

# Bayesian sequential data assimilation for COVID–19 forecasting

Maria L. Daza–Torres<sup>\*§</sup>    Marcos A. Capistrán<sup>†</sup>    Antonio Capella<sup>‡</sup>  
 J. Andrés Christen<sup>†</sup>

March 11, 2021

## Abstract

We introduce a Bayesian sequential data assimilation method for COVID–19 forecasting. It is assumed that suitable transmission, epidemic and observation models are available and previously validated and the transmission and epidemic models are coded into a dynamical system. The observation model depends on the dynamical system state variables and parameters, and is cast as a likelihood function. We elicit prior distributions of the effective population size, the dynamical system initial conditions and infectious contact rate, and use Markov Chain Monte Carlo sampling to make inference and prediction of quantities of interest (QoI) at the onset of the epidemic outbreak. The forecast is sequentially updated over a sliding window of epidemic records as new data becomes available. Prior distributions for the state variables at the new forecasting time are assembled using the dynamical system, calibrated for the previous forecast. Moreover, changes in the contact rate and effective population size are naturally introduced through auto-regressive models on the corresponding parameters. We show our forecasting method’s performance using a SEIR type model and COVID–19 data from several Mexican localities.

## 1 Introduction

In this paper we introduce a Bayesian sequential data assimilation method for COVID–19 forecasting. Reliable model–based COVID–19 forecasting should be helpful to assist decision–making and planning for healthcare authorities. Consequently, the contributions of this work are aimed at building confidence in our methodology.

Compartmental epidemic models have proven to be adequate to assimilate epidemic data and making forecasts [1, 2]. However, epidemic outbreak predictability is limited due to the influence of human behavior, incomplete knowledge of the virus’s evolution, and weather [3,4], as well as delay and under–reporting of new cases and deaths [5,6]. A practical compromise

---

<sup>\*</sup>Department of Public Health Sciences, University of California, Davis, United States. *mdazatorres@cimat.mx*

<sup>†</sup>Centro de Investigación en Matemáticas (CIMAT), Jalisco S/N, Valenciana, Guanajuato, 36023, México. *marcos, jac at cimat.mx*

<sup>‡</sup>Instituto de Matemáticas, UNAM, Circuito Exterior, CU, CDMX, México *capella@im.unam.mx*

<sup>§</sup>Corresponding author

is to make probabilistic epidemic forecasts a few weeks ahead of time [7, 8] using a model that accounts explicitly for data delay and under-reporting.

In this paper we assume that transmission, epidemic and observation models are properly postulated, previously validated and available. The transmission and epidemic models are coded into a dynamical system. The observation model depends on the dynamical system state variables and parameters, and is cast as a likelihood function. In Section 3 we use a SEIR type epidemic model with Erlang [9] residence times in the exposed and infected compartments. We elicit prior distributions of the effective susceptible population size, dynamical system initial conditions, and infectious contact rates and use Markov Chain Monte Carlo to make inference and prediction of quantities of interest (QoI), such as hospital occupancy, at the onset of the epidemic outbreak in a metropolitan area. Namely, at the time when community transmission starts in the metropolitan area being analyzed. As new data becomes available, we update the forecast sequentially over a sliding window of epidemic records. We assemble prior distributions for the state variables at the new forecasting time using the dynamical system calibrated for the previous forecast. Moreover, we introduce changes in the contact rate and effective population size naturally through auto-regressive models on the corresponding parameters. We argue that this is a natural approach to data assimilation with an epidemic model.

We show our forecasting method's performance using a SEIR type model and COVID-19 data from several Mexican localities.

## 1.1 Related work

Real time epidemic forecasting is an emerging research field [10]. Many forecast modeling efforts study how to address data under-reporting and delays [8, 11]. Other efforts are directed at exploring what sources of information can be incorporated as covariates to make better forecasts. Mcgough *et al.* [12] incorporate traditional surveillance with social media data to forecast Zika in Latin America. The RAPIDD ebola forecasting challenge [13] explored how to integrate different sources of data for Ebola forecasting. Hii *et al.* [14] use temperature and rainfall to forecast dengue incidence.

In a related work, [15] present a COVID19 prediction model. Using a SEIR type dynamical model, and including hospital dynamics and Erlang compartments [9] to properly model residence times, [15] model and predict the COVID19 epidemic in the Mexican 32 states and several metropolitan areas, from the epidemic onset in Mexico in March 2020 (and until February 2021, see <https://coronavirus.conacyt.mx/proyectos/ama.html> (in Spanish), model *ama2*). However, fitting the whole of the epidemic, to infer initial state values, for an epidemic lasting several months, ceases to be useful and adds to the numerical complexity and reliability of the system. In fact, given the generation interval of COVID19, data beyond one month in the past should start to have less importance for current nowcasting and predictions.

## 1.2 Contributions and limitations

The probabilistic forecasting method introduced in this paper allows us to forecast the incidence of new cases and deaths one to four weeks in advance

Once we are near or after a local incidence maximum, our forecasting method disentangles the role of infectious contact rate and effective population size. Other quantities of interest such as hospital occupancy can be calculated as a byproduct of the forecast using suitable

renewal equations. More general data analysis, e.g. by age groups, is not presented in this work. However our results may be applicable on those cases, provided suitable transmission and epidemic models are available.

This manuscript is organized as follows. In Section 2 we make a summary of the modeling decisions taken to implement our forecasting method. In Section 3 we apply our method to COVID–19 epidemic data. Finally, in Section 4 we present the analysis of the Mexico City data. Other examples are provided in the supplementary material.

## 2 Bayesian Sequential Forecasting Method

Let us assume that community transmission starts at time  $t = t_0$  at the metropolitan area where the outbreak is being analyzed. Set  $k = 0$  and denote by  $[t_k, t_k + L]$  the learning period. Namely, the period when we collect epidemic records  $z^{(k)}$  to create a forecast. In the example presented in Section 3, these epidemic records are new hospital admittances and deaths. The delay period is  $[t_k + L, t_k + L + D]$ , i.e. the period when epidemic records are not mature and may include delays in reporting. The forecasting day is  $t_k + L + D$ . We refer to  $[t_k + L + D, t_k + L + D + F]$  as the forecasting period, and  $[t_k, t_k + L + D + F]$  is the forecasting window as illustrated in Figure 1.

Let  $x(t) = (S(t), E(t), I(t), \dots)^T$  denote the time-dependent vector of state variables. We shall assume that the epidemic and transmission models are posed as an initial value problem for a nonlinear system of ordinary differential equations

$$\begin{aligned} \dot{x}(t) &= f(x(t), \theta_k) \\ x(t_k) &= x_k, \end{aligned} \tag{1}$$

where  $t_k$  and  $x_k$  denote respectively the initial time and state in the forecasting window  $[t_k, t_k + L + D + F]$ , and  $\theta_k$  is a vector of model parameters (e.g. contact rate  $\beta$ , effective population size  $\omega$ , etc.) used to calibrate model (1). We shall denote  $p^{(k)} = (x_k, \theta_k)$  the joint vector of initial conditions and model parameters to be inferred.

If  $k = 0$ , we postulate a prior distribution  $\pi_{P^{(k)}}(p^{(k)})$ , a likelihood  $\pi_{Z^{(k)}|P^{(k)}}(z^{(k)}|p^{(k)})$  and use equation (1) and samples obtained through Markov Chain Monte Carlo of the corresponding posterior distribution  $\pi_{P|Z}(p^{(k)}|z^{(k)})$  to make a probabilistic prediction of  $x(t)$  in the forecasting period  $t \in [t_k + L + D, t_k + L + D + F]$ . Afterwards, we update the forecasting window by setting  $t_{k+1} = t_k + n$ , where  $n$  is the number of days until the next forecast. We assemble a new prior distribution  $\pi_{P^{(k+1)}}(p^{(k+1)})$  for the model parameters  $p^{k+1}$  in the new forecasting window  $[t_{k+1}, t_{k+1} + L + D + F]$  using the predicted values of  $x(t)$  at  $t = t_{k+1}$  obtained with equation (1) and samples of the posterior distribution  $\pi_{P^{(k)}|Z^{(k)}}(p^{(k)}|z^{(k)})$  of the previous forecast. Model parameters  $\theta_{k+1}$  have an autoregressive prior distribution in terms of  $\theta_k$ . Finally, we set  $k \leftarrow k + 1$  and repeat the above process to create a new forecast.

