bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Inferring a qualitative contact rate index of uncertain epidemics

Marco Tulio Angulo[1,*] and Jorge X. Velasco-Hernandez[1]

[1] *Institute of Mathematics, Universidad Nacional Autónoma de México (UNAM), Juriquilla 76230, México.*

We will inevitably face new epidemic outbreaks where the mechanisms of transmission are still uncertain, making it difficult to obtain quantitative predictions. Thus we present a novel algorithm that qualitatively predicts the start, relative magnitude and decline of uncertain epidemic outbreaks, requiring to know only a few of its "macroscopic" parameters. The algorithm is based on estimating exactly the time-varying contact rate of a canonical but time-varying Susceptible-Infected-Recovered epidemic model parametrized to the particular outbreak. The algorithm can also be extended to other canonical epidemic models. Even if dynamics of the outbreak deviates significantly from the underlying epidemic model, we show the predictions of the algorithm remain robust. We validated our algorithm using real time-series data of measles, dengue and the current zika outbreak, comparing its performance to existing algorithms that also use a few macroscopic parameters (e.g., those estimating reproductive numbers) and to those using a thorough understanding of the mechanisms of the epidemic outbreak. We show our algorithm can outperform existing algorithms using a few macroscopic parameters, providing an informative qualitative evaluation of the outbreak.

Keywords. Epidemic prediction; contact rate; qualitative; uncertainty.

I. INTRODUCTION

New epidemic outbreaks of influenza, ebola, dengue, zika or other novel diseases are waiting to emerge [1]. For some of these new outbreaks, the mechanisms determining key characteristics such as disease transmission, risk groups, contact rates, vectorial capacities and so on will be uncertain. A thorough understanding of such mechanisms is necessary to obtain quantitative predictions of an outbreak (e.g., its start, magnitude and decline), because they are used by the detailed mathematical models generating the predictions [2–4]. Since these mechanisms can rapidly become very complex due to genetics [5], human mobility [6] and many other factors such as climate change [1], developing accurate quantitative predictions of new outbreaks is a process that can take time —time to collect detailed data of the epidemic outbreak and time to fully understand and mathematically model its mechanisms [7]. For epidemic emergencies or in the unlawful release of infectious agents, time is critical. And thus, before we fully understand the mechanisms of an epidemic outbreak, it is useful to have an early evaluation of the timing and size of the outbreaks —even if the evaluation is just qualitative.

*Electronic address: mangulo@im.unam.mx

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The "macroscopic" dynamics of the majority of past and present epidemic diseases can be reasonable captured by simple canonical epidemic models, such as the Susceptible-Infected-Recovered (SIR) or Susceptible-Exposed-Infected-Recovered (SEIR) ones [8–10]. By contrast to detailed models, these canonical models require only a few "macroscopic" parameters of the outbreak (e.g., birth/mortality and recovery rates of the population). But even with these few parameters, the available data can result in a large uncertainty in their value [10]. This uncertainty might increase further in the case of emergent epidemics, as happened in the recent zika epidemic [11].

Here we propose an algorithm to qualitatively predict the start, relative magnitude and decline of an epidemic outbreak, requiring to know only a few of its "macroscopic" parameters. The algorithm is based on robustly estimating the time-varying contact rate —the fraction of the susceptible population that an average infectious individual successfully infects per day— of a canonical but time-varying SIR (tvSIR) epidemiological model parameterised to the particular outbreak. The contact rate and the related basic/effective reproduction numbers are key quantities for predicting and managing the start, magnitude and end of epidemic outbreaks [8, 12–14]. Hence, a significant number of studies have approached the problem of estimating these quantities from time-series of infection incidence. Most of these efforts focus on estimating constant contact rates and the basic reproductive number [15, 16], with less frequent attempts to estimate time-varying contact rates [17–20]. Some of these estimation methods, such as those in [18, 20], also require few macroscopic parameters of the outbreak. However we show they tend to incorrectly predict, even qualitatively, the start, relative magnitude and decline of an outbreak.

Our algorithm is similar to these last works in that it also relies on incidence time-series data. Indeed, despite being an aggregated measurement of possibly very complex transmission processes —involving immunity, infectiousness, resistance and so on— this is often the first and only available data of an emergent outbreak. We rigorously prove that the proposed algorithm estimates exactly the contact rate in the ideal case when the outbreak dynamics follows exactly a tvSIR model (Section II-A); numerically show its robustness to parametric uncertainties and unmodeled dynamics (Section II-B); and experimentally demonstrate on data of measles, dengue and zika that our algorithm can outperform the qualitative predictions of similar algorithms that only use macroscopic parameters of the outbreak (Section II-C). The algorithm also provides information on how much the actual dynamics of the outbreak differs from its proposed tvSIR model. This allows, for example, detecting if the outbreak is decaying faster than expected by the model. We end the paper offering some concluding remarks and discussing the limitations of the algorithm (Section III).

