [27, 28]. As explained before, a latent model probably exists where the population under surveillance, the pathogens, the surveillance system and

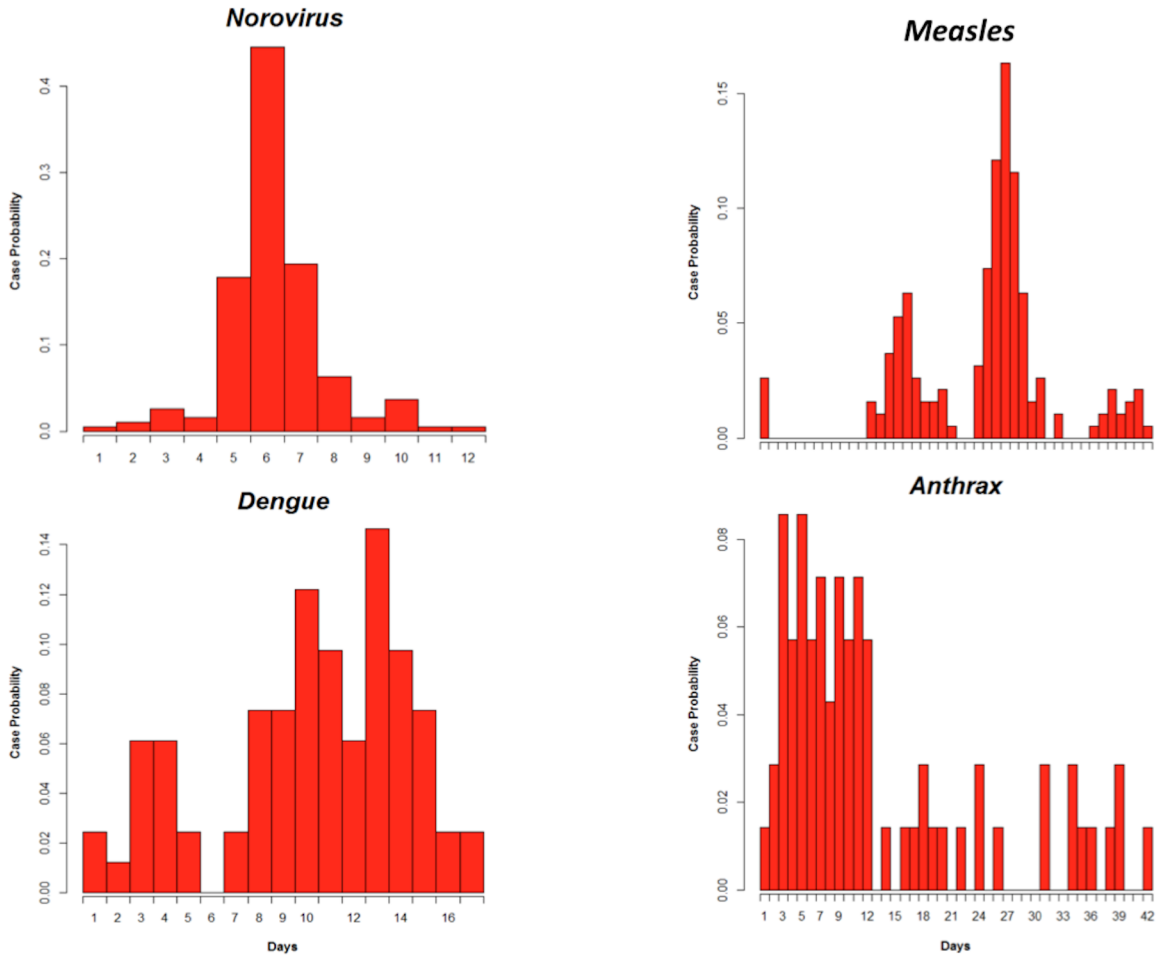


Fig 1. Probability mass function (PMF) for four published epidemic curves (norovirus, measles, anthrax and dengue outbreaks) represented by the case probability according to the time step of the outbreak [30–33].

https://doi.org/10.1371/journal.pone.0183992.g001

many other factors closely interact. An epicurve can be seen as the directly observable result of all these co-acting factors in association with unknown co-factors. We also need to take into consideration the risk of model misspecification, which may produce unrealistic or inaccurate curves. This topic is developed in the discussion proposed by Nishiura [29] and Horner [30], associated with a possible misspecification of the lognormal model proposed by Sartwell in 1966 [31] and largely used by the community. If we adopt this approach to consider the cruise outbreak, Gaussian simulation of all norovirus curves should be considered as a misspecification of the "atypical" curve (more adequately represented by an exponential decay or a Gamma model with different parameters).

Another possibility is to use simulation solutions that do not require modeling beforehand, such as Monte Carlo resampling methods. Indeed, these methods are tailored for simulating complex curves, especially when mathematical formalizations are not accessible, which is frequently the case for epidemic curves. S2 Fig describes what Monte Carlo resampling methods could add to the existing solutions used in the literature.

This study aims at verifying if a new and simple approach based on historical outbreak curves can be used to produce a range of outbreak signals (in terms of diversity of curve shapes) including very unlikely signals. The generated dataset of signals should reflect the

likely distribution and the diversity of authentic signals faced by the surveillance system, while conserving as much as possible the shape realism that must conform to the apparent outbreak characteristics conveyed by the curve shape. To achieve this goal, we will evaluate if this new approach using homothetic transformation and resampling methods applied to historical disease outbreaks can be used to produce customized curves by controlling outbreak duration and size, while preserving curve shape (identified as critical for an ODA evaluation).

2 Material and methods

If we want to test algorithm ability and limitations under a large range of real surveillance situations, an adequate simulation tool must also be able to create the entire spectrum of outbreak curves, even those considered to be "atypical" but already-observed curves (qualified as "outliers") must be considered, provided their atypical nature is recognized. The risk of a statistical modeling approach is that it might restrict the scope of the simulation used for the algorithm evaluation to an average behavior of the epidemics, which could be non-representative of the large range of shape varieties observed for the same pathogen. Using outlier curves precisely allows algorithm limitations to be evaluated against unusual outbreaks. This is why we chose real datasets from published disease outbreaks, representing different modes of transmission (Fig 1) and a range of shape varieties.

Probability mass functions (PMF) and cumulative distribution functions (CDF) will be extracted from the outbreak curves by computing the probability of the number of cases (or the cumulative probability) for each day of the epidemic. We propose that homothetic transformation (known to be a uniform scaling technique that preserves form) be used to control outbreak duration, together with a resampling method based on the Monte Carlo approach to manage its size. S2 Fig gives an example of the simulation and evaluation process (detailed above) in the case of a measles outbreak shape, for a 14-day duration and 30-case outbreak simulation.