The Bayesian sequential data assimilation method consists of three parts; a dynamical system that codes the transmission and an epidemiological model, a probabilistic model for the observed incident cases and deaths, and an informed prior distribution for the parameter space in each forecasting period. In Section 3, we show how to postulate each model component for a forecasting model of covid-19 using data from several Mexico localities.

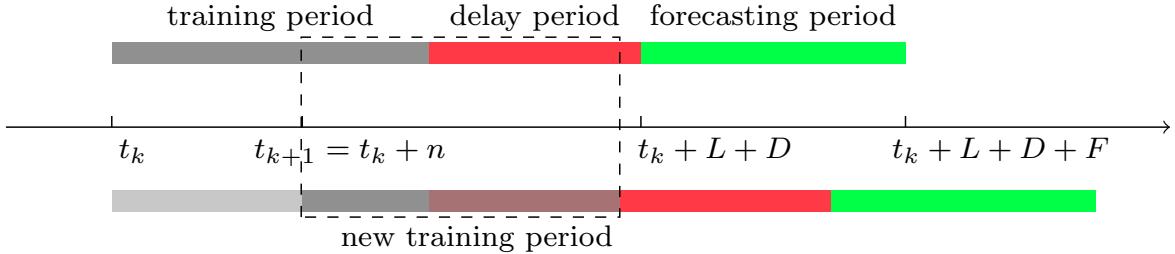


Figure 1: **Bayesian Sequential data assimilation.** We propose a Bayesian filtering method that predicts along the dynamical system (1) evaluated in sample points of the posterior distribution  $\pi_{P^{(k)}|Z^{(k)}}(p^{(k)}|z^{(k)})$  in the current forecasting window  $[t_k, t_k + L + D + F]$ .

### 3 Example: A SEIR type model

#### 3.1 Dynamical model

We consider a variation on the SEIRD epidemic model for susceptible, exposed, infectious, removed, and dead individuals. We have added a compartment for unobserved infectious individuals.

We assume that the total population of the metropolitan area being analyzed is  $N$ . We assume further that there is only a small number of infected individuals at the onset of community transmission. Susceptible individuals  $S$  become exposed  $E$  with force of infection  $\lambda$ . The transmission model is coded into  $\lambda$  as follows, we assume that only unobserved ( $U$ ) and observed ( $O$ ) infectious individuals spread the infection, that is

$$\lambda = \frac{(U + kO)\beta}{N},$$

where  $\beta$  is the infectious contact rate. We have assumed that the contact rate for observed infectious is a factor ( $k$ ) of the contact rate for unobserved infectious. A fraction  $f$  of exposed individuals proceeds to the observed infected class ( $O$ ) at rate  $\sigma_1$ , while the remainder goes directly to an unobserved infective stage ( $U$ ), also at rate  $\sigma_1$ . Individuals leave the infectious class at rate  $\sigma_2$ , with a fraction  $1 - g$  recovering and going to the removed class ( $R$ ) and the remainder ( $g$ ) dying of infection. Unobserved go the removed stage at rate  $\gamma$ . We split the  $E$ ,  $I$ , and  $O$  compartments into two sub-compartments to model residence rates explicitly as Erlang distributions [9], see Table 1.

The dynamics of the epidemic process is governed by the following nonlinear system of ordinary differential equations

$$\begin{aligned}\dot{S} &= -\lambda S \\ \dot{E} &= \lambda S - \sigma_1 E \\ \dot{O} &= f\sigma_1 E - \sigma_2 O \\ \dot{U} &= (1 - f)\sigma_1 E - \gamma U \\ \dot{R} &= (1 - g)\sigma_2 O + \gamma U \\ \dot{D} &= g\sigma_2 O,\end{aligned}$$

---

**Algorithm 1:** Bayesian sequential data assimilation for COVID-19 forecasting

---

Input. Initial time ( $t_0$ ), data ( $z^{(k)}$ ) for  $k = 0, 1, \dots$ , learning period size size ( $L$ ), number of days to forecast ( $F$ ), number of days to move the forecasting window  $n$ , and number of the delays days ( $D$ ),

Output. 

- Posterior distribution  $\pi_{P^{(k)}|Z^{(k)}}(p^{(k)}|z^{(k)})$  for  $k = 0, 1, \dots$
- Prediction of QoI, e.g. hospital occupancy, report of new cases, etc, in the forecasting period  $[t_k + L + D, t_k + L + D + F]$  for  $k = 0, 1, \dots$

Step 1. If  $k = 0$ :

    Set a prior distribution for  $p^{(k)}$ ,  $\pi_{P^{(k)}}(p^{(k)})$

    Go to Step 3

else:

    Go to Step 2

Step 2. For  $k > 0$ , the prior distribution for  $p^{(k)} = (x_k, \theta_k)$  is set using the MCMC output of the period  $k - 1$ :

- For the  $k$ -initial state ( $x_k$ ), the MCMC output of the state variable at time  $t_0 + nk$  is fitted a known distribution.
- For the model parameters  $\theta_k$ , the MCMC output of  $\theta_{k-1}$  is fitted a known distribution.

Step 3. Compute posterior distribution,  $\pi_{Z^{(k)}|P^{(k)}}(z^{(k)}|p^{(k)})$

Step 4. Forecast QoI up to time  $t = t_0 + nk + L + D + F$ ;

Step 5. Save the MCMC output for the next forecasting time.

---

with initial conditions;  $E(0) = E_0$ ,  $O(0) = O_0$ ,  $U(0) = U_0$ ,  $R_0 = R(0)$ ,  $D_0 = D(0)$ , and  $S(0) = N - E_0 - O_0 - U_0 - R_0 - D_0$ . Here  $N = S + E + O + U + R + D$ .

A flow diagram for the model is shown in Figure 2.

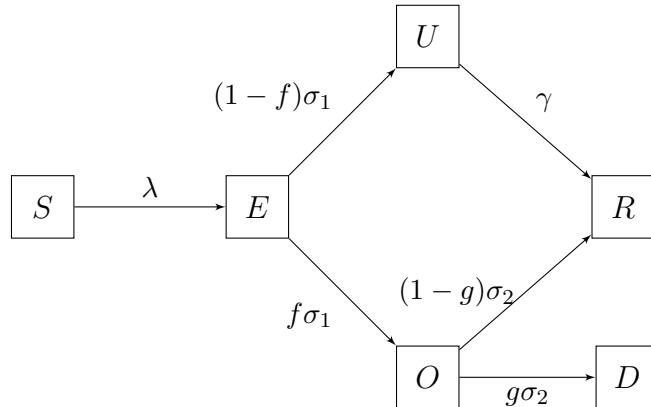


Figure 2: A SEIR type model that into account both observed and unobserved infections.

Table 1: Average times and Erlang shape parameters for sub-compartments.

Variable	Rates	Average time	Erlang shape	Reference
$S$	$\beta$	Inferred	1	
$E$	$1/\sigma_1$	5 days	2	[16, 17]
$A$	$1/\gamma$	7 days	2	[18]
$I$	$1/\sigma_2$	14 days	2	[19, 20]

### 3.2 Model parameters

The model has two kinds of parameters that have to be calibrated or inferred; the ones related to COVID-19 disease (such as residence times and proportions of individuals that split at each bifurcation of the model) and those associated with the public response to mitigation measures such as the contact rate Beta and the proportion of effective population size during the outbreak ( $\omega$ ). Some of these parameters can be found in recent literature (see Table 1) or inferred from reported cases and deaths, but some remain mostly unknown. In the latter category, we have the fraction  $1 - f$  of unobserved infections. We assume  $1 - f = 0.2$ , which means that 80% of cases of symptomatic/asymptomatic infectious go unreported.

### 3.3 Observational model and data

The observed data used to fit the model is based on time series of incident confirmed cases and deaths. We consider daily deaths counts  $d_i$  and its theoretical expectation that is estimated in terms of the dynamical model as

$$\mu_D(t_i) = D(t_i) - D(t_{i-1}).$$

Analogously, we consider daily case  $c_i$  and its corresponding  $\mu_c(t_i)$  given by the daily flux entering the  $O$  compartment [15], namely

$$\mu_c(t_i) = \int_{t_{i-1}}^{t_i} f\sigma_1 E_2(t) dt$$

where  $E_2(t)$  is the last state variable in the  $E$  Erlang series. We calculate the above integral using a simple trapezoidal rule with 10 points.