II. RESULTS

Assume that incidence time-series data of the epidemic is available, reporting the proportion of the infected population $I(t)$ at discrete time instants $t \in \mathbb{T} = \{t_0, \cdots, t_f\}$. Using this data, our qualitative contact rate index takes

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the form

$$\hat{\beta}(t) = \frac{\dot{y}(t) + [\gamma + \mu]y(t)}{[\hat{M}(t) - y(t)][y(t) + \varepsilon]}, \tag{1}$$

where $\hat{M}(t)$ is the solution to the differential equation $\dot{\hat{M}}(t) = \mu - \mu\hat{M}(t) - \gamma y(t)$ with initial condition $\hat{M}(t_0) = 1 - (\gamma/\mu)y(t_0)$. The continuous-time signal $y(t)$ in (1) is obtained by interpolating (and filtering, if needed) the data $\{I(t), t \in \mathbb{T}\}$, and $\dot{y}(t)$ is an estimate of its derivative at time t. Note that (1) is negative when the outbreak develops in a faster time-scale than the one imposed by the parameters of $\hat{M}$ (see Section II-B for details). The parameters $\mu > 0, \gamma > 0$ and $\varepsilon > 0$ represent the rate of birth/mortality, infected recovering rate, and a small migration term of the outbreak, respectively. These three "macroscopic" parameters depend on the disease and population only, and thus can be adjusted according to previous outbreaks of the same disease, or heuristically adjusted using existing data (see Section II-C).

In what follows we rigorously analyse the performance of (1) when the outbreak dynamics follows an ideal tvSIR model; validate its robustness to deviations of the outbreak dynamics from this model; and analyse and compare its performance using real data.

A. Theoretical analysis. Consider that the outbreak dynamics follow a canonical tvSIR model [21] with parameters μ, γ and ε, and a time-varying contact rate $\beta(t)$:

$$\dot{S}(t) = \mu - \mu S(t) - \beta(t)S(t)[I(t) + \varepsilon], \quad \dot{I}(t) = \beta(t)S(t)[I(t) + \varepsilon] - \gamma I(t) - \mu I(t), \quad \dot{R}(t) = \gamma I(t) - \mu R(t), \tag{2}$$

where the population is composed of susceptible S, infected I and recovered R individuals —scaled as proportions of the total population. In this case, using algorithm (1) with $y(t) = I(t)$ guarantees that $\hat{\beta}(t) \to \beta(t)$ as $t \to \infty$ exponentially fast with rate $\mu(t - t_0)$, Fig 1a. In order to prove this claim, let $M = S + I$ and note its time derivative is $\dot{M} = \mu - \mu M - \gamma I$. Hence, using the choice for $\dot{\hat{M}}$ in (1), we obtain

$$\hat{M}(t) - M(t) = [\hat{M}(t_0) - M(t_0)]e^{-\mu(t-t_0)},$$

implying that $\hat{M}(t)$ converges to $M(t)$ exponentially fast with rate $\mu(t - t_0)$. Consequently, the quantity $\hat{S}(t) = \hat{M}(t) - y(t)$ —obtained from data and the numerical solution of (1)— also converges exponentially fast to $S(t)$, regardless of the particular contact rate, Fig.1b. Borrowing the terminology of system and control theory, $\hat{S}(t)$ and the dynamics for $\hat{M}(t)$ are an unknown input observer for (2), see e.g. [22] and [23]. This means that, by measuring $y = I$ only, the S variable of the tvSIR model is detectable, implying that it is impossible to estimate the proportion of susceptible population faster than the exponential rate provided by our algorithm. Finally, using $\hat{S}(t)$, we can obtain the contact rate $\beta(t)$ using the second equation in (2), yielding the formula for $\hat{\beta}$ in (1) and thus proving the claim.

