Supplementary material

S1 Testing policy, testing bias, and positivity rate

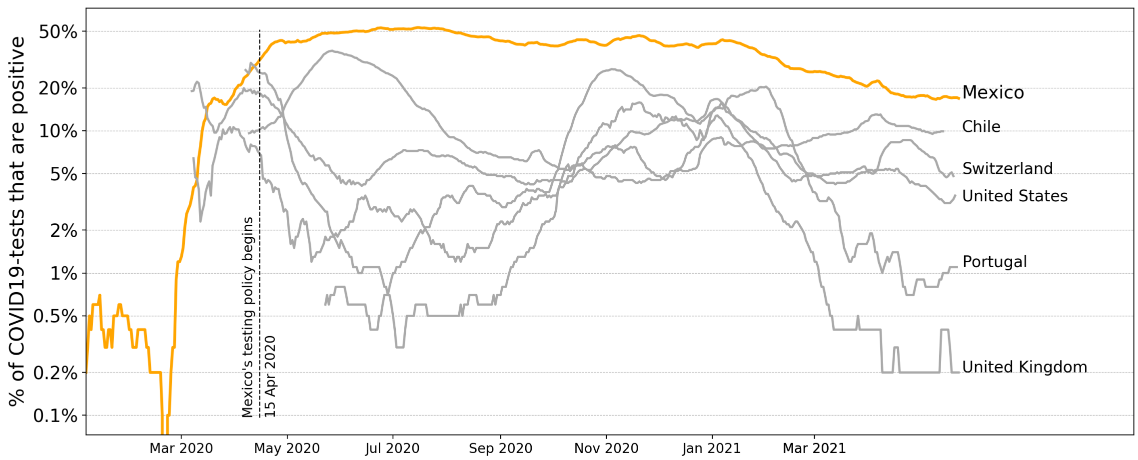


Figure S1: **Percentage of COVID19-tests that are positive for selected countries**. Testing policies during the pandemic have been different among countries and correspond to distinct strategies and objectives. While testing in the open population for contact tracing has been standard in many countries, Mexico's federal health authorities COVID-19 test-policy targets all public hospital admissions exclusively. Starting April 2020, Mexico's testing policy [32] (p. 19, in Spanish) is to test 100% and randomly select 10% of severe and mild suspected cases, respectively. Severe cases correspond to hospital admitted patients and mild cases to ambulatory infections. This restricted testing policy produces a biased sample of positive COVID-19 cases. The upside is that this bias is consistent and constant in time, as the figure confirms. Since infection rates can be estimated better from constant bias samples than from more extensive ones of unknown bias, we have - with this testing policy- a more reliable proxy of the pandemic evolution, better suited for modeling and forecasting. Reasonable estimates of the true number of infected individuals at a given time are impossible to obtain until a characterization of the asymptomatic infection is available. The latter is true for any testing policy besides the rather unrealistic case where the whole population is continuously tested.

S2 Linear observation operators

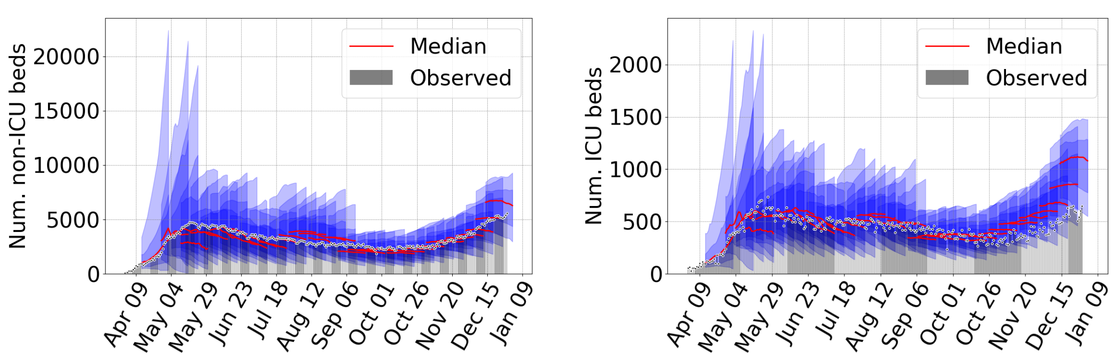
As explained in the main paper, linear observation operators have the form of a convolution between a kernel G with the estimated individuals in a particular compartment. The goal of these operators is to implement the record-keeping and counting needed to keep track of the individuals in the linear compartments of the model that are not required for the inference problem. This approach has further advantages: first, we do not have to use exponential or Erlang waiting-time distributions only (see [12]). Second, we can use these linear observation operators offline, at any compartment outside the nonlinear term in the differential equations system. Third, the linear part of the system can be as complex as needed, as is the case of many SEIRD types of epidemiological models.