### 3.4 Estimating model parameters with MCMC

We consider daily confirmed cases  $c_i$  of patients with a positive test ( $O$ ) and daily report deaths  $d_i$ , for the area being analyzed. To account for over dispersed counts, we use a negative binomial (NB) distribution  $NB(\mu, \omega, \theta)$  with mean  $\mu$  and over dispersion parameters  $\theta$  and  $\omega$ . For data  $y_i$ , we let

$$y_i \sim NB(p\mu(t_i), \omega, \theta),$$

with fixed values for the over dispersion parameters  $\omega, \theta$  and an additional reporting probability  $p$ . We assume conditional independence in the data and therefore from the NB model we obtain a likelihood.

The parameters to infer are the contact rate ( $\beta$ ), the proportion of the effective population ( $\omega$ ), the fraction of infected dying ( $g$ ), and crucially we also infer the initial conditions for  $E(0)$ ,  $O(0)$ ,  $U(0)$ ,  $R(0)$ ,  $D(0)$ . Letting  $S(0) = \omega \cdot N - (E(0) + O(0) + U(0) + R(0))$ . We have all initial conditions defined and the model can be solved numerically to obtain  $\mu_D$  and  $\mu_c$  to evaluate our likelihood.

Regarding the elicitation of the parameters' prior distribution for the first forecast, we use Gamma distributions for the initial conditions  $E_0$ ,  $O_0$ , and  $U_0$ , with scale 1 and shape parameter 10. This for modeling the low, near to 10, and close to zero counts for the number of initial infectious conditions. For the initial conditions  $R_0$  and  $D_0$ , we also use Gamma distributions with scale and shape parameters equal to 1. This because at the beginning of the outbreak, both parameters are close to zero. The prior distributions for the remaining parameters are summarized in Table 2.

To sample from the posterior, we resort to MCMC using the t-walk generic sampler [21]. The MCMC runs semi-automatic, with consistent performances in most data sets.

Table 2: Parameters and prior distributions used for Bayesian inference.

Parameter	Prior distribution
Contact rate ( $\beta$ )	$LogNorm(1, 1)$
Fraction of infected dying ( $g$ )	$Beta(1 + 1/6, 1 + 1/3)$
Proportion of the effective population ( $\omega$ )	$Beta(1 + 1/6, 1 + 1/3)$

## 4 Results

This section displays the results to apply Algorithm 1 with covid-19 data set from Mexico's city metropolitan area. Further examples of other Mexican metropolitan areas can be found in the supplementary material. The data set reports the incident number of confirmed cases and deaths for each location at a daily frequency starting in early 2020.

The uncertainty in the early forecast is high because we do not yet know the effective size of the population participating in the epidemic. Further, these forecasts are prone to further errors given the uncertainty in disease spread parameters and the initial state of the disease. We use the Bayesian Sequential Forecasting Method to predict trajectories, given weekly updates. The model starts with inaccurately predicted trajectories, where the median of the trajectories overestimate the future data (See Figure 3). Further, the prediction has a high initial cone of uncertainty.

The next predictions using previous information and update data quickly increase accuracy, i.e., the median of the forecast becomes close to the future data, and the cone of uncertainty shrinks, Figures 4-5. These results point to opportunities in capturing behavioral changes during sequential forecasts and predicting in real-time the future trajectory of the disease after sufficient observations. Figures 4-6 show how the model captures the tendency changes in the pandemic dynamic. Also, changes in the parameters associated with the public health measures such as the contact rates ( $\beta$ ), the proportion of effective population size during the outbreak ( $\omega$ ), and the proportion of observed infectious dying ( $g$ ) are exhibited in Figure 7.

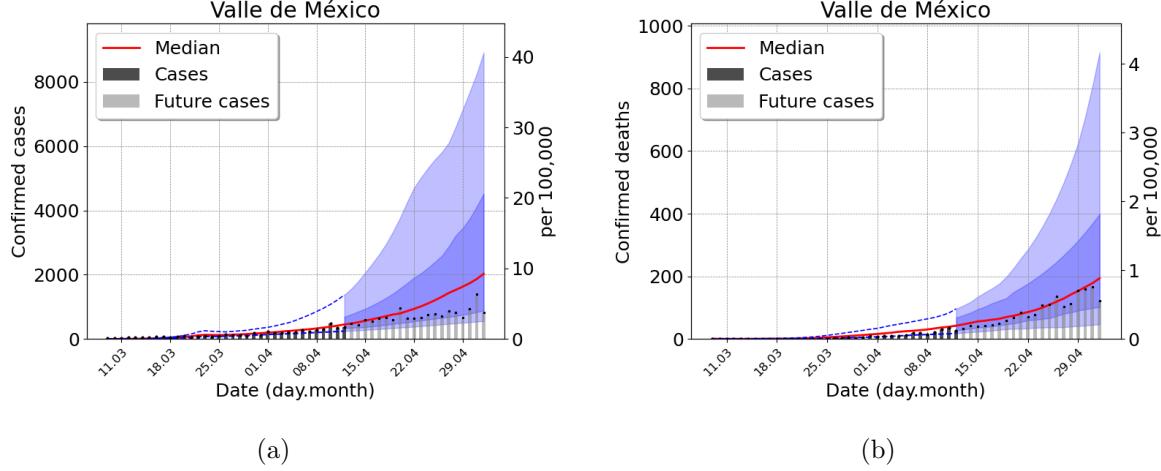


Figure 3: Outbreak analysis for Mexico city metropolitan area, using data from March 8 to April 12. (a) Confirmed cases (b) Confirmed deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are ten days from the point of prediction. Total population 21,942, 666 inhabitants.

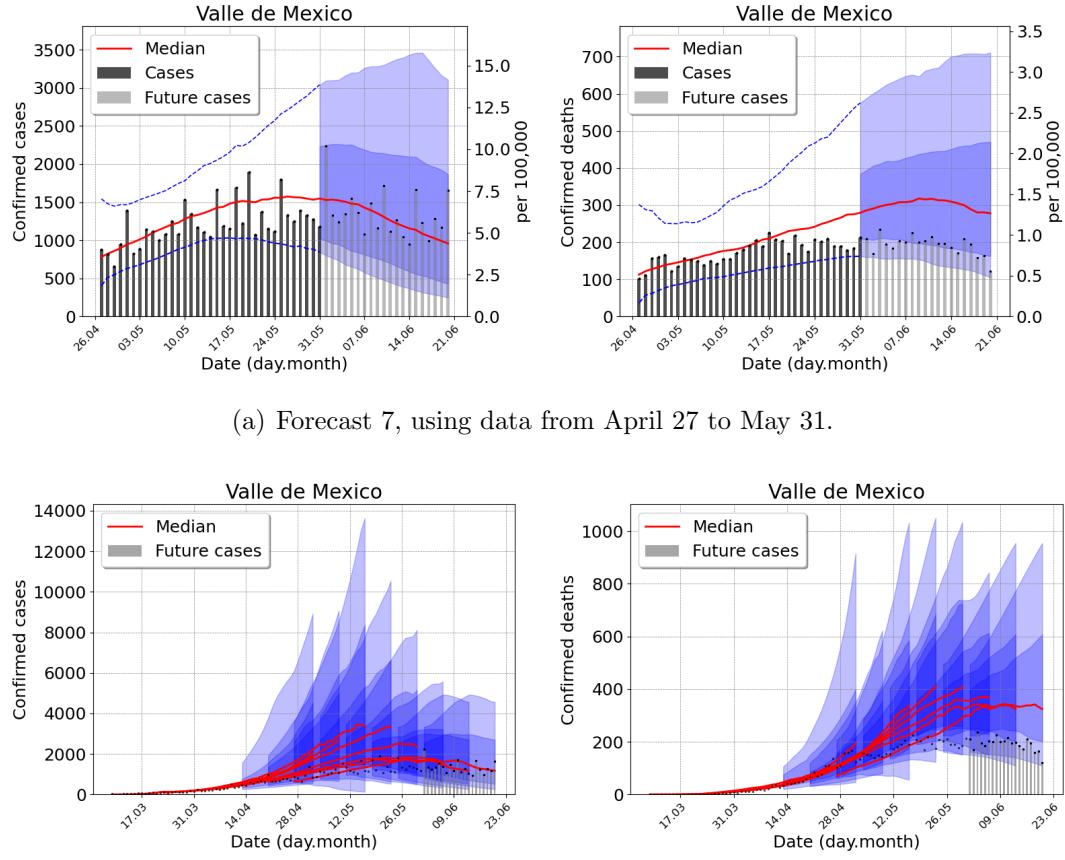
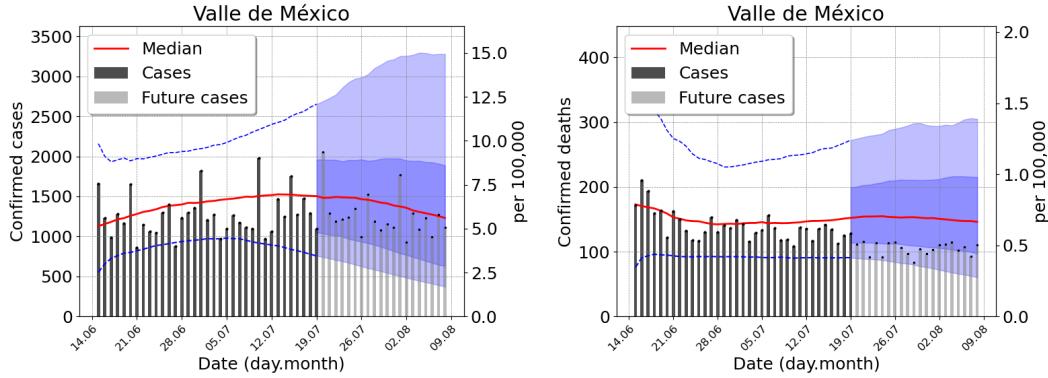
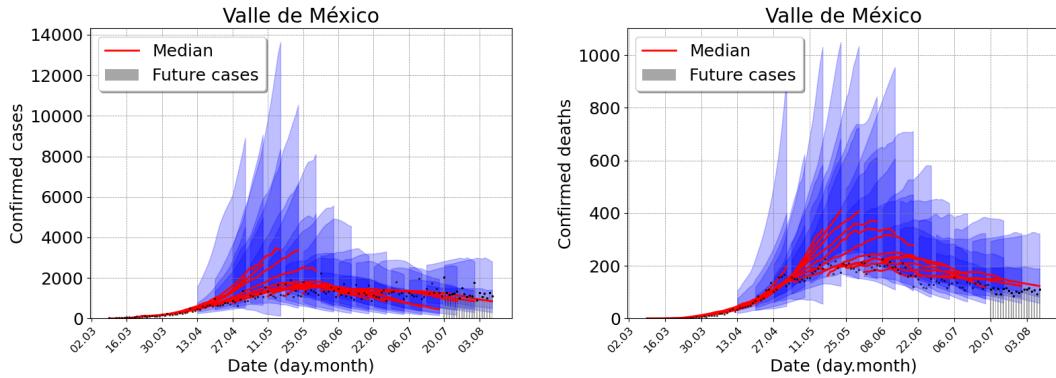