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

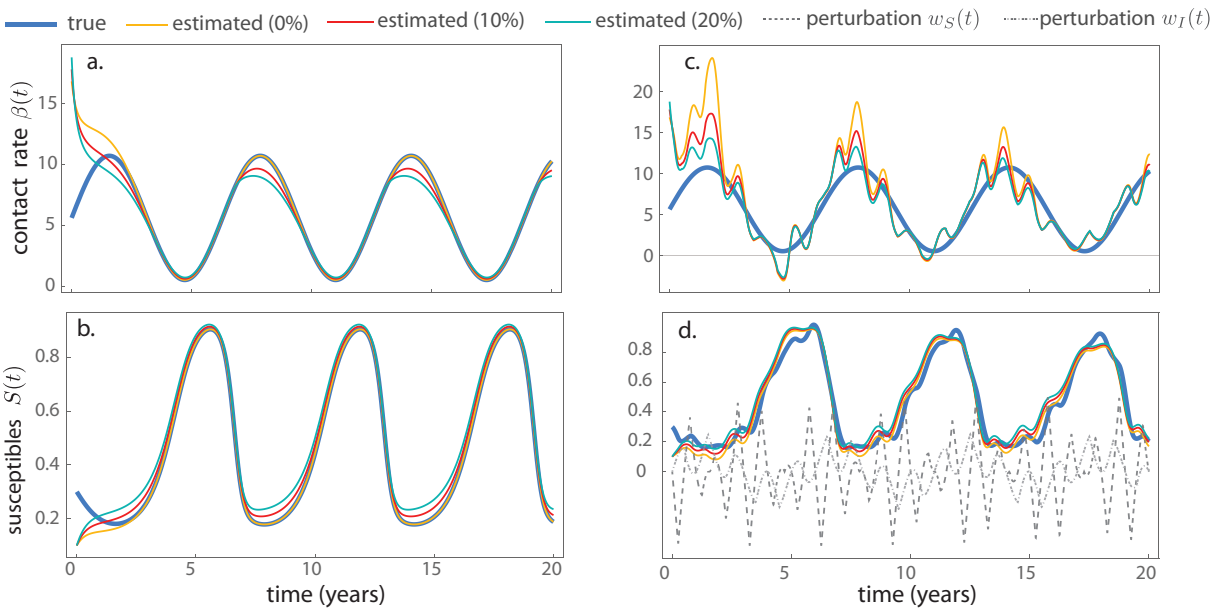

FIG. 1: **Numerical results.** Performance of (1) on the tvSIR model (2) with $\gamma = 0.5, \mu = 1.5, \varepsilon = 10^{-2}$ and $\beta(t) = 5.1(\sin(t) + 1.1)$. **a. and b.** When the proposed algorithm uses exactly the parameters of the tvSIR model, the estimation of $\beta(t)$ and $S(t)$ is exact and exponentially fast. And even when there exists 10% or 20% mismatch between the parameters of the tvSIR model and the ones used by the algorithm (i.e. (1) uses the parameters $1.1\mu, 0.9\gamma, 1.1\varepsilon$ or $1.2\mu, 0.8\gamma, 1.2\varepsilon$), the proposed algorithm provides a robust estimation. **c. and d.** The robustness of the algorithm allows qualitatively tracking the trend of $\beta(t)$ despite the perturbations $w_S(t) = -0.5|\cos(t)|w(t)$ and $w_I(t) = 0.3\sin(\cos(2t))w(0.5t)$ acting on the model (here $w(t)$ is a triangular wave with period 1 varying in $[-1, 1]$).

By using $\hat{\beta}(t)$, we can also construct an index for the effective (or instantaneous) reproductive number:

$$\hat{R}(t) = \frac{\hat{\beta}(t)}{\gamma + \mu}\left[1 - y(t) - \gamma e^{-\mu t}\int_{t_0}^{t} e^{\mu\tau} y(\tau) d\tau\right], \quad t \geq t_0, \tag{3}$$

This quantity preserves the usual property of reproductive numbers: the condition $\hat{R}(t) < 1$ is sufficient for the global asymptotic stability of the infected-free solution $I(t) \equiv 0$ of the tvSIR model (2), see Supplementary Information 1. Note that at the start of an epidemic (i.e., when $I \approx 0$) and assuming a constant contact rate β, the formula (3) reduces to the well-known basic reproductive number $R(0) = \hat{\beta}/(\gamma + \mu)$ for the (time-invariant) SIR model [15]. Additionally, for the experimental results discussed in Section II-C, the approximation $\hat{R}(t) \approx \hat{\beta}(t)/(\gamma + \mu)$ is reasonable, since the magnitude of y is of the order 10^{-3}.

B. Numerical validation of its robustness. No outbreak dynamics can be exactly captured by the tvSIR model (2). Thus, in general, we cannot expect to obtain an exact estimation of the contact rate of the outbreak using (1). Instead, the proposed algorithm must qualitatively follow the contact rate of an epidemic outbreak, even if its dynamics significantly deviate from the tvSIR model due to incomplete information on cases, underreporting, demographic and environmental stochastic effects, etc. This requires robustness of the estimation of $\beta(t)$ to both

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

changes in the parameters of the tvSIR, and the presence of unmodeled dynamics or perturbations in dynamics of the outbreak.

On one hand, by the continuity of Eq. (1) with respect to its parameters, the estimator $\hat{\beta}(t)$ is robust to changes in the parameters γ, μ and ε, see Fig. 1a-b. This means that a small deviation of these parameters from the "true ones" will only produce a small error in the estimation of the contact rate. Numerical experiments suggest that (1) performs better when the birth/mortality rate μ is overestimated and the recovering rate γ is underestimated (Fig. S1).

On the other hand, in order to analyse the performance of the proposed indicator in more general scenarios were the dynamics of the outbreak significantly deviates from tvSIR model, we considered the perturbed tvSIR model

$$\dot{S}(t) = \mu - \mu S(t) - \beta(t)S(t)[I(t)+\varepsilon] + w_S(t), \quad \dot{I}(t) = \beta(t)S(t)[I+\varepsilon] - \gamma I(t) - \mu I(t) + w_I(t), \quad \dot{R}(t) = \gamma I(t) - \mu R(t) + w_R(t),$$

where $w_S(t), w_I(t)$ and $w_R(t)$ are perturbations representing unmodeled mechanisms of infection, demography, vaccination and variability in recovery infection and mortality rates. Even when these perturbations have the same order magnitude as the model (i.e., they have the same order of magnitude as S itself), we find that the estimation $\hat{S}(t)$ of $S(t)$ from (1) remains acceptable and robust to changes in the parameters also (Fig. 1d). In addition, provided that the perturbations occur at a faster time scale than changes in the contact rate, the indicator $\hat{\beta}(t)$ follows closely the trend of the true contact rate, and it is also robust to changes in the parameters of the algorithm (Fig. 1c).