We present an application to estimate bed and ICU occupancy with a linear observation operator to make these ideas precise. After the inference and forecasting process, we have an estimate of the observed individuals leaving the last Erlang box $O_{last}(t)$ at time t presented in Figure 2. From hospital records, we know the fraction of hospital admitted individuals $h(t)$ at time t, and we can estimate the waiting-time distribution of hospital bed occupancy $G(t, t_0)$. The value of $G(t_1, t_0)$ is equal to the fraction of individuals admitted at time t_0 that still occupy a hospital bed at time t_1. Hence, the number of beds occupied at time t is given by

$$H(t) = \int_{-\infty}^{t} h(\tau)G(t, \tau)O_{last}(\tau)d\tau. \tag{2}$$

The integral 2 defines the linear operator as a renewal equation. Note that, if the function $G(t, t_0)$ is an Erlang distribution, there exists a system of ordinary differential equations as depicted in Figure S3 that can be added to our original SEIRD model to estimate $H(t)$ by the results in [12]. In this case, the model constant ($h, k, g, \sigma_2, \sigma_3$ and σ_4) have be adapted consistently to achieve the equivalence.

In Figure S2 we show the result of the observation operators applied to the weakly forecasts of the hospital bed and ICU demand. Health authorities in the Mexican Federal Government have used these forecasts to assist their decision-making processes. The kernel function G was estimated from public hospital records. For further examples, see [35] (in Spanish) for the forecast's weekly updates.



(a) Non-ICU bed demand

(b) ICU bed demand

Figure S2: Weakly estimates of Mexico City metropolitan area hospital demand. Grey bars represent the observed occupancy of non-ICU and ICU hospital beds.

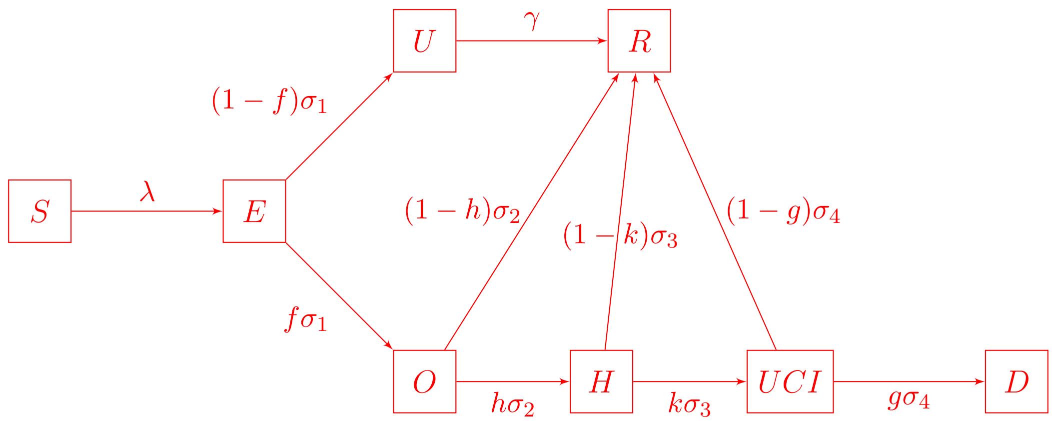


Figure S3: Extension of the original SEIR model presented in the main paper with the compartments corresponding to hospital beds and ICU-beds. Note that, only if G is an Erlang distribution this model will produce the same estimates obtained by the observation operator (2) .

S3 State codes and population considered in epidemic models

State name	Code	Pop.	State name	Code	Pop.
Aguascalientes	AS	1,434,635	Nayarit	NT	1,288,571
Baja California	BC	3,634,868	Nuevo León	NL	5,610,153
Baja California Sur	BS	804,708	Oaxaca	OC	4,143,593
Campeche	CC	1,000,617	Puebla	PL	6,604,451
Chiapas	CS	5,730,367	Querétaro	QO	2,279,637
Chihuahua	CH	3,801,487	Quintana Roo	QR	1,723,259
Coahuila	CL	3,218,720	San Luis Potosí	SP	2,866,142
Colima	CM	785,153	Sinaloa	SL	3,156,674
Durango	DG	1,868,996	Sonora	SR	3,074,745
Guanajuato	GT	6,228,175	Tabasco	TC	2,572,287
Guerrero	GR	3,657,048	Tamaulipas	TS	3,650,602
Hidalgo	HG	3,086,414	Tlaxcala	TL	1,380,011
Jalisco	JC	8,409,693	Veracruz	VZ	8,539,862
Estado de México(*)	MC	4,640,934	Yucatán	YN	2,259,098
Michoacán	MN	4,825,401	Zacatecas	ZS	1,666,426
Morelos	MS	2,044,058	Mexico city area(*)	ZVMX	21,942,666

Table S1: State names, their corresponding codes and population sizes used in our examples, see `https://en.wikipedia.org/wiki/Administrative_divisions_of_Mexico` for maps an further details. (*) Mexico's City metropolitan area, with about 22 million inhabitants, includes some counties that do not belong to the official Mexico's City federal division which only has 8 million inhabitants. As this does not make sense for epidemic modeling, in this study, we define ZVMX as Mexico's City metropolitan area and include the corresponding population of the nearby states. Notably, for the "Estado de Mexico" (MC), we include 12 million inhabitants who live in the ZVMX and remove them from MC.