Figure 4: Outbreak analysis for Mexico city metropolitan area. From left to right, confirmed cases and deaths.

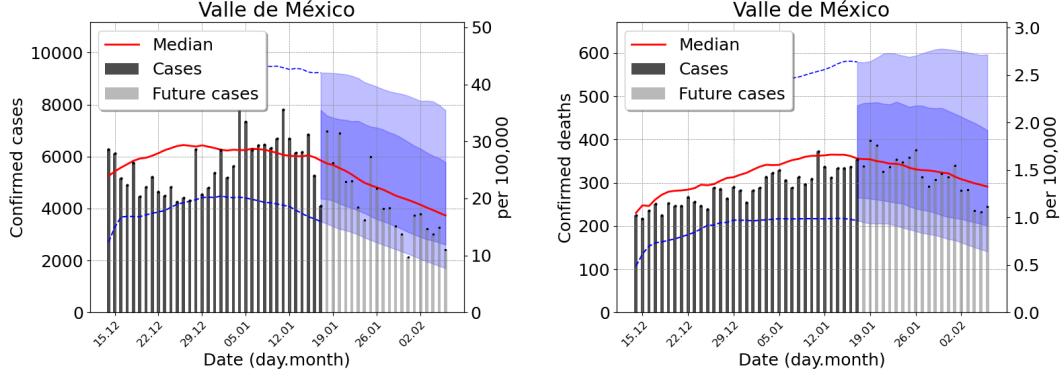


(a) Forecast 14, using data from June 15 to July 19

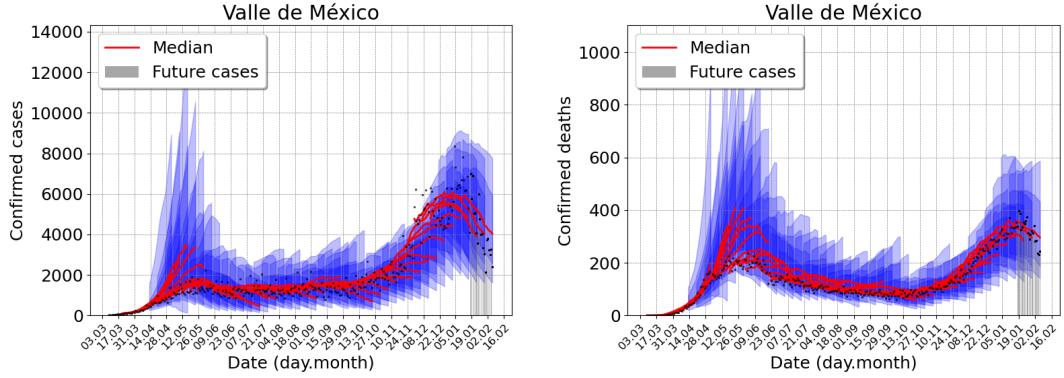


(b) The previous forecastings were superimposed.

Figure 5: Outbreak analysis for Mexico city metropolitan area. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.



(a) Forecast 40, using data from December 14 to January 17



(b) The previous forecastings were superimposed.

Figure 6: Outbreak analysis for Mexico city metropolitan area. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.

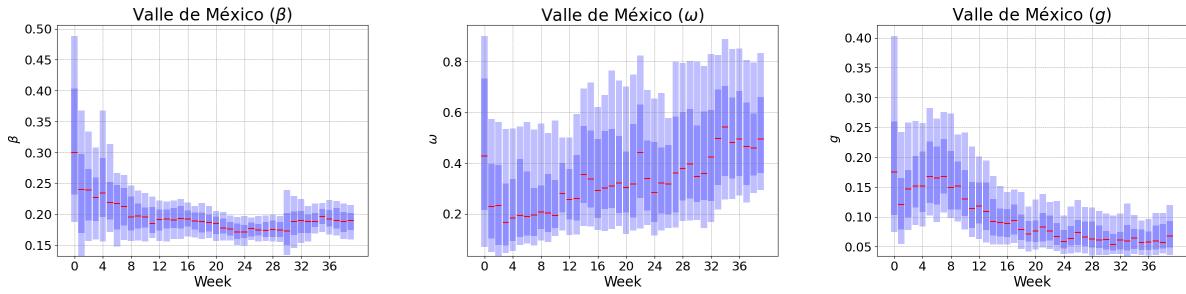


Figure 7: Outbreak analysis for Mexico city metropolitan area. From left to right, contact rate after lockdown ( $\beta$ ), proportion of the effective population ( $\omega$ ), and the fraction of infected dying ( $g$ ). Central red lines indicate median incidence forecast. Darker shaded region indicates forecast interquartile range, and lighter shaded region indicates 5–95th percentile range.

## Acknowledgments

The authors are partially founded by CONACYT CB-2016-01-284451 grant. AC was partially supported by UNAM PAPPIT-IN106118 grant. MLDT was funded by FORDECYT 296737 “CONSORCIO EN INTELIGENCIA ARTIFICIAL”.