Note that due to the presence of these perturbations, the quantity $\hat{\beta}(t)$ can be negative. This happens when $I(t)$ is decreasing faster than the exponential $e^{-(\gamma+\mu)t}$. Hence, algorithm (1) also measures the discrepancy between the data and the model assumptions. Although this can complicate its interpretation as a true "contact rate", in Section II-C we argue this is compensated by its better predictive properties compared to other standard contact rate approximations.

C. Experimental results. We tested the proposed indicator on time-series data of measles in New York, dengue in Rio de Janeiro and the current zika epidemic in Brazil and Colombia (see Methods for details of the data and parameters of the algorithm). We also compared its performance with recent methods for estimating the time-varying reproductive number [14, 18, 20]. In the first two references, the estimation is based on the quantity

$$\hat{R}'(t) = [I(t+1)/I(t)]^{1/r}, \quad t \in \mathbb{T}, \tag{4}$$

for some exponent r (typically equal to one), together with an additional statistical treatment for improving the estimation. In the last reference [14], the authors use the more sophisticated statistical inference method described in [19]. The methods of the first two references require few "macroscopic" parameters of the outbreak, similar to the proposed algorithm. The method in the third reference [20] thoroughly exploits the detailed mechanisms of the zika epidemics, and thus it is expected to be quantitatively and qualitatively better. The results of the comparison are