S4 Estimates for model parameters for all Mexican states and Mexico City's metropolitan area

S4.1 Posterior distributions for the time dependent susceptible pool ω

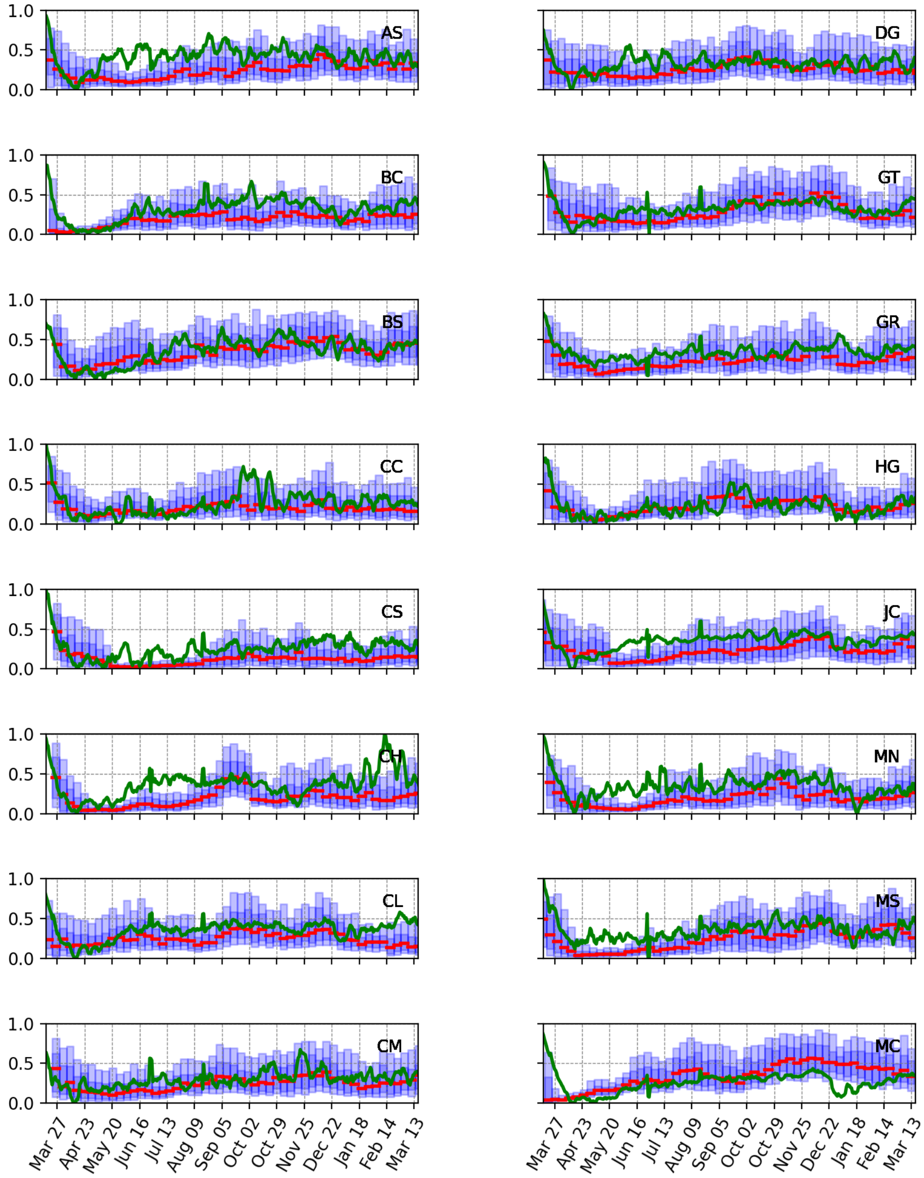


Figure S4: **Posterior distributions for the time dependent susceptible pool** ω. We present the weekly posterior distributions for the first 16 Mexican states, with colored vertical box-plots. Light blue is the 10% to 90% quantile range and dark blue are the interquantile ranges; the red lines are the medians. We also added a mobility index (green line, arbitrary units) derived from social media tracking. The susceptible pool of people participating in the epidemic in our model ω, indeed only estimated using the epidemic data, seems to correlate (sometimes remarkably well) the mobility index in the corresponding areas. See also Figure S5.