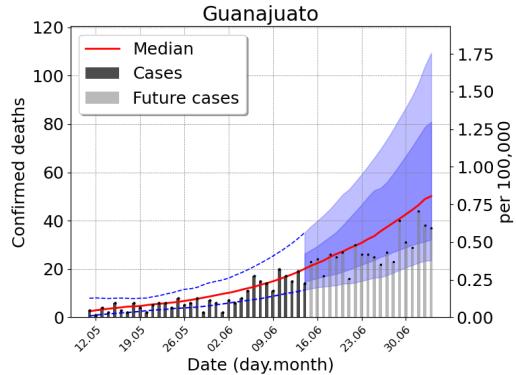
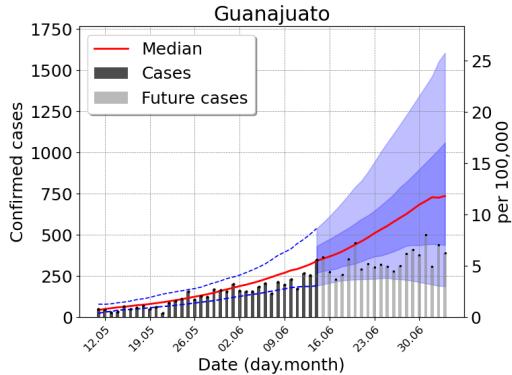
## References

- [1] Jason Asher. Forecasting ebola with a regression transmission model. *Epidemics*, 22:50–55, 2018.
- [2] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *arXiv preprint arXiv:2004.04741*, 2020.
- [3] Mario Castro, Saúl Ares, José A Cuesta, and Susanna Manrubia. The turning point and end of an expanding epidemic cannot be precisely forecast. *Proceedings of the National Academy of Sciences*, 117(42):26190–26196, 2020.
- [4] Claus O Wilke and Carl T Bergstrom. Predicting an epidemic trajectory is difficult. *Proceedings of the National Academy of Sciences*, 2020.
- [5] Steven G Krantz and Arni SR Srinivasa Rao. Level of underreporting including under-diagnosis before the first peak of covid-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling. *Infection Control & Hospital Epidemiology*, 41(7):857–859, 2020.
- [6] Hien Lau, Tanja Khosrawipour, Piotr Kocbach, Hirohito Ichii, Jacek Bania, and Veria Khosrawipour. Evaluating the massive underreporting and undertesting of covid-19 cases in multiple global epicenters. *Pulmonology*, 27(2):110–115, 2021.
- [7] Logan C Brooks, Evan L Ray, Jacob Bien, Johannes Bracher, Aaron Rumack, Ryan J Tibshirani, and Nicholas G Reich. Comparing ensemble approaches for short-term probabilistic covid-19 forecasts in the us. *International Institute of Forecasters*, 2020.
- [8] Ralf Engbert, Maximilian M Rabe, Reinhold Kliegl, and Sebastian Reich. Sequential data assimilation of the stochastic seir epidemic model for regional covid-19 dynamics. *Bulletin of mathematical biology*, 83(1):1–16, 2021.
- [9] David Champredon, Jonathan Dushoff, and David JD Earn. Equivalence of the erlang-distributed seir epidemic model and the renewal equation. *SIAM Journal on Applied Mathematics*, 78(6):3258–3278, 2018.
- [10] Angel N Desai, Moritz UG Kraemer, Sangeeta Bhatia, Anne Cori, Pierre Nouvellet, Mark Herringer, Emily L Cohn, Malwina Carrion, John S Brownstein, Lawrence C Madoff, et al. Real-time epidemic forecasting: challenges and opportunities. *Health security*, 17(4):268–275, 2019.
- [11] Graham C Gibson, Nicholas G Reich, and Daniel Sheldon. Real-time mechanistic bayesian forecasts of covid-19 mortality. *medRxiv*, 2020.
- [12] Sarah F McGough, John S Brownstein, Jared B Hawkins, and Mauricio Santillana. Forecasting zika incidence in the 2016 latin america outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS neglected tropical diseases*, 11(1):e0005295, 2017.

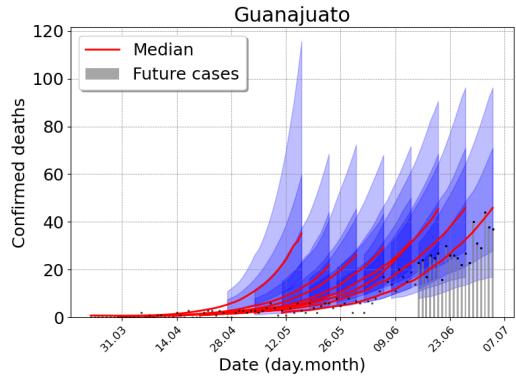
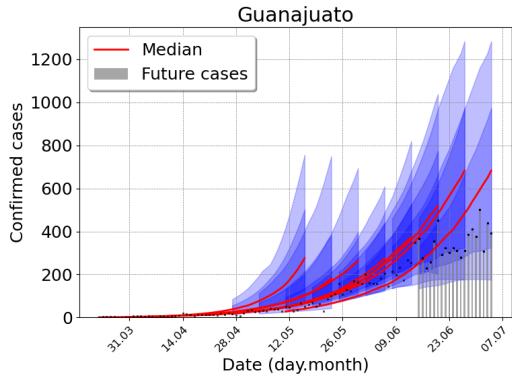
- [13] Cécile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, et al. The rapid ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.
- [14] Yien Ling Hii, Huaiping Zhu, Nawi Ng, Lee Ching Ng, and Joacim Rocklöv. Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis*, 6(11):e1908, 2012.
- [15] Marcos A. Capistran, Antonio Capella, and J. Andrés Christen. Forecasting hospital demand in metropolitan areas during the current covid-19 pandemic and estimates of lockdown-induced 2nd waves. *PLOS ONE*, 16(1):1–16, 2021.
- [16] SA Lauer, KH Grantz, Q Bi, FK Jones, Q Zheng, HR Meredith, AS Azman, and NG Reich. 181 lessler j. the incubation period of coronavirus disease 2019 (covid-19) from publicly 182 reported confirmed cases: estimation and application. *Ann Intern Med*, 2020.
- [17] Xuan Jiang, Simon Rayner, and Min-Hua Luo. Does sars-cov-2 has a longer incubation period than sars and mers? *Journal of medical virology*, 92(5):476–478, 2020.
- [18] Quan-Xin Long, Xiao-Jun Tang, Qiu-Lin Shi, Qin Li, Hai-Jun Deng, Jun Yuan, Jie-Li Hu, Wei Xu, Yong Zhang, Fa-Jin Lv, et al. Clinical and immunological assessment of asymptomatic sars-cov-2 infections. *Nature medicine*, 26(8):1200–1204, 2020.
- [19] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Nat-suko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, 2020.
- [20] Qifang Bi, Yongsheng Wu, Shuijiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, et al. Epidemiology and transmission of covid-19 in shenzhen china: Analysis of 391 cases and 1,286 of their close contacts. *MedRxiv*, 2020.
- [21] J. A. Christen and C. Fox. A general purpose sampling algorithm for continuous distributions (the t-walk). *Bayesian Analysis*, 5(2):263–282, 2010. cited By 60.

## 4.1 Other examples

### Guanajuato

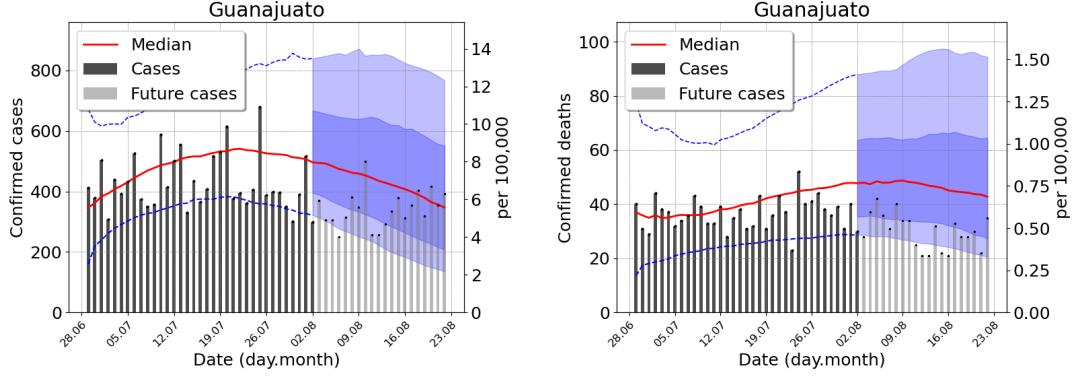


(a) Forecast 7, using data from May 11 to June 14

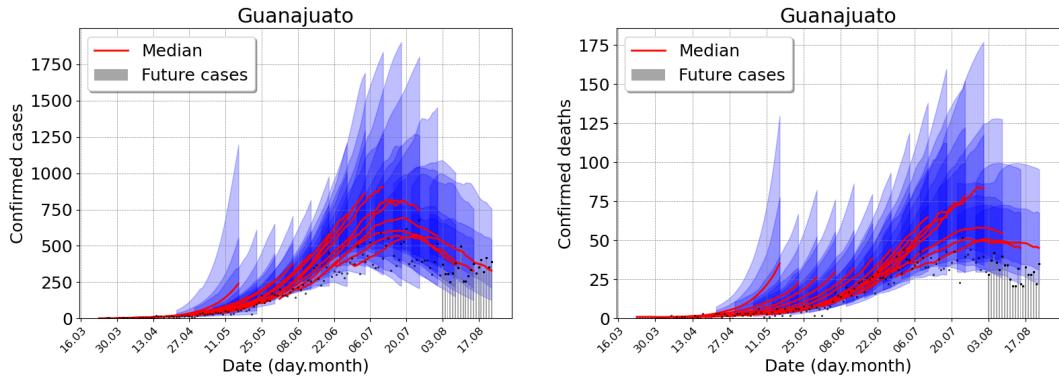


(b) The previous forecastings were superimposed..