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

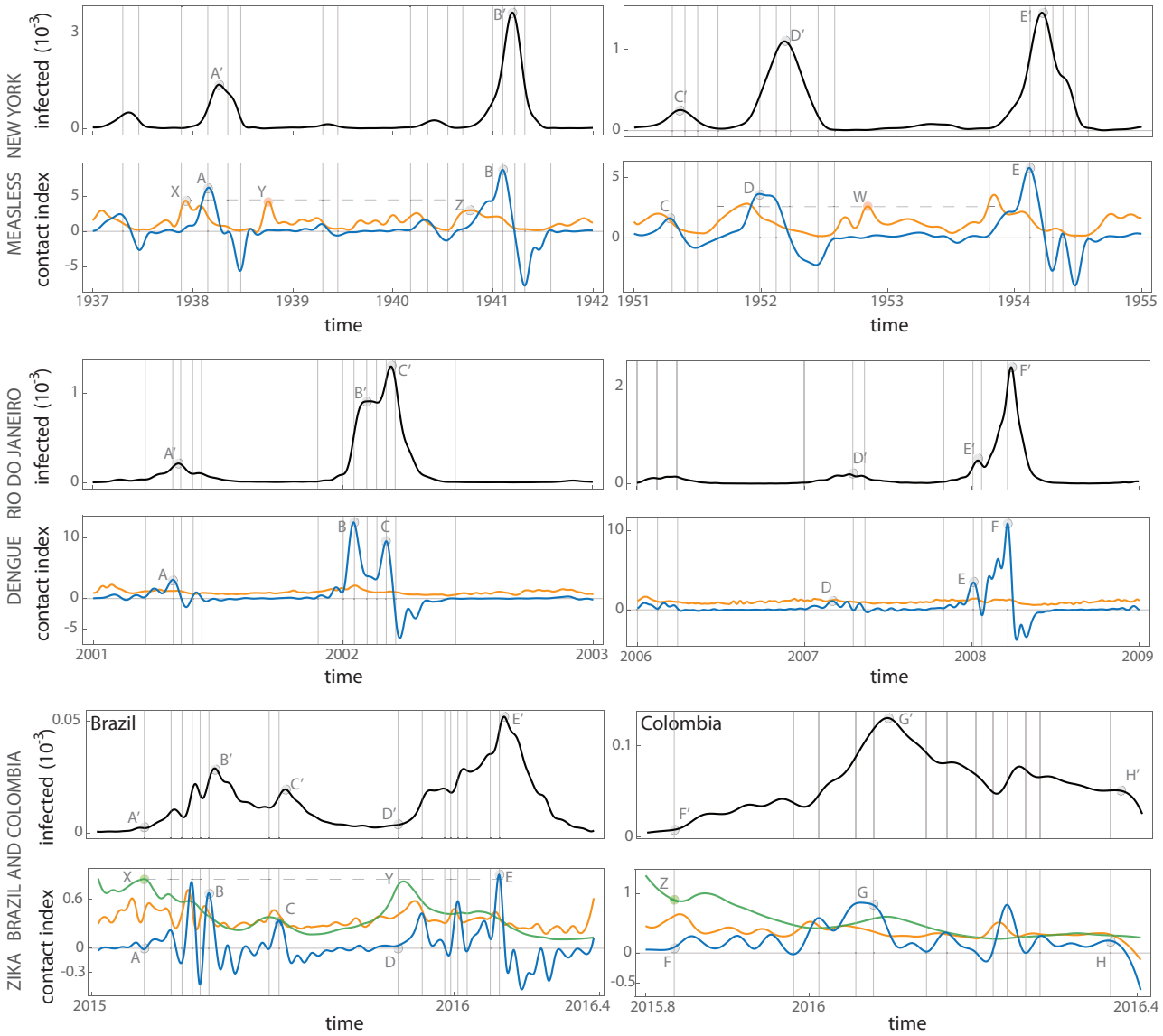

FIG. 2: **Experimental results on reported cases of measles in New York, dengue in Rio do Janeiro (Brazil) and the current zika epidemic in Brazil and Colombia.** Blue: $\hat{\beta}(t)$ in (1). Orange: results from the algorithm for estimating the (time-varying) reproductive number $\hat{R}'(t)$ based in Eq. (4). Green: the (mean) estimation for the time varying reproductive number obtained in [14] for the zika epidemic.

shown in Fig 2.

In the case of measles —with outbreak dynamics reasonably modeled by a SIR model— we observe that $\hat{\beta}(t)$ is able to predict early the maximum of the epidemic outbreaks: points A to E mark months ahead the maxima observed in the data for outbreaks A' to E'. Note that the algorithm (4) predicts earlier than ours those maxima (e.g., the point X precede point A), but this algorithm also predicts non-existent outbreaks (points Y and W, for example). Furthermore, the magnitude of $\hat{R}'(t)$ does not correlate correctly with the magnitude of the outbreak. For instance, point X has larger magnitude than point Z, but the corresponding outbreaks have the oposite trend: A' is way smaller

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

than B'. By contrast, for the proposed algorithm, the larger the magnitude of a peak in $\hat{\beta}(t)$ the larger the outbreak is, showing it correctly predicts the relative magnitude of the outbreaks. For instance, $A < B$ and $A' < B'$.

In the case of dengue, we expect the tvSIR model to be a poor approximation of the outbreak dynamics, since it ignores important mechanisms such as serotype diversity and vector dynamics. Nevertheless, our indicator still qualitatively predicts reasonably well the relative magnitude, start and end of the outbreaks; although its predictive power (e.g., how far ahead can the start or end of the outbreak be predicted) decreases compared to the case of measles. In this case, we also find that the algorithm (4) provides no useful information of the outbreak.

Finally, in the case of zika, for the same reasons as with dengue, we also expect that a tvSIR poorly models its mechanisms. Yet, we observe that our indicator can predict weeks ahead the start (e.g., A and A' or D and D') and the decline (e.g., B and B' or E and E') of the epidemic. We note that its predictive power is similar to the estimation obtained using the algorithm [19] (e.g., points X and A similarly predict the start of the epidemic A'), but our indicator still outperforms the other since it gives a better estimation of the relative magnitude of the outbreak. For example, points X and Y given by the algorithm in [19] suggests that the magnitude of the outbreaks should be very similar in both cases, but in reality $B' < E'$. On the other hand, the predicted magnitude given by our indicator satisfies $B < E$.

The proposed indicator can also be used to obtain a quantitative prediction of the infected population in a given time interval. To do this, the value of $\hat{\beta}(t)$ on a time interval $[t_n, t_n + \varepsilon]$ is estimated from its values on $[t_0, t_n]$ by using e.g. splines, echo-state neural networks or autoregressive discrete-time linear models. Then, the expected proportion of infected population $\hat{I}(t)$, $t \in [t_n, t_n + \varepsilon]$, can be obtained by simulating the tvSIR model (2) over the interval $[t_n, t_n + \varepsilon]$ with initial conditions $\hat{I}(t_n) = I(t_n)$, $\hat{S}(t_n) = \hat{M}(t_n) - I(t_n)$ and contact rate $\hat{\beta}(t)$. We tested this method to predict the proportion of infected population in the big measles outbreak in New York near the year $t_n = 1941.18$, and the dengue outbreak in Rio do Janeiro near the year $t_n = 2002.18$, see Methods for details. Our method provides an adequate prediction for $\varepsilon = 3.6$ months in the case of measles, and $\varepsilon = 18$ days in the case of dengue, Fig. 3. We can also correctly predict the increase and latter decline of the infected population.

III. DISCUSSION AND CONCLUDING REMARKS

By aiming for a qualitative instead of a quantitative prediction, we have proposed an algorithm that only uses macroscopic parameters of the outbreak, based on an elementary tvSIR epidemiological model. The proposed estimator (1) can be interpreted as an indicator closely associated to the true contact rate. The design of the proposed algorithm borrows ideas from the so-called "unknown input observers", which have been very effective algorithms for solving estimation problems in engineering [22, 23]. The main property of the tvSIR allowing the design of such observers is its detectability with respect to the infected population. Since other common but more elaborate epidemi-