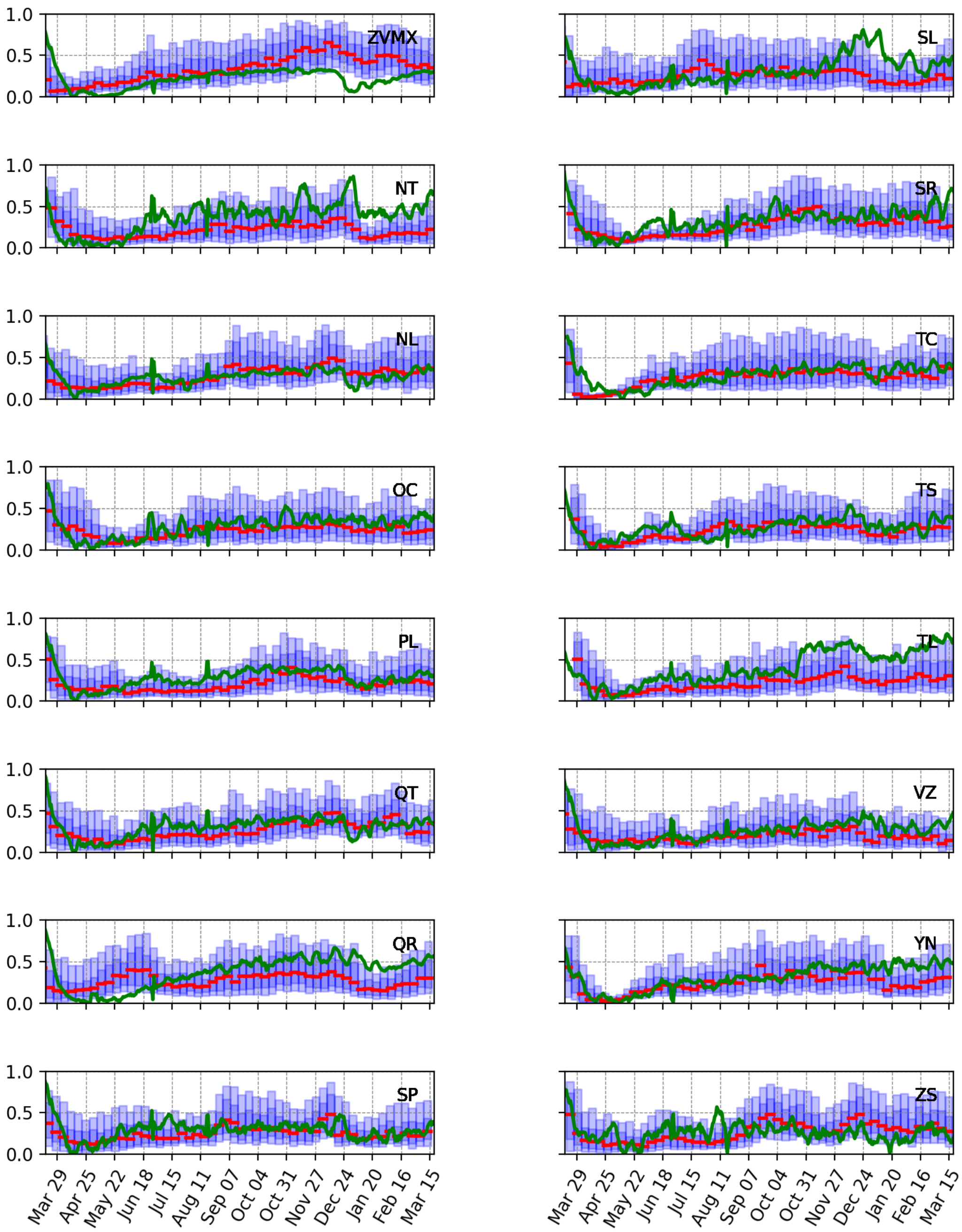


Figure S5: **Posterior distributions for the time dependent susceptible pool** ω. We present the weekly posterior distributions for Mexico city metro area and the remaining 15 Mexican states, see Figure S4 for details.