Figure 8: Outbreak analysis for Guanajuato. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction. Total population 6,228,175 inhabitants.

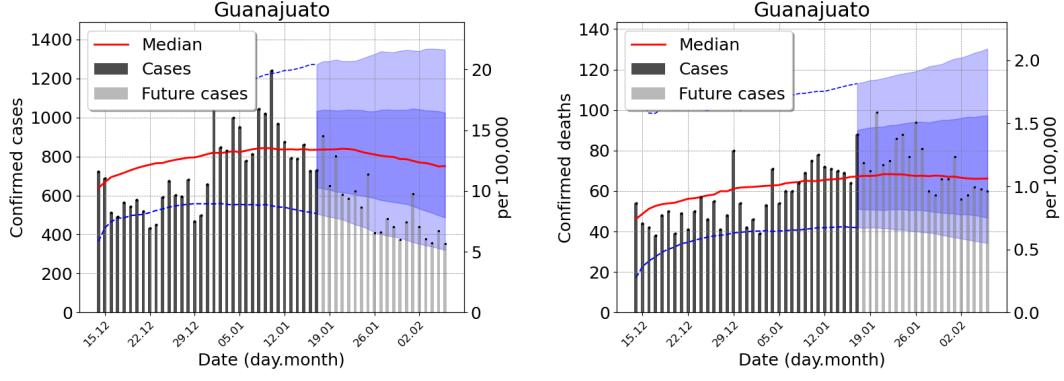


(a) Forecast 14, using data from June 27 to August 2.

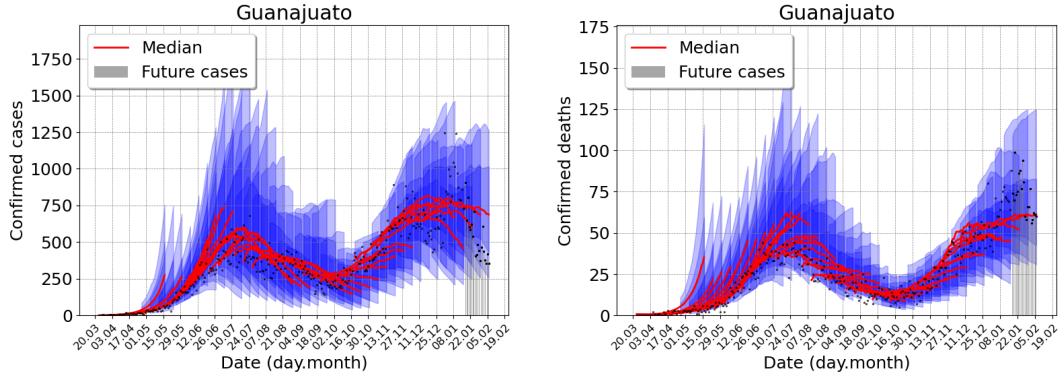


(b) The previous forecastings were superimposed.

Figure 9: Outbreak analysis for Guanajuato. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.



(a) Forecast 14, using data from December 14 to January 17.



(b) The previous forecastings were superimposed.

Figure 10: Outbreak analysis for Guanajuato. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.

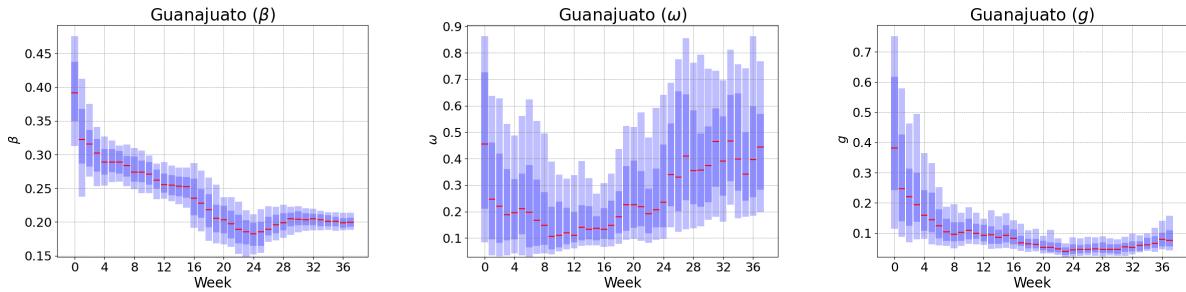
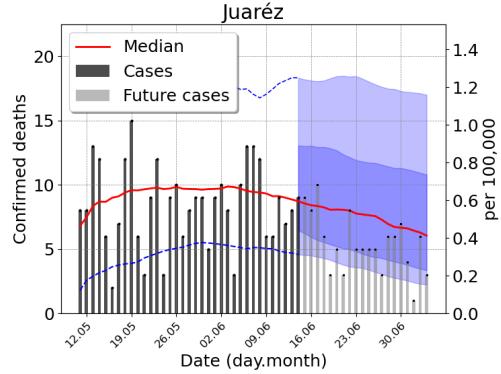
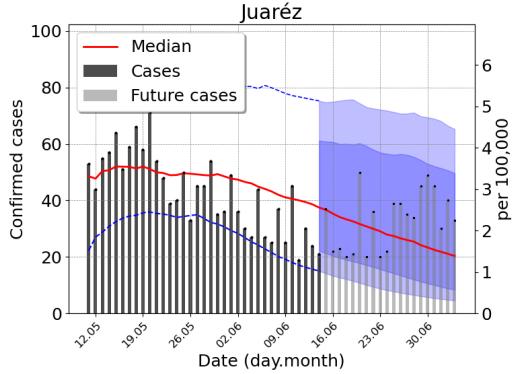
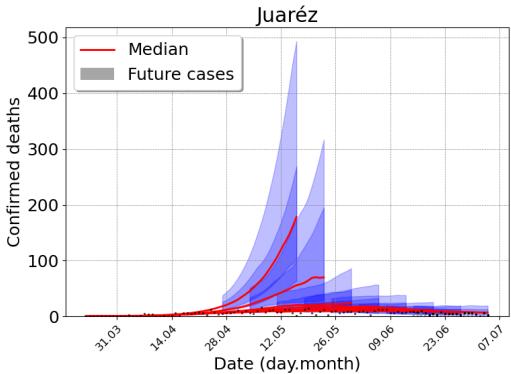
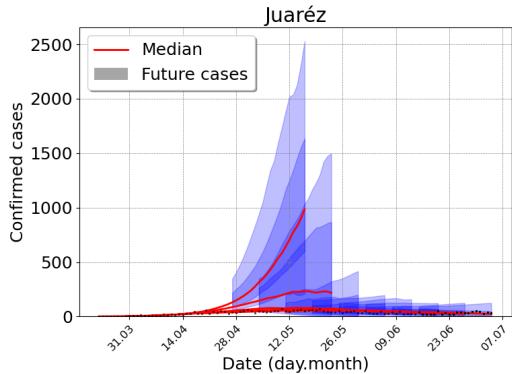


Figure 11: Outbreak analysis for Guanajuato. From left to right, contact rate after lockdown ( $\beta$ ), proportion of the effective population ( $\omega$ ), and the fraction of infected dying ( $g$ ). Central red lines indicate median incidence forecast. Darker shaded region indicates forecast interquartile range, and lighter shaded region indicates 5–95th percentile range.