We used R 3.0.1 (R Core team 2013) to implement all the methods described below [32].

2.1 Scaling outbreak duration

We apply a homothetic transformation to a PMF covering n time units to compute a new probability distribution lasting m time units. This transformation is a uniform dilation of the number of days using a scaling factor r, such as:

$$P_j = \sum_{i=1}^{n} f_{ij}.\pi_i \quad with \quad \sum_{i=1}^{n} \pi_i = 1 \; ; \; \sum_{j=1}^{m} P_j = 1 \; ; \; r = \; n/m$$

Where:

- π_i is the initial probability at time i in the PMF;
- P_j is the final probability at time j in the new probability distribution;
- f_{ij} is the fraction of π_i at time i participating in P_j and depending on r = n/m.

A detailed R script is proposed (S1 Script) for $r \leq 1$.

2.2 Scaling outbreak size

Until now, outbreak curves have mostly been produced by using a certain explicit outbreak propagation model. This model building is thus considered as a prerequisite, even if the final goal of the simulation is not to understand the mechanisms and factors shaping the curve, nor

to implement an explicative model of the outbreak. We adopt a pragmatic, sparing approach by skipping the usual steps of modeling that are required to produce an outbreak curve (S2 Fig). The aim of scaling is to give the outbreak size an intended value while preserving curve shape. We propose that this goal be achieved by testing six Monte Carlo resampling methods.

2.2.1 Inverse transform sampling method (ITSM) for discrete random variables. This method requires that only the form of the density of interest (PMF) and its CDF be known. The random process of sampling needs to create a uniformly-distributed variate U in [0,1] for each simulated case. The case dates are found by reading on the CDF the random value (x-axis) corresponding to the variate U on the cumulative probability axis (y-axis). The process is repeated until the expected size of the outbreak is reached.

2.2.2 Binomial method. This method (described in Kachitvichyanukul [33]) makes it possible to simulate independent and identically-distributed (i.i.d.) variables according to an outbreak target distribution (PMF). It considers an outbreak day as a binomial event of the epidemic with a success probability equal to π_i (where i is the day of the outbreak). A case is assigned to a day according to the binomial distribution law. This process is repeated until the case sum is equal to the expected size of the outbreak.

2.2.3 Random walk Metropolis-Hastings. A Metropolis-Hastings algorithm [34, 35] is a Markov chain Monte Carlo sampling method that allows the progressive construction of a sample by taking into account a simulated value (x_{t-1}) in order to generate the next one (x_t). With Metropolis-Hastings, the Markov chain is built using the original PMF as a stationary density that limits the distribution of the Markov chain [36]. Metropolis-Hastings needs to build a conditional density $q(x_t|x_{t-1})$ (or transition kernel) associated with the original target density (the outbreak curve PMF).

Random Walk Metropolis-Hastings (RWMH) [36] is a practical approach to this method, which carries out a local exploration of the neighborhood of the current value of the Markov chain. The conditional density depends on the previously-simulated value and a random perturbation (ε_t):

$$x_t = x_{t-1} + \varepsilon_t.$$

This random perturbation is the random walk and follows a symmetric distribution (g) centered on zero. The transition kernel between simulated cases is defined by choosing a conditional density:

$$q(x_t|x_{t-1}) = g(x_t - x_{t-1}).$$

In our study, g followed a uniform distribution and, as proposed by Robert [37], the scale parameter of ε_t was defined as the integer part of the distribution variance.

The simulation process consists in choosing an initial outbreak date and another epidemic day by using the transitional kernel $q(x_t|x_{t-1})$. Adding a case at this new date will be accepted or rejected according to the Hastings ratio.

To set the starting point of the simulation, we used a burn-in period of 1,000 iterations. The same period was used for the other MCMC methods.

2.2.4 Independent Metropolis-Hastings Algorithm (IMHA). The Independent Metropolis-Hastings Algorithm (IMHA) [35] generates a candidate conditional density that depends only on the present state of the chain.

This density can thus be written: $q(x_t|x_{t-1}) = q(x_t) = g(x_t)$. This method ensures the correlation of the simulated cases.

The process is the same as for RWMH, but the choice of the new date depends on an independent transitional kernel. In our study, the conditional density is a discrete uniform distribution defined over the same range of days as the original PMF.

2.2.5 Gibbs sampling algorithm. This algorithm [36] is used to simulate Markov chain states from a joint distribution $q(x_t;y_t)$. This requires knowledge of the two conditional distributions $q(x_t|y_t)$ and $q(y_t|x_t)$ that must be built.

We chose slice sampling within the Gibbs sampler approach [38] since it is formally applied with no restriction to any density on its shape or dimension. We consider disease outbreak days (x) and simulated cases (y) as the variables and use a uniform distribution (x, y), which preserves the curve shape despite y changes:

$$p((X,Y)=(x,y)) = \begin{cases} \frac{1}{N} & \text{if } 0 < y < \pi(x) \\ & 0 \text{ else} \end{cases}.$$

According to this joint distribution, we construct the conditional distribution ($q(x_t|y_t)$ and $q(y_t|x_t)$) and use a Gibbs sampling process.

This sampling consists in selecting an initial date for the simulation and choosing the next case from it by using the conditional density $q(y_t|x_t)$. We then simulate the next date from this case by using the conditional density $q(x_t|y_t)$.

A shrinkage procedure is used in this context because we cannot directly simulate a case using $q(x_t|y_t)$. It consists in randomly choosing a simulated date (x) and testing by means of an interval whether this date can be generated according to the conditional density $q(x_t|y_t)$.

2.2.6 Hybrid Gibbs sampler (Metropolis-within-Gibbs). Sometimes the density function obtained from the disease outbreak curve may be unconventional, i.e. discontinuous or resulting from a mixing of different distribution laws (as in the Sverdlosk event or in a measles outbreak). It is difficult for Markov chains to move over discontinuous curves [25]. In theory, the Gibbs sampler should encounter certain difficulties in simulating the shape of this type of curve since some densities cannot be directly simulated. A hybrid Gibbs sampler, in the Metropolis-within-Gibbs strategy sense, may be a solution [39]. The idea is to use a preliminary step consisting of a random walk simulation of Markov chain states instead of a direct simulation within Gibbs sampling. Hybrid strategy associates the conditional density $q(x_t|y_t)$ with the Hastings ratio, allowing the exploration of all the outbreak curves.

2.3 Quality assessment

When the duration or the number of cases during an outbreak simulation is changed, the shape of the outbreak (representing a latent process) must be maintained within a range of plausible epidemiological curves, which could be defined as the range of real triplets (size, duration, disease) observed during real outbreaks (obtained from publications). This necessity is also true for mechanistic or statistical distribution-based methods that may also face the risk of simulating an unrealistic outbreak.