503 ## S4.2 Posterior distributions for the Infection Contact rate β

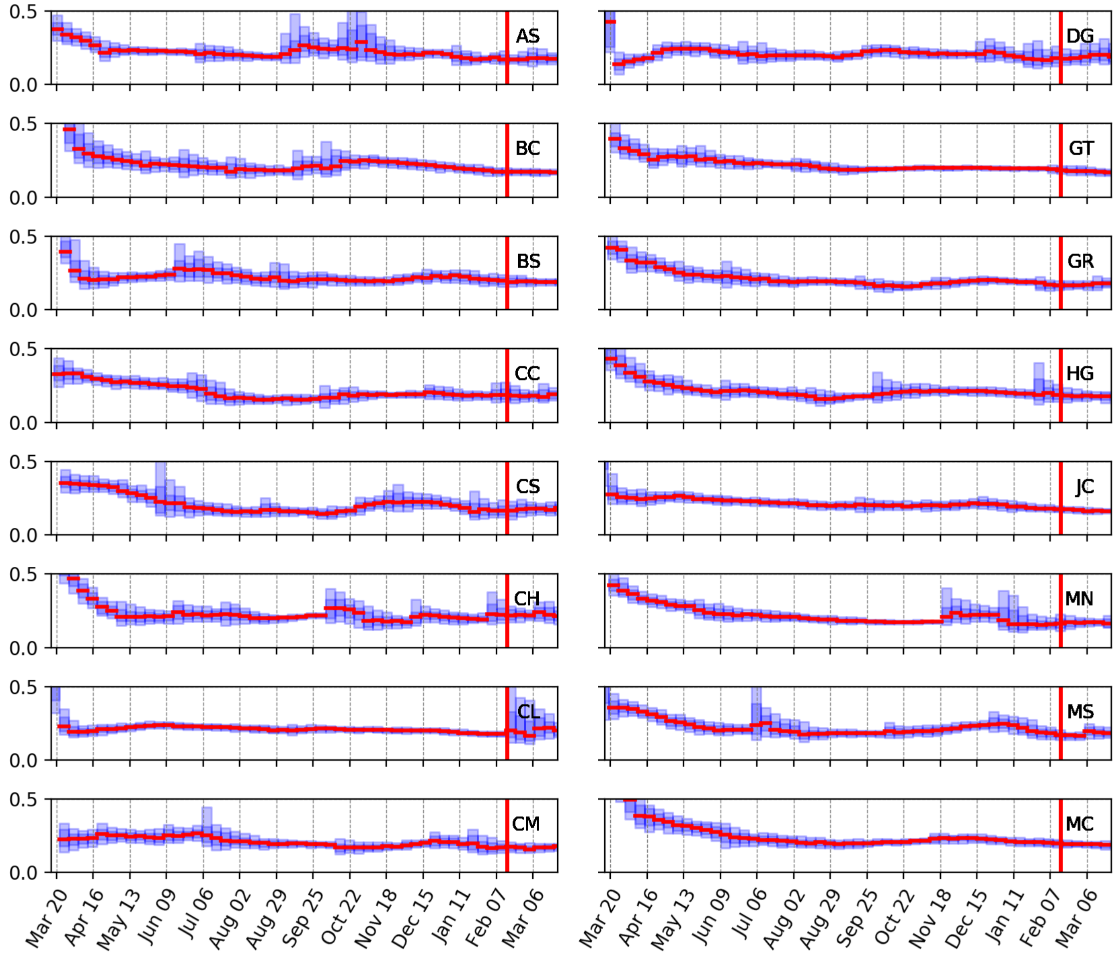


Figure S6: **Posterior distributions for the Infection Contact rate** β. We present the weekly posterior distributions for the first 16 Mexican states for β, displayed as vertical box plots (see Figure S4 for details). The vertical red line marks the day vaccination began. Despite the variability in population, epidemic outbreak history, and other socioeconomic factors, the estimate of Infection Contact rate β show a relatively equal value in all cases. This evidence, combined with the observed variability in the pool of susceptible individuals ω, would imply that this quantity is mainly disease dependent. The uncertainty of these posteriors increases in the wave's exponential growth periods, where a confounding effect between ω and β exists [7]. See also Figure S7.