## Juaréz

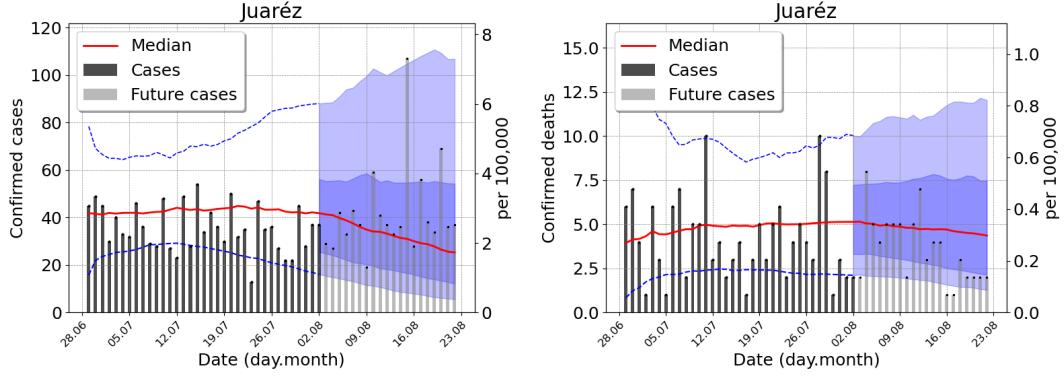


(a) Forecast 7, using data from May 11 to June 14.

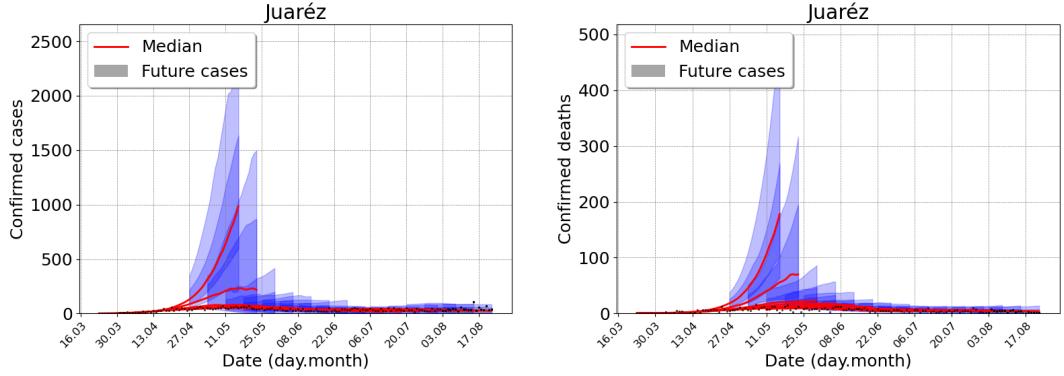


(b) The previous forecastings were superimposed.

Figure 12: Outbreak analysis for Juaréz. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction. Total population 1,464,930 inhabitants.

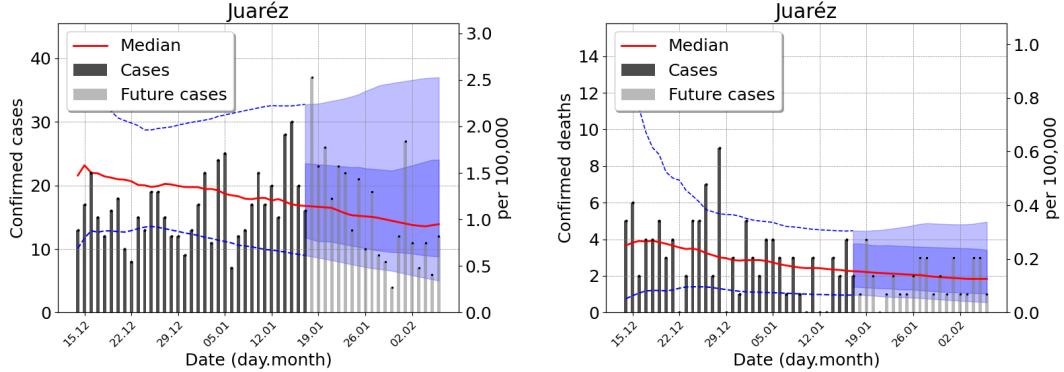


(a) Forecast 14, using data from June 29 to August 2

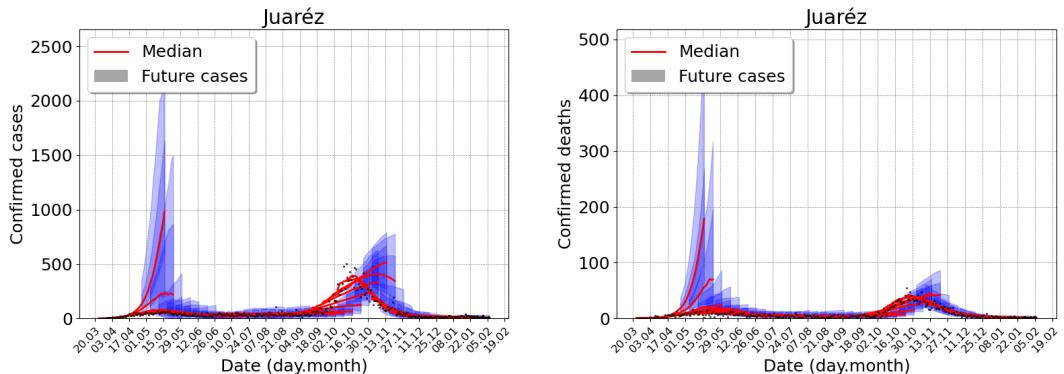


(b) The previous forecastings were superimposed.

Figure 13: Outbreak analysis for Juaréz. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.



(a) Forecast 38, using data from December 15 to January 18



(b) The previous forecastings were superimposed.

Figure 14: Outbreak analysis for Juaréz. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.

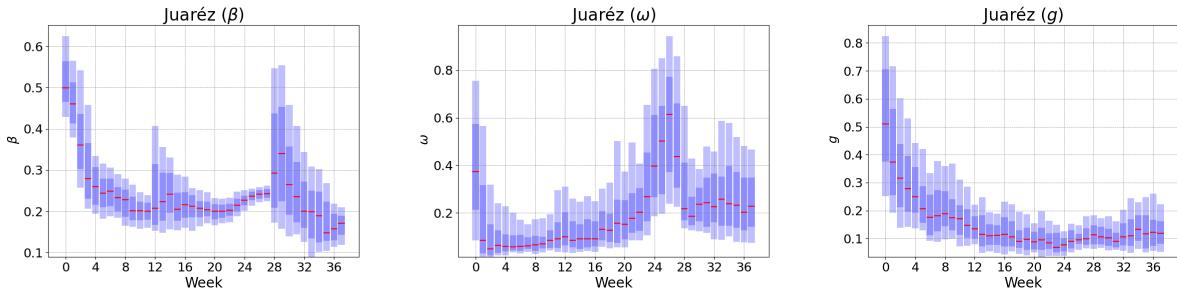
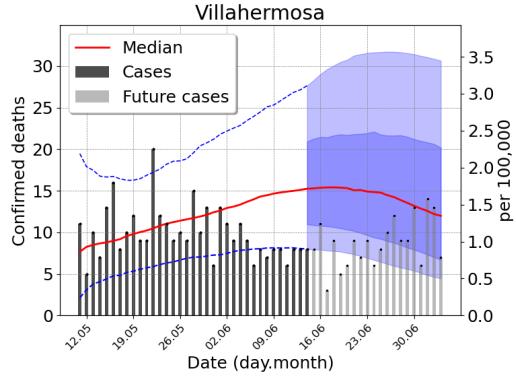
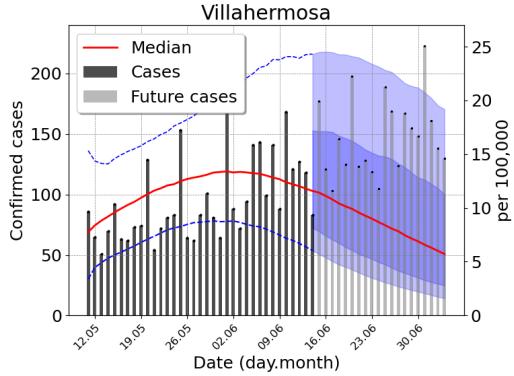
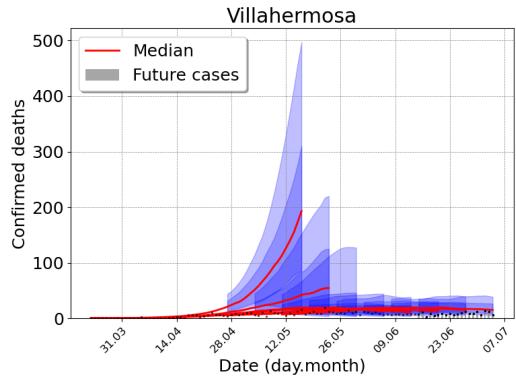
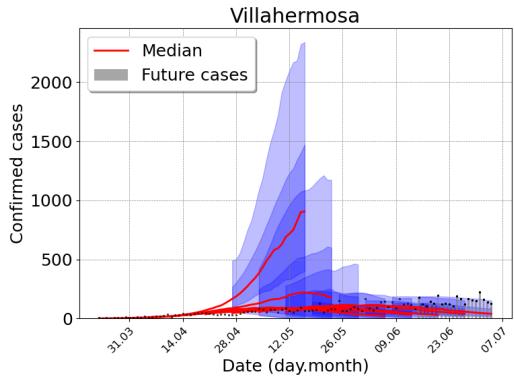


Figure 15: Outbreak analysis for Juaréz. From left to right, contact rate after lockdown ( $\beta$ ), proportion of the effective population ( $\omega$ ), and the fraction of infected dying ( $g$ ). Central red lines indicate median incidence forecast. Darker shaded region indicates forecast interquartile range, and lighter shaded region indicates 5–95th percentile range.