The same problem was highlighted by the High-Fidelity Injection Detectability Experiments (HiFIDE) project in 2005. It tried to extend the semi-synthetic method by conserving the relationship between "the magnitude of the real outbreak and the strength of the signal in the surveillance data collected during the real outbreak" [9]. Without forgetting this simulation constraint during a real evaluation, in a first theoretical approach (proof of concept) with the aim of testing the capacity and the limits of MC methods, we chose to generate each triplet resulting from the combination of 12 sizes (10, 20, 30, 50, 100, 300, 500, 1000, 2000, 3000, 4000 and 5000 cases), six durations (5, 7, 14, 28, 42, 56 days) and four real outbreak shapes

(Norovirus, Measles, Anthrax, Dengue) [24, 40–42]. We replicated each simulation 20 times and used a median for evaluating stochastic algorithm performance, as proposed by Birattari [43].

Since all the methods proposed in this study are able to precisely simulate duration and size, quality assessment will be based only on the degree of conformity to the real outbreak shape.

2.3.1 Goodness-of-fit (GOF). Curve shape preservation was evaluated using the distance between the original real curve and the simulated one. We chose a GOF approach to evaluate this distance considered to be an indicator of simulation quality.

A Goodness-of-Fit (GOF) test between two distributions is usually indicated to determine whether an observed distribution corresponds to a particular probability distribution (such as the historical outbreak distribution). Due to the constraints of the test approach (e.g. the necessity to define a significance level α that varies with the quantity of simulated data), we chose to use GOF metrics (distance, divergence, similarity) as proposed by several authors [44, 45] to compare simulation qualities.

There is currently no consensus in the literature on the choice of a metric. Moreover, since there is no certainty regarding the convergence of quality results according to different metrics, we selected a limited number of metrics corresponding to different GOF properties. Due to the discrete nature of data, we selected the L2-distance built for comparing distance between two densities. Pearson divergence and Kolmogorov distance were chosen since they are derived from the statistics used to evaluate GOF between two probability density functions. And to evaluate similarity, we chose the Overlap (delta) Coefficient and Matusita's (Rho) Measure [44, 45]. All the GOF metrics are detailed in the S1a and S1b Table.

To avoid scale effects, all metrics were normalized. A two-sample T-test was used to compare the metrics between two groups.

2.3.2 Overall scale factor. Simulation relies on a scaling process. We assumed this scaling process could affect the quality of results, and in particular lead to a possible loss of information resulting from scattering a small number of cases over the duration of the outbreak.

To estimate the influence of the scaling process on simulations, we first calculated the case density for each distribution as the number of cases divided by the number of days. We then calculated the scale factor as the case density ratio of the simulation divided by the case density of the original distribution. During our work, we studied the impact of this scale factor on the quality of the simulations evaluated by different metrics.

3 Results and discussion

We first identified the input parameters that contribute most to the accuracy of results and investigated the way a change of input parameters influences results (Table 1).

3.1 The influence of outbreak duration and size

Quality decreases with the duration of an outbreak and increases with its size for all combinations of curve, algorithm and metric chosen for the evaluation. An example for a measles outbreak is presented in Fig 2. Two groups of algorithms (the Binomial, ITSM and Gibbs group versus the Metropolis Algorithms and Hybrid group) can be distinguished based on their sensitivity to the outbreak duration. The first group seems less influenced by a change in duration (Fig 2a). The GOF difference (normalized Pearson metric) is significant between these two different groups ($p<0{,}001$).

Rho distance for the difference of inadequacy between binomial and independent Metropolis algorithms averages 11% [min = 0.5%—max = 31%] (Fig 2b). The maximum difference

Table 1. Evaluation of six resampling methods according to outbreak type, number of cases and number of days simulated.

Outbreak	Measles	Anthrax	Dengue	Norovirus		
Binomial	0.025* *(0.00)*†	0.033 *(0.00)*	0.042 *(0.00)*	0.036 *(0.00)*		
ITSM	0.027 *(0.64)*	0.033 *(0.70)*	0.042 *(0.73)*	0.041 *(0.99)*		
Metrop. RW	0.082 *(0.53)*	0.071(0.57)	0.086 *(0.58)*	0.062 *(0.81)*		
Metrop. Ind	0.085 *(0.54)*	0.089 *(0.59)*	0.080 *(0.64)*	0.097(0.80)		
Gibbs	0.037 *(0.77)*	0.039 *(0.70)*	0.045 *(0.66)*	0.038 *(0.95)*		
Hybrid	0.083 *(0.81)*	0.088 *(0.90)*	0.075 *(1.04)*	0.096 *(1.00)*		
Cases	10	50	100	500	1000	5000
Binomial	0.185 (0.00)	0.038 (0.00)	0.020 (0.00)	0.004 (0.00)	0.002 (0.00)	0.000 (0.00)
ITSM	0.190 (0.00)	0.043 (0.03)	0.020 (0.06)	0.004 (0.28)	0.002 (0.57)	0.000 (2.87)
Metrop. RW	0.327 (0.21)	0.099 (0.23)	0.060 (0.24)	0.015 (0.36)	0.008 (0.51)	0.002 (1.77)
Metrop. Ind	0.411 (0.22)	0.107 (0.23)	0.056 (0.25)	0.011 (0.38)	0.006 (0.53)	0.001 (1.82)
Gibbs	0.201 (0.26)	0.047 (0.28)	0.024 (0.29)	0.005 (0.44)	0.003 (0.62)	0.000 (2.23)
Hybrid	0.401 (0.34)	0.104 (0.35)	0.057(0.38)	0.011 (0.55)	0.005 (0.77)	0.001 (2.65)
Days	5	14	28	42	56	
Binomial	0.005 (0.00)	0.018 (0.00)	0.036 (0.00)	0.057 (0.00)	0.081 (0.00)	
ITSM	0.006 (0.64)	0.021(0.71)	0.039 (0.77)	0.057 (0.87)	0.085 (0.93)	
Metrop. RW	0.016 (0.60)	0.046 (0.62)	0.088 (0.60)	0.124 (0.63)	0.155 (0.65)	
Metrop. Ind	0.011 (0.65)	0.046 (0.65)	0.096 (0.63)	0.156 (0.64)	0.199 (0.64)	
Gibbs	0.008 (0.72)	0.024 (0.78)	0.048 (0.81)	0.058 (0.83)	0.087 (0.74)	
Hybrid	0.011 (0.78)	0.046 (0.86)	0.092 (0.96)	0.155 (1.12)	0.190 (1.13)	

*Normalized Pearson distance averaged for all other parameters

† (Average time of simulation in seconds)

Binomial = Binomial method / ITSM = Inverse transform sampling method / Metrop. RW = Metropolis-Hasting Random Walk /Metrop. Ind = Metropolis-Hasting independent sampler / Hybrid = Hybrid Gibbs sampler / Gibbs = Gibbs sampling algorithm

https://doi.org/10.1371/journal.pone.0183992.t001

between the two groups was observed for outbreaks of 20 cases, with an average difference of 24%. Difference increases continuously and linearly with outbreak duration.