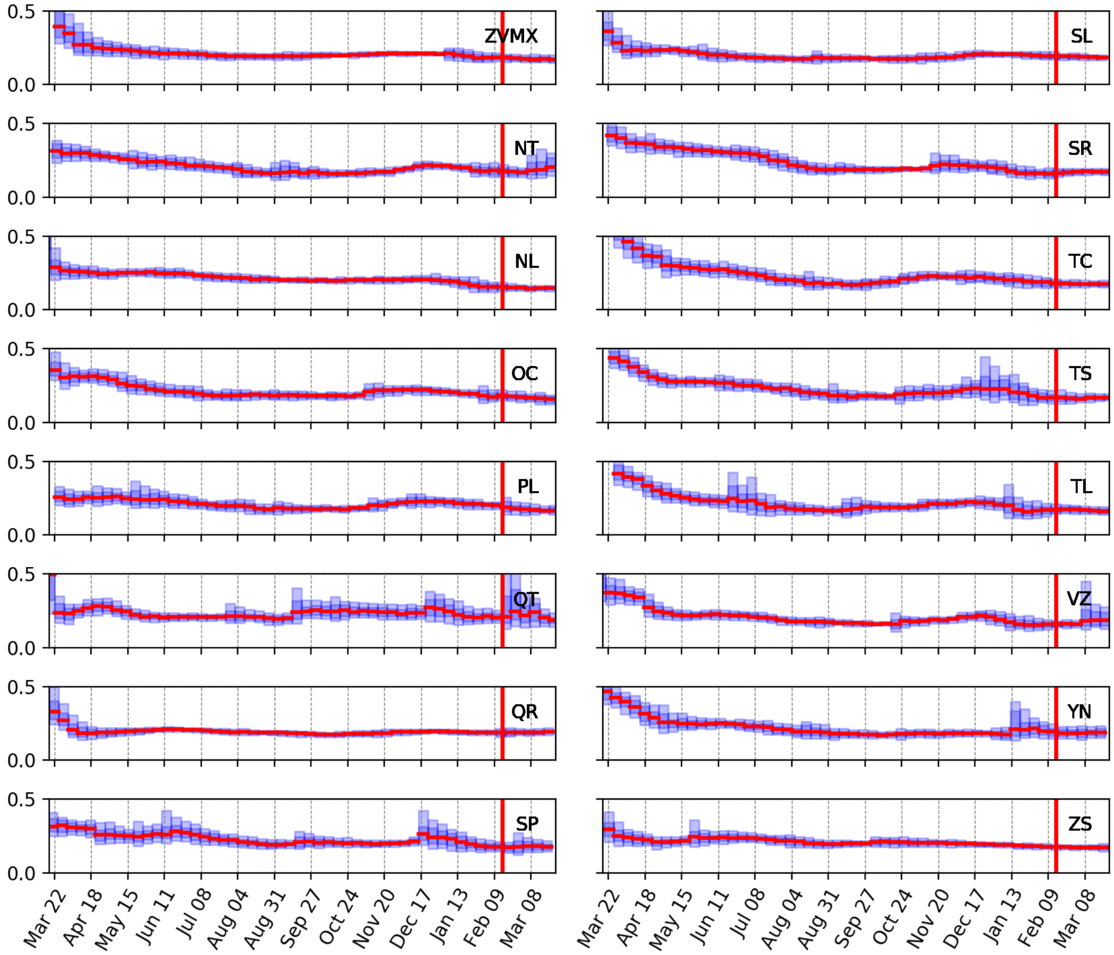


Figure S7: **Posterior distributions for the Infection Contact rate** β. Weekly posterior distributions for β, for Mexico city metro area and the remaining 15 Mexican states, see Figure S6 for details.