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

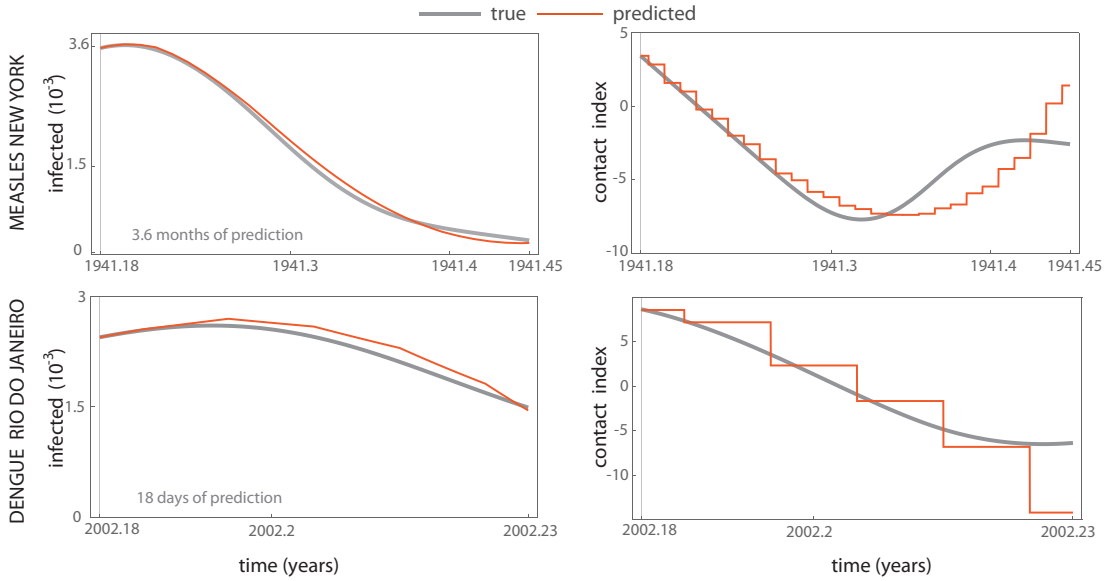

FIG. 3: **Prediction of proportion of infected population for measles in New York and dengue in Rio do Janeiro.** The proposed method provides an acceptable prediction for 22 days in measles, and 2 days in dengue.

ological models such as SEIR, SIS and so on have this same detectability property, Eq. (1) can be straightforwardly modified to replace the tvSIR model by any of those other models. This would allow incorporating more information of the mechanisms of the epidemic outbreak on the qualitative predictions generated by the proposed algorithm. For example, in the case of dengue, it is reasonable that the model includes state variables for each serotype and the mosquito population. We note that despite the success of observers in solving diverse engineering problems, their application to epidemiological models seems rather unexplored. Indeed, we were only able to locate the recent work [24] that considers the classical SIR model. Note also that our algorithm is similar in spirit to [17], in which the authors statistically fit the parameters and initial conditions of a time-invariant SIR model to time-series data in order to estimate the effective reproductive number.

An intrinsic limitation for applying the proposed algorithm is that the data needs to have enough samples to adequately represent the continuous time evolution of the infected population. Rigorously, according to the Nyquist criterion, this require sampling at more than twice the frequency of the bandwidth of this continuous time signal. Note also that for highly noisy data, the estimation will improve by using better differentiation algorithms (e.g., [25] and [26]).

Authors' contributions. M.T.A. conceived the project, did the analysis and numerical calculations. All authors analyzed the results. M.T.A. wrote the manuscript with the help of J.X.V-H.

Acknowledgements. We thank the financial support of CONACyT, México.

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

[1] A.J. McMichael, D.H. Campbell-Lendrum, C.F. Corvaln, and et al. Climate change and human health: Risks and responses. report, World Health Organization, 2003.

[2] Helen J Wearing, Pejman Rohani, and Matt J Keeling. Appropriate models for the management of infectious diseases. PLoS Med, 2(7):e174, 2005.

[3] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. Journal of computational science, 1(3):132–145, 2010.

[4] Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore-Piontti, Natalie E Dean, Diana P Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, et al. Projected spread of zika virus in the americas. bioRxiv, page 066456, 2016.

[5] Erik Carrillo-Valenzo, Rogelio Danis-Lozano, JX Velasco-Hernandez, Gilma Sánchez-Burgos, Celia Alpuche, Irma López, Claudia Rosales, Cécile Baronti, Xavier de Lamballerie, Edward C Holmes, and José Ramos-Castañeda. Evolution of dengue virus in Mexico is characterized by frequent lineage replacement. Archives of Virology, 155(9):1401–12, sep 2010.

[6] Magali Teurlai, Rekol Huy, Bernard Cazelles, Raphaël Duboz, Christophe Baehr, and Sirenda Vong. Can Human Movements Explain Heterogeneous Propagation of Dengue Fever in Cambodia? PLoS Neglected Tropical Diseases, 6(12):e1957, dec 2012.

[7] Neil M Ferguson, Matt J Keeling, W John Edmunds, Raymond Gani, Bryan T Grenfell, Roy M Anderson, and Steve Leach. Planning for smallpox outbreaks. Nature, 425(6959):681–685, 2003.

[8] Roy M Anderson, Robert M May, and B Anderson. Infectious diseases of humans: dynamics and control, volume 28. Wiley Online Library, 1992.

[9] Raymond Gani and Steve Leach. Transmission potential of smallpox in contemporary populations. Nature, 414(6865):748–751, 2001.