Simulation quality increases exponentially ($R^2 > 0.97$) with the number of cases (Fig 2b) and with decreasing outbreak duration.

With regard to the interaction between size and duration, the Metropolis-Hasting independent sampler and Hybrid Gibbs sampler are nearly equivalent in terms of accuracy. Random Walk Metropolis-Hasting performs slightly better in extreme simulation values. The real gap appears with Gibbs sampling, followed by ITSM and binomial methods for all simulations (Fig 3) and curve shapes.

3.2 Comparison of algorithms

This work examines how to simulate realistic infectious disease outbreak curves with a minimum number of parameters. As for any temporal simulation, we needed to provide two parameters: duration and size, and as is the case for a simple mathematical distribution approach, we had to choose a distribution shape among a set of real outbreaks. The resampling methods we tested, especially the binomial method and ITSM, are easy to implement even for a non-specialist. Although all these methods are able to generate a simulated curve from real outbreaks, their accuracy is algorithm-dependent.

Simulation quality for all diseases considered was measured (using different Goodness-of-fit measures) by calculating the difference between the generated outbreak curve and its

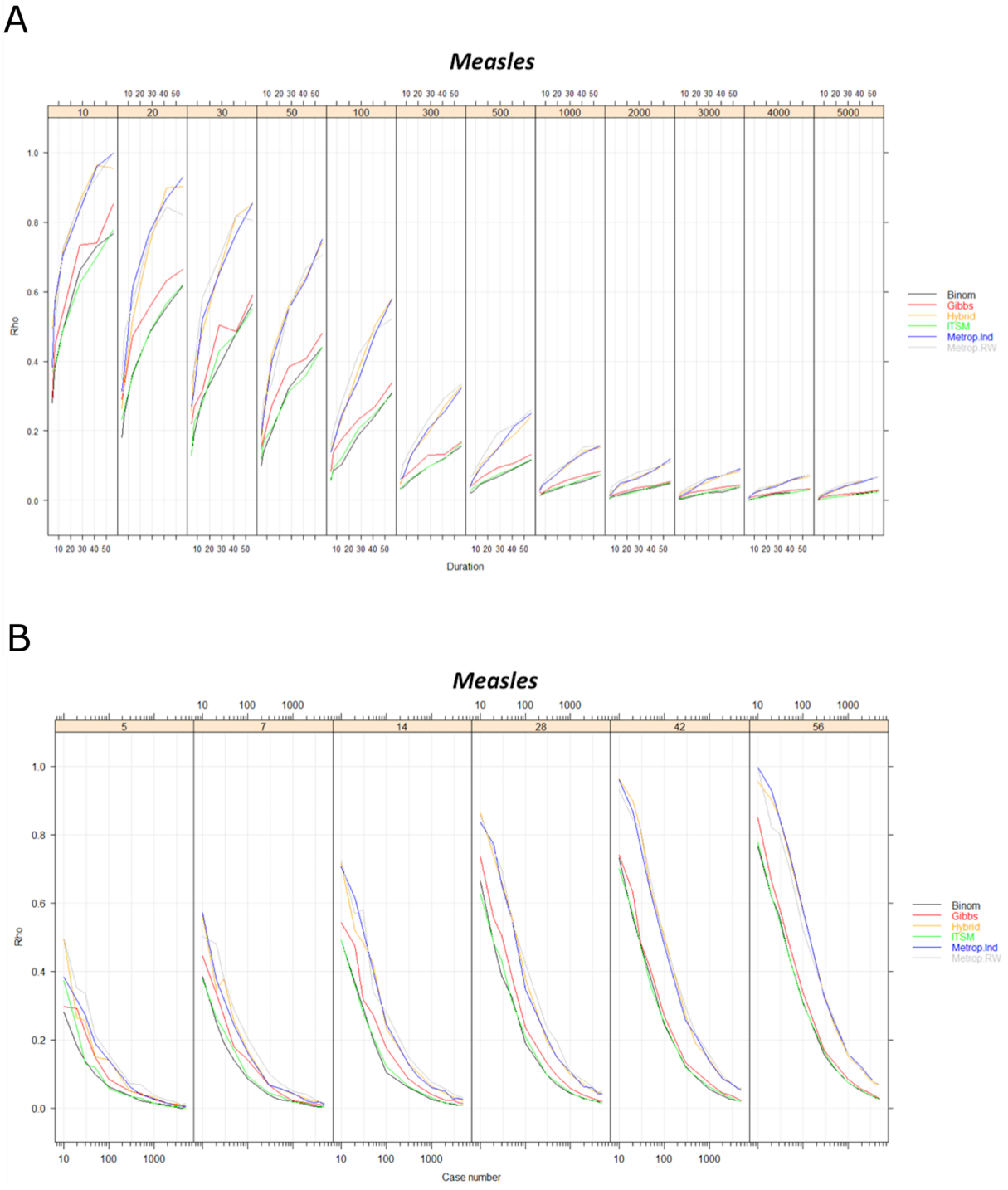


Fig 2. Measles outbreak simulation assessment. (a) according to the outbreak duration (in days). (b) according to the number of simulated cases. Rho: Matusita's Measure used for evaluation / Binom = Binomial method / ITSM = Inverse transform sampling method / Metrop. RW = Metropolis-Hasting Random Walk / Metrop. Ind = Metropolis-Hasting independent sampler / Hybrid = Hybrid Gibbs sampler / Gibbs = Gibbs sampling algorithm.