S4.3 Posterior distributions for the Hospital Fatality rate g

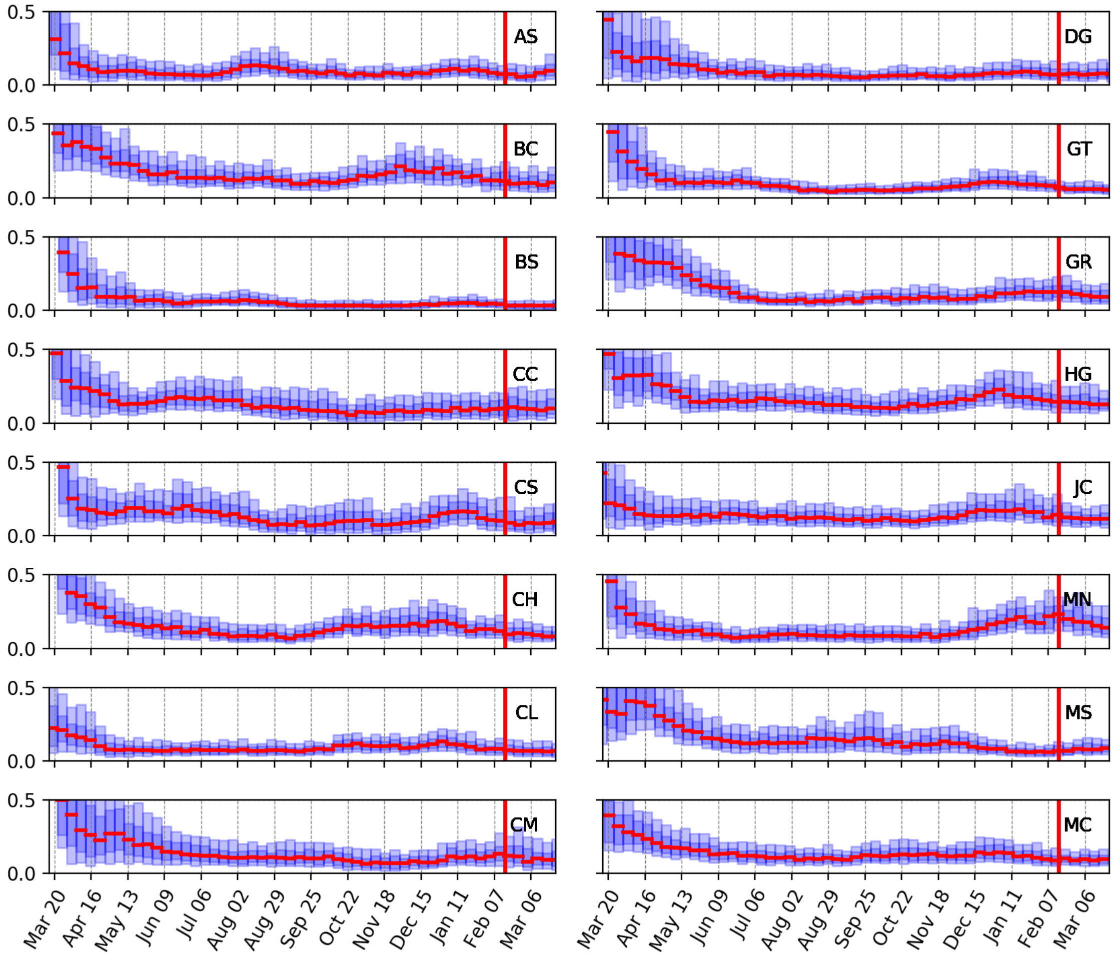


Figure S8: **Posterior distributions for the Hospital Fatality rate** g. We present the weekly posterior distributions for the first 16 Mexican states for g, displayed as vertical box plots (see Figure S4 for details). Interestingly, its value declines over time, with a slight and steady further decline starting in February 2021 in some cities. It could be argued that last reduction is consistent with the local vaccination campaigns on the elderly population. See also Figure S9.

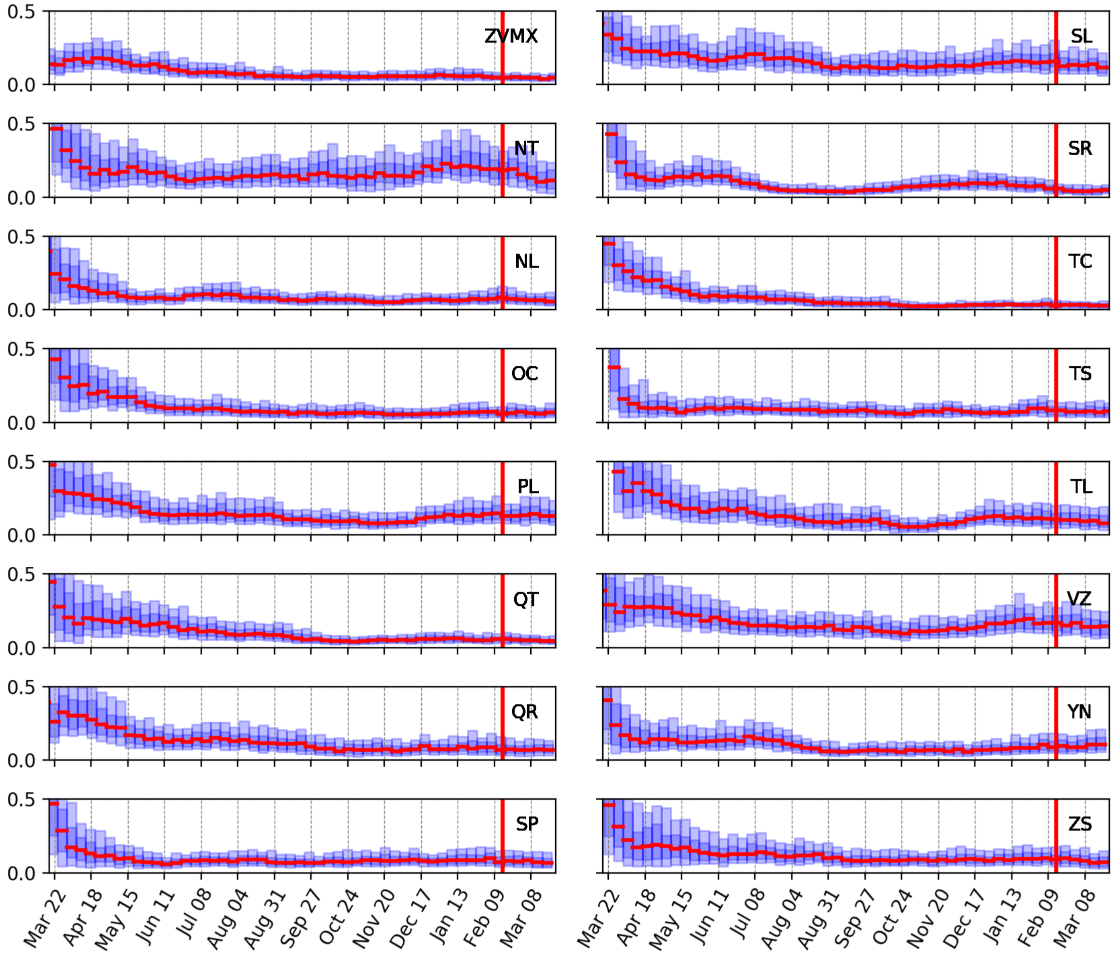


Figure S9: **Posterior distributions for the Hospital Fatality rate** g. Weekly posterior distributions for g, for Mexico city metro area and the remaining 15 Mexican states, see Figure S8 for details.