[10] Bret D Elderd, Vanja M Dukic, and Greg Dwyer. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. Proceedings of the National Academy of Sciences, 103(42):15693–15697, 2006.

[11] Bogoch II, Brady OJ, Kraemer MU, German M., and et al. Anticipating the international spread of zika virus from brazil. Lancet, 387(10016):335–336, 2016.

[12] Christophe Fraser, Steven Riley, Roy M Anderson, and Neil M Ferguson. Factors that make an infectious disease outbreak controllable. Proceedings of the National Academy of Sciences of the United States of America, 101(16):6146–6151, 2004.

[13] Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. Nature, 442(7101):448–452, 2006.

[14] Neil M. Ferguson, Zulma M. Cucunubá, Ilaria Dorigatti, Gemma L. Nedjati-Gilani, Christl A. Donnelly, Maria-Gloria Basáñez, Pierre Nouvellet, and Justin Lessler. Countering zika in latin america. Science, 2016.

[15] JM Heffernan, RJ Smith, and LM Wahl. Perspectives on the basic reproductive ratio. Journal of the Royal Society Interface, 2(4):281–293, 2005.

[16] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

numbers. Proceedings of the Royal Society of London B: Biological Sciences, 274(1609):599–604, 2007.

[17] Ariel Cintrón-Arias, Carlos Castillo-Chávez, Luis MA Bettencourt, Alun L Lloyd, and HT Banks. The estimation of the effective reproductive number from disease outbreak data. Math Biosci Eng, 6(2):261–282, 2009.

[18] Hiroshi Nishiura, Gerardo Chowell, Hans Heesterbeek, and Jacco Wallinga. The ideal reporting interval for an epidemic to objectively interpret the epidemiological time course. Journal of The Royal Society Interface, page rsif20090153, 2009.

[19] Anne Cori, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. American journal of epidemiology, 178(9):1505–1512, 2013.

[20] Flavio Codeço Coelho and Luiz Max de Carvalho. Estimating the attack ratio of dengue epidemics under time-varying force of infection using aggregated notification data. Scientific reports, 5, 2015.

[21] R Engbert and FR Drepper. Chance and chaos in population biology—models of recurrent epidemics and food chain dynamics. Chaos, Solitons & Fractals, 4(7):1147–1169, 1994.

[22] Jie Chen, Ron J Patton, and Hong-Yue Zhang. Design of unknown input observers and robust fault detection filters. International Journal of Control, 63(1):85–105, 1996.

[23] Denis Dochain. State and parameter estimation in chemical and biochemical processes: a tutorial. Journal of process control, 13(8):801–818, 2003.

[24] Pierre-Alexandre Bliman and Bettina D'Avila Barros. Interval observer for sir epidemic model subject to uncertain seasonality. 2016.

[25] Luma K Vasiljevic and Hassan K Khalil. Error bounds in differentiation of noisy signals by high-gain observers. Systems & Control Letters, 57(10):856–862, 2008.

[26] Emmanuel Cruz-Zavala, Jaime A Moreno, and Leonid M Fridman. Uniform robust exact differentiator. IEEE Transactions on Automatic Control, 56(11):2727–2733, 2011.

IV. METHODS

A Mathematica notebook to replicate and apply the results of the paper to other time-series is available as Supplementary Material.

Data. Infection data for measles was obtained from International Infectious Disease Data Archive at McMaster University*. Data for dengue and Zika was obtained from the supplementary information of [20] and [14], respectively. The proportion of infected population was obtained by dividing the reported number of cases $I(t_k)$ over the total population provided by a census. The interpolation to obtain $y(t)$ was done using fourth order splines.

Used parameters of the algorithm. The parameters of the algorithm were chosen as $\mu = 1.4, \gamma = 1.1$ and $\varepsilon = 10^{-3}$ for measles, $\mu = 3.4, \gamma = 5.1$ and $\varepsilon = 10^{-2.6}$ for dengue and $\mu = 1.4, \gamma = 1.1$ and $\varepsilon = 10^{-3}$ for zika.

Prediction. For prediction of $\hat{\beta}(t)$ over the interval $[t_n, t_n + \varepsilon]$, we used the function `TimeSeriesModelFit` provided

* `https://ms.mcmaster.ca/ bolker/measdata.html`

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in the Mathematica software to build a discrete-time linear model of the contact index. In the case of measles, we used the value $\hat{\beta}(t)$ from 1940 to $t_n = 1941.18$ (sampled at rate 0.0001) as training data. The result was the ARIMA discrete-time model `ARIMAProcess`$[8.47663\times10^{-13}, \{0.0908973\}, 4, \{0.784174\}, 5.06941\times10^{-19}]$. In the case of dengue, the training data $\hat{\beta}(t)$ was from 2000.7 to $t_n = 2002.18$ with sampling rate 0.0001 years. The result was the ARIMA discrete-time model `ARIMAProcess`$[-8.94849 \times 10^{-11}, \{1.09715, -0.102606\}, 3, \{0.518601\}, 9.91838 \times 10^{-16}]$.