https://doi.org/10.1371/journal.pone.0183992.g002

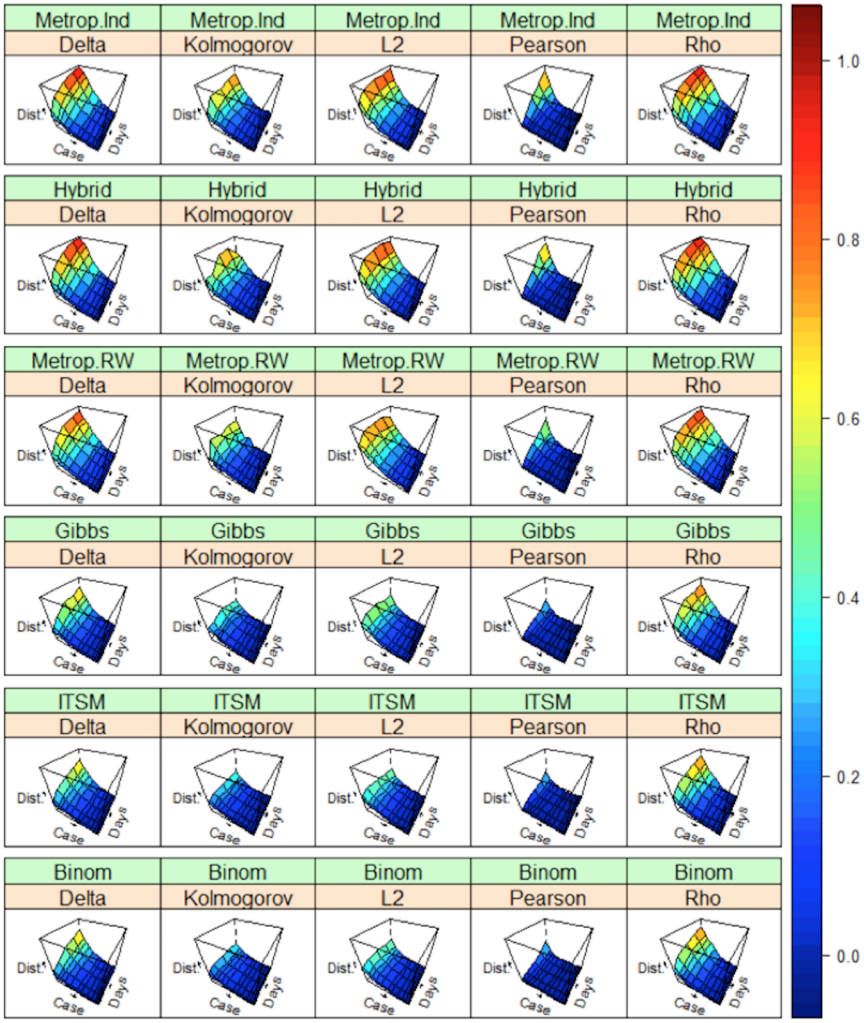


Fig 3. Average quality (between the generated outbreak curves and the corresponding historical epidemic curve) for all diseases considered. Quality was evaluated by different metrics, according to the number of cases and duration. Dist. = normalized distance used for evaluation detailed in pink line; Delta = Overlap Coefficient; Kolmogorov = Kolmogorov distance;; L2 = L2-distance; Pearson = Pearson distance, Rho = Matusita's Measure; Technique used for simulation in green line: Binom = Binomial Method / ITSM = Inverse Transform Sampling Method / Metrop. RW = Metropolis-Hasting Random Walk /Metrop. Ind = Metropolis-Hasting Independent Sampler / Hybrid = Hybrid Gibbs Sampler / Gibbs = Gibbs Sampling Algorithm.

https://doi.org/10.1371/journal.pone.0183992.g003

corresponding historical outbreak curve while taking into account the influence of the number of cases and duration of the outbreak simulated. In Fig 3, for a given GOF, the lower the surface (greens and blues) in the cube, the better the quality of the simulation. We can observe that algorithms based on MCMC are less precise, regardless of the GOF chosen, than those based on independent sampling (Table 1, Fig 3).

The binomial algorithm had the best performance and was fast, since we implemented it using Rbinom, an optimized function based on C language directly implemented in the core package of R.

While generally accurate and fast, the ITSM algorithm was slower (Table 1) than most other algorithms with increasing size or duration of the simulated curve. In theory, the "burn-in" process penalizes all MCMC algorithms, but when the number of cases and days increases,

the study shows that ITSM takes twice as much time as RWMH. In fact, ITSM is slower than other methods [46], especially for a large sample, due to the number of comparisons it requires. However, this method can be used for any type of distribution function, including a mixture of discrete distributions.

Since an outbreak curve may have several local minima/maxima and regions of low probability, a Markov chain may fail to explore all of this complex distribution, except for Gibbs-based methods. For dependent variables, the convergence of Markov chains with any particular distribution is not guaranteed [47].

Gibbs sampling with a shrinkage procedure is designed to manage complex distributions [36]. This two-stage Gibbs sampler ensures the convergence of the Markov chain [36]. The hybrid algorithm and Metropolis group has an acceptance rate of less than 1, thus convergence cannot be ensured or is reached more slowly than for a Gibbs algorithm. This probably explains the difference between Gibbs and the other MCMC algorithms. The accuracy of the Gibbs algorithm was close to the independent (i.i.d.) algorithms, but Gibbs has a slower convergence. In theory, Gibbs produces samples of disease outbreak distributions that are more realistic, reproducing the temporal correlation of data.

The proposal distribution has a strong influence on IMHA performance [36], which is problematic with the complexity of outbreak curves. For example, this algorithm performed best with dengue outbreaks, which have a curve that resembles a uniform distribution.

Random Walk Metropolis-Hastings (RWMH) often appears to be a generic algorithm. But a frequent issue of RWMH is whether or not the random walk should explore all the regions of the distribution to be sampled. Robert [36] explains that RWMH "usually is not necessarily the most efficient choice because it requires many iterations to overcome difficulties (such as low-probability regions of curve) and because of its symmetric features, it spends roughly half the simulation time revisiting regions it has already explored". Our study confirms that RWMH is more efficient for simulating a symmetric jumping distribution, such as that of the norovirus outbreak, than other asymmetric distributions.

The fundamental advantage offered by a Metropolis-within-Gibbs (hybrid) structure is that it breaks down a complex curve into a number of smaller and simpler targets, where local MH algorithms can be tailored at little cost. While these algorithms are very successful for many target distributions, they cannot work efficiently when many wasteful jumps or small moves are generated. The phenomenon is generally encountered in high-dimensional cases or some low-dimensional extreme cases [48]. The hybrid algorithm performed best for dengue and measles outbreaks (outbreaks with non-connected zones), probably because of the ability of this method to explore these types of complex curves.

3.3 Influence of the overall scale factor

The overall scale factor (OSF) describes how outbreak size and duration interact by comparing the original and resulting curves. The results of the OSF study are presented in Fig 4.