## Villa Hermosa

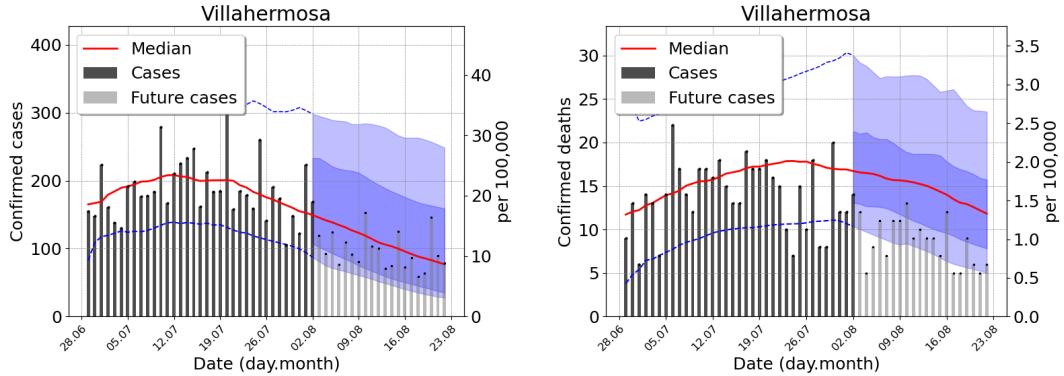


(a) Forecast 7, using data from May 11 to June 14.

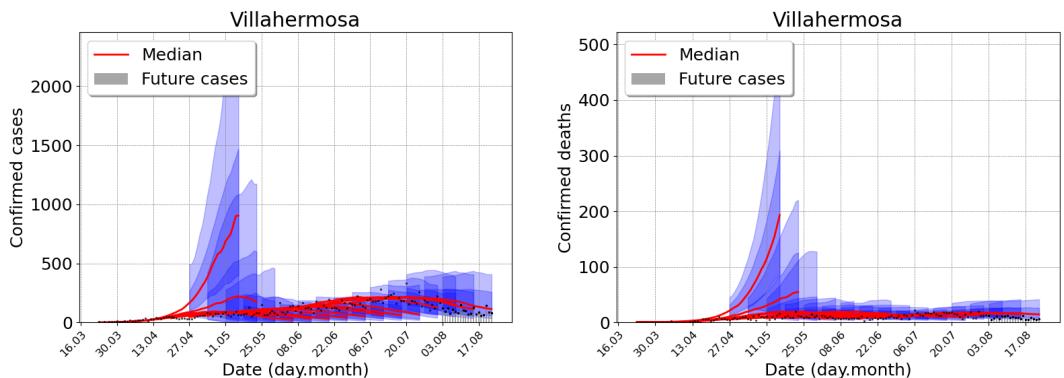


(b) The previous forecastings were superimposed.

Figure 16: Outbreak analysis for Villa Hermosa. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction. Total population 888,867 inhabitants.

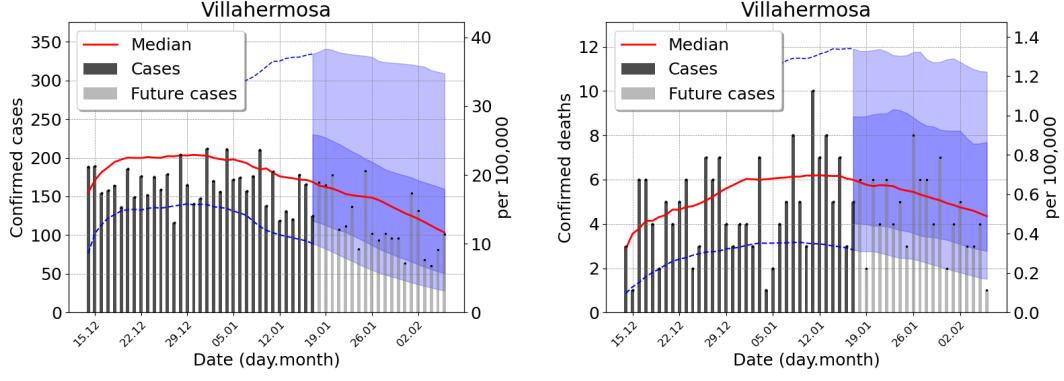


(a) Forecast 14, using data from June 29 to August 2.

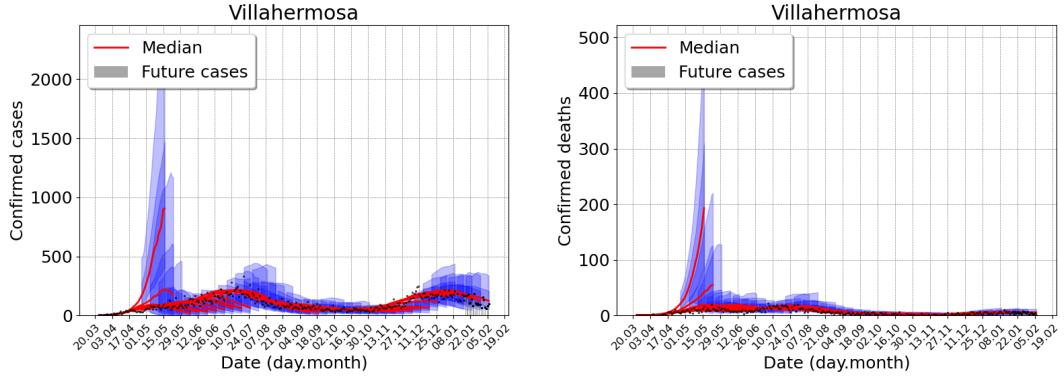


(b) Previous forecast.

Figure 17: Outbreak analysis for Villa Hermosa. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.



(a) Forecast 38, using data from December 14 to January 17.



(b) Previous forecast.

Figure 18: Outbreak analysis for Villa Hermosa. From left to right, confirmed cases and deaths. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 5–95th percentile range. All displayed forecast durations are 20 days from the point of prediction.

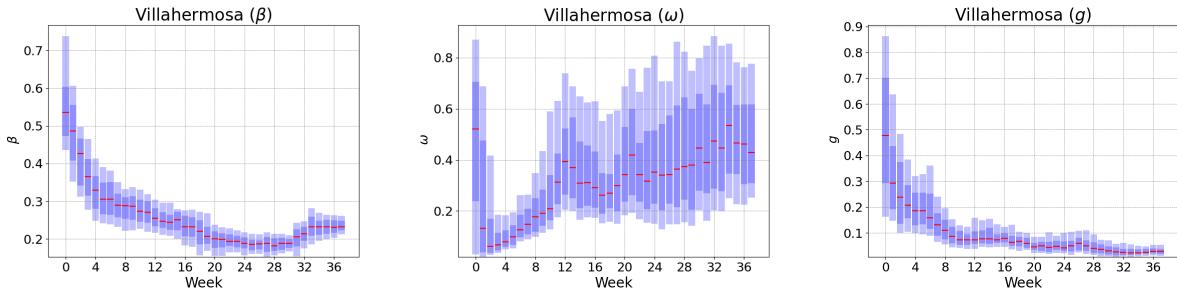


Figure 19: Outbreak analysis for Villa Hermosa. From left to right, contact rate after lockdown ( $\beta$ ), proportion of the effective population ( $\omega$ ), and the fraction of infected dying ( $g$ ). Central red lines indicate median incidence forecast. Darker shaded region indicates forecast interquartile range, and lighter shaded region indicates 5–95th percentile range.