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Inferring a qualitative contact rate index of uncertain epidemics —Supplementary Information—

S1. OBTAINING AN EFFECTIVE REPRODUCTIVE NUMBER FROM THE PROPOSED INDICATOR

The effective reproductive number characterizes a sufficient condition for the end of the epidemic. Hence, the objective in this section is to derive a sufficient condition for the stability of the infected-free solution $I(t) \equiv 0$ in the tvSIR model (2). By the continuous dependence of the the solutions of this model on its parameters, we consider the limit case $\varepsilon = 0$ only.

Since the differential equation for $M = S + I$ is linear

$$\dot{M}(t) = -\mu M(t) + \mu - \gamma I(t), \quad M(0) = M_0,$$

we can compute its solution as

$$M(t) = M_0 e^{-\mu(t-t_0)} + e^{-\mu t} \int_{t_0}^{t} e^{\mu\tau}[\mu - \gamma I(\tau)]d\tau = (M_0 - 1)e^{-\mu(t-t_0)} + 1 - \gamma e^{-\mu t} \int_{t_0}^{t} e^{\mu\tau} I(\tau)d\tau.$$

Then, since $S = M - I$ we obtain

$$S(t) = (M_0 - 1)e^{-\mu(t-t_0)} + 1 - \gamma e^{-\mu t} \int_{t_0}^{t} e^{\mu\tau} I(\tau)d\tau - I(t).$$

Substituting this expression in (2) means we can rewrite the differential equation for the infected population as

$$\dot{I}(t) = \beta(t) \left[(M_0 - 1)e^{-\mu(t-t_0)} + 1 - \gamma e^{-\mu t} \int_{t_0}^{t} e^{\mu\tau} I(\tau)d\tau - I(t) \right] I(t) - (\gamma + \mu)I(t).$$

Since $\mu > 0$, the term $(M_0 - 1)e^{-\mu(t-t_0)} \to 0$ as $t \to \infty$ exponentially fast. By definition of stability, only the large time behavior of the system is considered, and hence we obtain the approximation

$$\dot{I}(t) \approx [\beta(t) - \gamma - \mu]I(t) - \beta I^2(t) - \gamma\beta(t)e^{\mu t} I(t) \int_{t_0}^{t} e^{\mu t} I(\tau)d\tau. \tag{S1}$$

A sufficient condition for stability of the solution $I(t) \equiv 0$ is: $I > 0$ implies $\dot{I} < 0$. The condition $\dot{I} < 0$ in (S1) is equivalent to

$$\beta(t) \left[1 - I(t) - \gamma e^{-\mu t} \int_{t_0}^{t} e^{\mu\tau} I(\tau)d\tau \right] < \gamma + \mu,$$

which can be rewritten as $R(t) < 1$ with

$$R(t) = \frac{\beta(t)}{\gamma + \mu} \left[1 - I(t) - \gamma e^{-\mu t} \int_{t_0}^{t} e^{\mu t} I(\tau)d\tau \right].$$

Replacing β by $\hat{\beta}$ obtained in (1) leads to equation (3). Furthermore, since $\hat{\beta}(t) \to \beta(t)$ exponentially fast, then $\hat{R}(t) \to R(t)$ exponentially fast also.

Remark 1. A less conservative condition for stability could be obtained by applying the theory of Lyapunov stability (e.g., Lyapunov functionals) to the Volterra differential equation (S1). The analysis here is equivalent to using the Lyapunov function $V(I) = \frac{1}{2}I^2$ to characterize stability.

bioRxiv preprint doi: https://doi.org/10.1101/107474; this version posted February 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

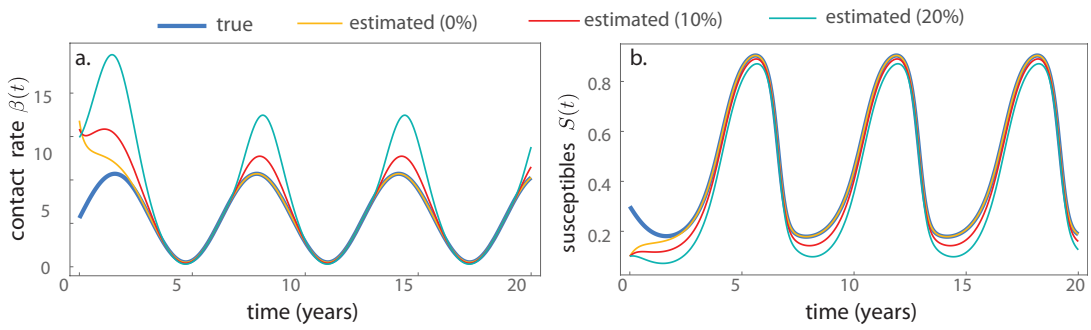

FIG. S1: **Simulation results of the proposed algorithm for the SIR model.** Comparison of the performance of (1) on the SIR model (2) using the exact parameters $\gamma = 0.5, \mu = 1.5, \varepsilon = 10^{-2}$, and introducing a 10% error (i.e. (1) uses the parameters $0.9\mu, 1.1\gamma, 0.9\varepsilon$ instead of γ, μ, ε) and 20% error (i.e. $0.8\mu, 1.2\gamma, 0.8\varepsilon$) in these parameters.