When the OSF is below 1 (i.e. when the number of cases per day in the simulation is less than that of the original curve), the accuracy and realism of the simulation, evaluated by GOF metrics, decrease rapidly (significant difference, $p<0,001$). For example, the median loss of accuracy below this threshold is equal to 14.6% for the binomial method (15.7% for ITSM, 16.1% for Gibbs) and 43.8% for Metropolis Independent (42.3% for hybrid, 31.2% for RWMH). Over 1 and for the same algorithms, the loss of accuracy is respectively 0.22% (0.23% for ITSM, 0.29% for Gibbs) and 0.63% (0.64% for hybrid, 1.0% for RWMH). The accuracy is 2 to 3 times better for the first group of algorithms. Moreover, an OSF below 1 is associated with an increase in the variance and duration of the simulation.

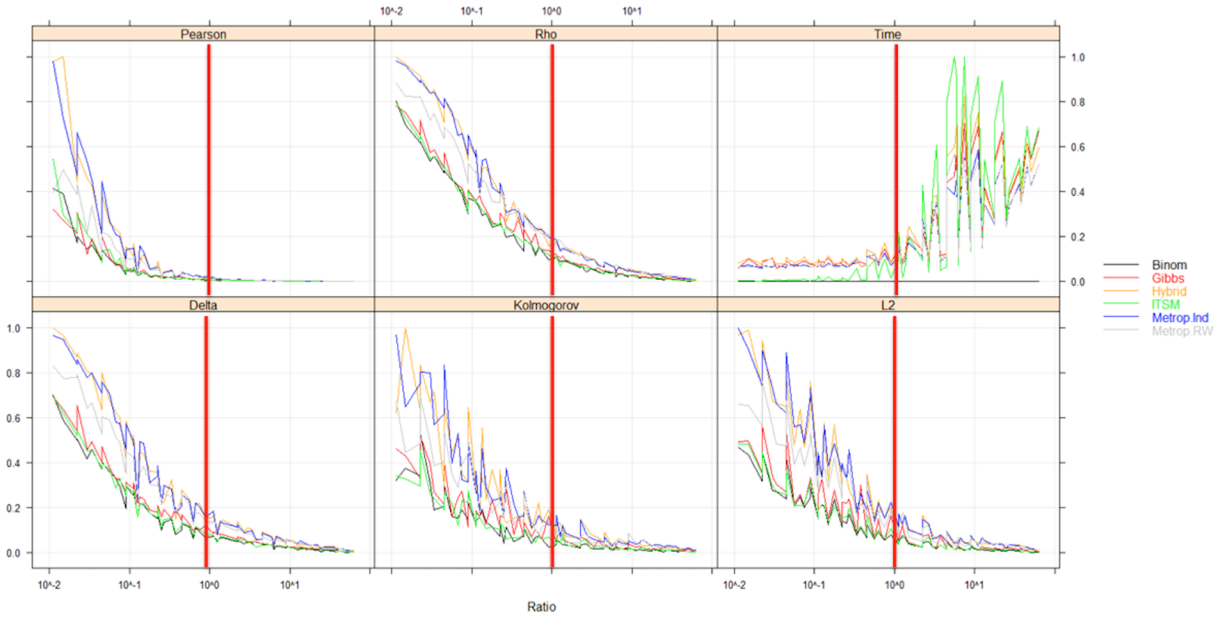


Fig 4. Algorithm simulation accuracy measured by goodness-of-fit measures* and time according to the ratio of simulation. *Delta = Overlap Coefficient, Kolmogorov = Kolmogorov distance; L2 = L2-distance, Pearson = Pearson distance, Rho = Matusita's Measure.

https://doi.org/10.1371/journal.pone.0183992.g004

In fact, an overall scale factor (OSF) below 1 is associated with a significant loss of information during the simulation process. Regardless of the outbreak simulated, OSF = 1 can be considered as a kind of threshold below which the loss of information increasingly impairs simulation quality in terms of goodness-of-fit and variance. As an overall measure of simulation GOF, we used and recommend the χ^2-divergence, which is also a measure of the loss of information during the simulation process [49, 50].

3.4 Final considerations

The purpose of our study is not to evaluate existing algorithms but to propose a simple way to generate realistic and controlled (with regard to the most important characteristics influencing algorithm evaluation) outbreak signals. This preliminary step is required in order to carry out reproducible evaluations expected by system designers/practitioners and to evaluate the influence of the shape and the pathogen on the ODA performance. Our approach follows a pragmatic principle of parsimony by skipping the usual steps of modeling required to produce an outbreak curve (S2 Fig). Indeed, resampling methods are often used as a robust alternative to those based on parametric assumptions when they are dubious, or when parametric inference is impossible or requires very complex formulas for computation. Another advantage of resampling methods resides in the fact that our goal can be achieved without generating and validating a plausible model of transmission and surveillance.

The strength of Monte Carlo resampling methods is their ability to produce realistic and customized curves by rebuilding the original underlying distribution directly from a single observed curve (the epicurve probability distribution), even if the outbreak curve is unusual, provided that the outbreak PMF (probability mass function) can be extracted from real data. Creating a database of outbreak curves (Disease or shape curve/Number of cases/Number of days) supplemented with contextual information (such as population under surveillance/time/location) can provide the material for various simulations and may help to define what

constitutes a realistic and plausible epidemiological curve faced by the surveillance system. In this study, we present only four shapes for a proof of concept, but our database is currently made up of 893 real outbreaks resulting from 58 diseases under surveillance, allowing the generation of many other curves. However, the quality (accuracy) of the choice of curve to support the resampling process during the simulation is clearly the major limitation of the method.

As for all simulation approaches, an evaluation of the epidemiological situation (population, agent, system, location, period, etc.) of the original observed curve/model is required when choosing the one that will support the simulation. Until our proposal, simulation of the complex shape of the measles, dengue or anthrax outbreaks (presented in Fig 1) for evaluating a disease surveillance system and ODA was based on the curve/shape set presented in S1 Fig, which is considered the reference [6, 7, 16–19] for simulating outbreaks with controlled size and duration. The strength of resampling methods is their ability to preserve as much as possible the general shape of real outbreak curves (see an example provided in S3 Fig), which is known to influence the results of ODA evaluations. However, it is clear that one limitation of our study is the difficulty of proving that each produced curve is realistic. Ideally, each curve should be compared to curves observed during real surveillance activity, capturing all specific patterns of the outbreak observed, and not only to the published curve (considered as the reference curve) as in our case.

Our approach can be considered as an easier method than implementing a propagation model or using existing code from compartmental epidemic models, which require numerous real parameters. For example, in the case of the binomial method (the simplest method used in our study), a simulation of an outbreak of 30 cases over 10 days assigns each new case to one of the 10 days according to the binomial distribution law (the PMF extracted from the real curve providing the probability of success on each trial). The corresponding r code is rbinom (number of days, number of cases, probability mass function vector extracted from a real curve) or in our example rbinom(10, 30, PMF). This code is the same, irrespective of the disease outbreak.

4 Conclusion

Given the constraint of keeping the simulation within a range of plausible epidemiological curves faced by the surveillance system, our study confirms that MC resampling methods applied to historical disease outbreaks can be used for outbreak curve production. It also confirms the influence of the algorithm and outbreak duration, size and curve shape on simulation quality. It suggests that Gibbs sampling with a shrinkage procedure is a convenient approach to simulating outbreak curves, since it can be viewed as a fair trade-off between accuracy, simulation speed and data dependency. If dependency is of little importance, binomial and ITSM methods are accurate. This study also shows that χ^2-divergence is a useful tool, not only for evaluating simulation quality but also for quantifying the loss of information observed throughout the simulation process.

Finally, we confirm that the signal generation method proposed in this work must be able to simulate a set of signals for evaluation simply, including very unlikely ones (for testing special or extreme situations). Even if our approach can generate a large spectrum of possible signals, we do not find it necessary to test algorithm performance on all these unlikely signals, but only on the authentic signals faced by the surveillance system.

Supporting information

S1 Script. Function Dayscale (in R language) for homogeneous dilation of number of time steps in the case of a scaling factor r≤1 (number of time steps simulated superior to the

number of time steps in the initial curve).
(DOCX)

S1 Fig. Canonical shapes of simulated outbreaks generally used for ODA evaluation.
(TIF)

S2 Fig. Comparison between different approaches used to generate outbreak curves.
(TIF)

S3 Fig. Example of simulation and evaluation process in the case of a measles outbreak shape [40], for a 14-day duration and 30-case outbreak simulation.
(TIF)

S1 Table. Goodness of fit metrics. (a) Distance and divergence measures used for evaluation. (b) Measure of similarity used for evaluation. with P_i and P_{i0}: Probability Mass Function to compare.
(DOCX)

Acknowledgments

The authors thank Dr. Jean Cleophas ONDO, at ISSEA (Institut Sous-regional de Statistique et d'Economie Appliquee—Yaounde) for his statistical advice and Dr. Howard Burkom (Johns Hopkins University Applied Physics Laboratory) for all his valuable exchanges and input on an earlier version of the manuscript.

Author Contributions

Conceptualization: Gaetan Texier, Michael L. Jackson, Jean-Baptiste Meynard, Herve Chaudet.

Data curation: Gaetan Texier, Michael L. Jackson, Leonel Siwe.

Formal analysis: Gaetan Texier, Michael L. Jackson, Leonel Siwe, Herve Chaudet.

Investigation: Gaetan Texier.

Methodology: Gaetan Texier, Michael L. Jackson, Leonel Siwe, Herve Chaudet.

Project administration: Jean-Baptiste Meynard, Xavier Deparis, Herve Chaudet.

Resources: Leonel Siwe.

Software: Gaetan Texier, Michael L. Jackson, Leonel Siwe, Herve Chaudet.

Supervision: Jean-Baptiste Meynard, Herve Chaudet.

Validation: Gaetan Texier, Michael L. Jackson, Jean-Baptiste Meynard, Herve Chaudet.

Visualization: Gaetan Texier, Michael L. Jackson, Herve Chaudet.

Writing – original draft: Gaetan Texier, Herve Chaudet.

Writing – review & editing: Gaetan Texier, Michael L. Jackson, Leonel Siwe, Jean-Baptiste Meynard, Xavier Deparis, Herve Chaudet.

References

1. Chen H, Zeng D, Yan P. Public Health Syndromic Surveillance Systems. Infectious Disease Informatics: Syndromic Surveillance for Public Health and BioDefense. Boston, MA: Springer US; 2010. p. 9–31.

2. Enki DG, Garthwaite PH, Farrington CP, Noufaily A, Andrews NJ, Charlett A. Comparison of Statistical Algorithms for the Detection of Infectious Disease Outbreaks in Large Multiple Surveillance Systems. PLoS One. 2016; 11(8):e0160759. https://doi.org/10.1371/journal.pone.0160759 PMID: 27513749.
3. Buckeridge DL. Outbreak detection through automated surveillance: a review of the determinants of detection. J Biomed Inform. 2007; 40(4):370–9. https://doi.org/10.1016/j.jbi.2006.09.003 PMID: 17095301.
4. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Med Inform Decis Mak. 2007; 7(1):6. https://doi.org/10.1186/1472-6947-7-6 PMID: 17331250.
5. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. J Biomed Inform. 2005; 38(2):99–113. https://doi.org/10.1016/j.jbi.2004.11.007 PMID: 15797000.
6. Mandl KB, Reis B, Cassa C. Measuring outbreak-detection performance by using controlled feature set simulations. Morb Mortal Wkly Rep. 2004; 53(Suppl) 130–6. https://doi.org/10.1037/e307182005-024
7. Lotze T, Shmueli G, Yahav I. Simulating Multivariate Syndromic Time Series and Outbreak Signatures. 2007. In: Robert H Smith SSRN Electronic Journal [Internet]. http://dx.doi.org/10.2139/ssrn.990020.
8. Buehler JW, Hopkins OJMRS, Sosin DM, Tong V, CDC working group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC working group. MMWR Recomm Rep. 2004; 53 (RR-5): 1–11. PMID: 15129191.
9. Wagner MM, Wallstrom G. Methods for Algorithm Evaluation. Handbook of Biosurveillance. Wagner MM, Moore AW, Aryel RA editors ed. New York: Academic Press; 2006.
10. Giraldo JO, Palacio DH. Deterministic SIR (Susceptible—Infected—Removed) models applied to varicella outbreaks. Epidemiol Infect. 2008; 136(5):679–87. https://doi.org/10.1017/S0950268807009260 PMID: 17655783.
11. Anderson RM, May RM. Infectious Diseases of Humans—Dynamics and Control. Oxford: Oxford University Press; 2005 2005.
12. Mniszewski S, Valle DS, Stroud P, Riese J. EpiSimS simulation of a multi-component strategy for pandemic influenza. In Proceedings of the 2008 Spring simulation multi-conference 2008. p. 556–63.
13. Nsoesie EO, Beckman RJ, Shashaani S, Nagaraj KS, Marathe MV. A Simulation Optimization Approach to Epidemic Forecasting. PLoS ONE. 2013; 8(6): e67164. https://doi.org/10.1371/journal.pone.0067164 PMID: 23826222
14. Lewis B, Eubank S, Abrams AM, Kleinman K. in silico surveillance: evaluating outbreak detection with simulation models. BMC Med Inform Decis Mak. 2013; 13(1):12. https://doi.org/10.1186/1472-6947-13-12 PMID: 23343523.
15. Held L, Hofmann M, Hohle M, Schmid V. A two-component model for counts of infectious diseases. Biostatistics. 2006; 7(3):422–37. https://doi.org/10.1093/biostatistics/kxj016 PMID: 16407470.
16. Reis BY, Pagano M, Mandl KD. Using temporal context to improve biosurveillance. Proc Natl Acad Sci. 2003; 100(4):1961–5. https://doi.org/10.1073/pnas.0335026100 PMID: 12574522.
17. Stoto MA, Schonlau M, Mariano LT. Syndromic Surveillance: Is it Worth the Effort? Chance. 2003; 17 (1):19–24.
18. Burkom HS. Development, adaptation, and assessment of alerting algorithms for biosurveillance. Johns Hopkins APL Technical Digest. 2003; 24(4):335–42.
19. Reis BY, Kohane IS, Mandl KD. An Epidemiological Network Model for Disease Outbreak Detection. PLoS Med. 2007; 4(6):e210. https://doi.org/10.1371/journal.pmed.0040210 PMID: 17593895.
20. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC Med Inform Decis Mak. 2007; 7:6. https://doi.org/10.1186/1472-6947-7-6 PMID: 17331250.
21. Bollen KA, Curran PJ. Latent Curve Models: A Structural Equation Perspective: Ed. John Wiley and Sons; 2006 2006.
22. Baker GA. Factor analysis of relative growth. Growth. 1954; 18:137–43.
23. Tucker LR. Determination of parameters of a functional relation by factor analysis. Psychometrika. 1958; 23(1):19–23. https://doi.org/10.1007/bf02288975
24. Centers for Disease Control and Prevention. Outbreaks of Norwalk-like viral gastroenteritis-Alaska and Wisconsin. Morb Mortal Wkly Rep. 2000; 49(10):207–11. PMID: 10738840.
25. Centers for Disease C, Prevention. Outbreaks of gastroenteritis associated with noroviruses on cruise ships—United States, 2002. MMWR Morb Mortal Wkly Rep. 2002; 51(49):1112–5. PMID: 12530708.
26. Raoult D. Microbe interactions undermine predictions. Science. 2011; 331(6014):144–5; author reply 5–7. https://doi.org/10.1126/science.331.6014.144-c PMID: 21233365.

27. Neuberger A, Paul M, Nizar A, Raoult D. Modelling in infectious diseases: between haphazard and hazard. Clin Microbiol Infect. 2013; 19(11):993–8. https://doi.org/10.1111/1469-0691.12309 PMID: 23879334.

28. Raoult D. Molecular, epidemiological, and clinical complexities of predicting patterns of infectious diseases. Front Microbiol. 2011; 2:25. https://doi.org/10.3389/fmicb.2011.00025 PMID: 21687417.

29. Nishiura H. Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. Emerg Themes Epidemiol. 2007; 4:2. https://doi.org/10.1186/1742-7622-4-2 PMID: 17466070.

30. Horner RD, Samsa G. Criteria for the use of Sartwell's incubation period model to study chronic diseases with uncertain etiology. J Clin Epidemiol. 1992; 45(10):1071–80. PMID: 1474403.

31. Sartwell PE. The incubation period and the dynamics of infectious disease. Am J Epidemiol. 1966; 83 (2):204–6. PMID: 5930773.

32. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: http://www.R-project.org/; 2013.

33. Kachitvichyanukul V, Schmeiser BW. Binomial random variate generation. Communications of the ACM. 1988; 31(2):216–22. https://doi.org/10.1145/42372.42381

34. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. J Chem Physics 1953; 21(6):1087. https://doi.org/10.1063/1.1699114

35. Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika. 1970; 57(1):97. https://doi.org/10.1093/biomet/57.1.97

36. Robert C, Casella G. Introducing Monte Carlo Methods with R. New York: Springer-Verlag; 2010. 284 p. https://doi.org/10.1007/978-1-4419-1576-4

37. Roberts G, Rosenthal JJ. Coupling and Ergodicity of Adaptive Markov Chain Monte Carlo Algorithms. J Appl Prob. 2007; 44:458–75. https://doi.org/10.1239/jap/1183667414

38. Neal R. Slice sampling (with discussion). Ann Statist. 2003; 31(3):705–67.

39. Tierney L. Markov Chains for Exploring Posterior Distributions. Ann Statist. 1994; 22(4):1701–28. https://doi.org/10.1214/aos/1176325750

40. Centers for Disease Control and Prevention. Outbreak of measles among Christian Science students—Missouri and Illinois, 1994. Morb Mortal Wkly Rep. 1994; 43(25):463–5. PMID: 8208236.

41. Meselson M, Guillemin J, Hugh-Jones M, Langmuir A, Popova I, Shelokov A, et al. The Sverdlovsk anthrax outbreak of 1979. Science. 1994; 266(5188):1202–8. PMID: 7973702.

42. Xu G, Dong H, Shi N, Liu S, Zhou A, Cheng Z, et al. An outbreak of dengue virus serotype 1 infection in Cixi, Ningbo, People's Republic of China, 2004, associated with a traveler from Thailand and high density of Aedes albopictus. Am J Trop Med Hyg. 2007; 76(6):1182–8. PMID: 17556633.

43. Birattari M, Dorigo M. How to assess and report the performance of a stochastic algorithm on a benchmark problem: mean or best result on a number of runs? Optim Lett. 2006; 1(3):309–11. https://doi.org/10.1007/s11590-006-0011-8

44. Sung-Hyuk C. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. Int J Math Model Meth Appl Sci. 2007; 4(1):300–7.

45. Karian Z, Dudewicz EJ. Handbook of Fitting Statistical Distributions with R. CRC Press, Taylor and Francis Group ed. Boca Raton: Chapman & Hall; 2011 2012/02/29. https://doi.org/10.1080/09332480.2012.668473

46. Dagpunar JS. Simulation and Monte Carlo with Applications in Finance and MCMC. New York: John Wiley and Sons; 2007 2007.

47. Wilkinson D. Stochastic modeling for system biology. Boca Raton: Chapman and Hall/CRC Press; 2006 2006.

48. Bai Y. An Adaptive Directional Metropolis-within-Gibbs Algorithm. 2009. In: Technical Report [Internet]. http://www.utstat.toronto.edu/wordpress/WSFiles/technicalreports/0903.pdf.

49. Kumar P, Chhina S. A symmetric information divergence measure of the Csiszár's f-divergence class and its bounds. Comput Math Appl. 2005; 49(4):575–88. https://doi.org/10.1016/j.camwa.2004.07.017

50. Reid MD, Williamson RC. Information, Divergence and Risk for Binary Experiments. JMLR. 2011; 12:731–817.