S5 Forecast results for all Mexican states and Mexico City's metropolitan area

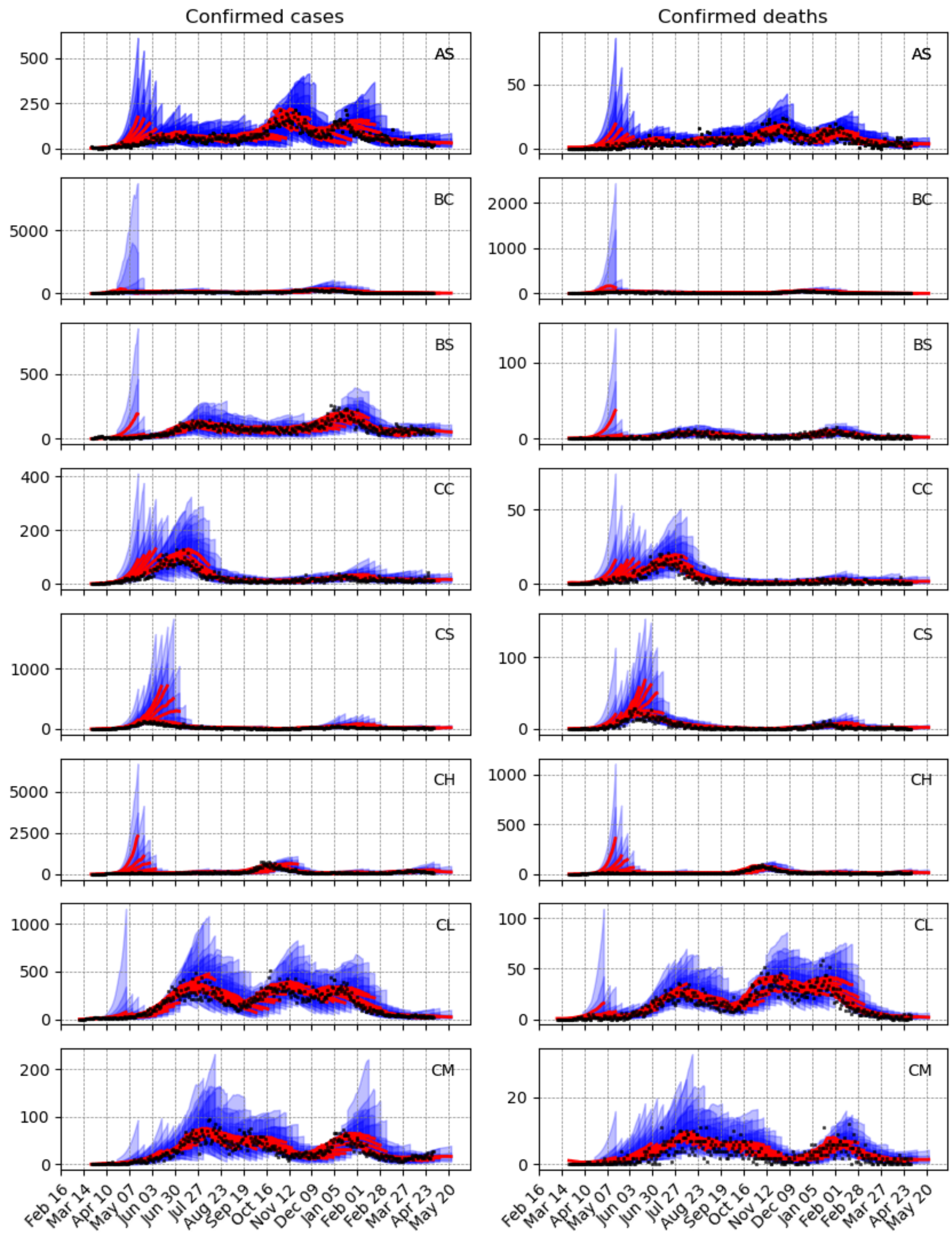


Figure S10: **Forecast results** for all Mexican states and Mexico City's metropolitan area. Left column, confirmed cases, and right column confirmed deaths. The three week posterior predictive distribution is depicted, for each of the weekly moving forecasts windows. Central red lines indicate the median incidence forecast. The darker shaded region indicates the interquartile forecast range, and the lighter shaded region indicates the 10% to 90% quantile range. See also Figures S11, S12 and S13.

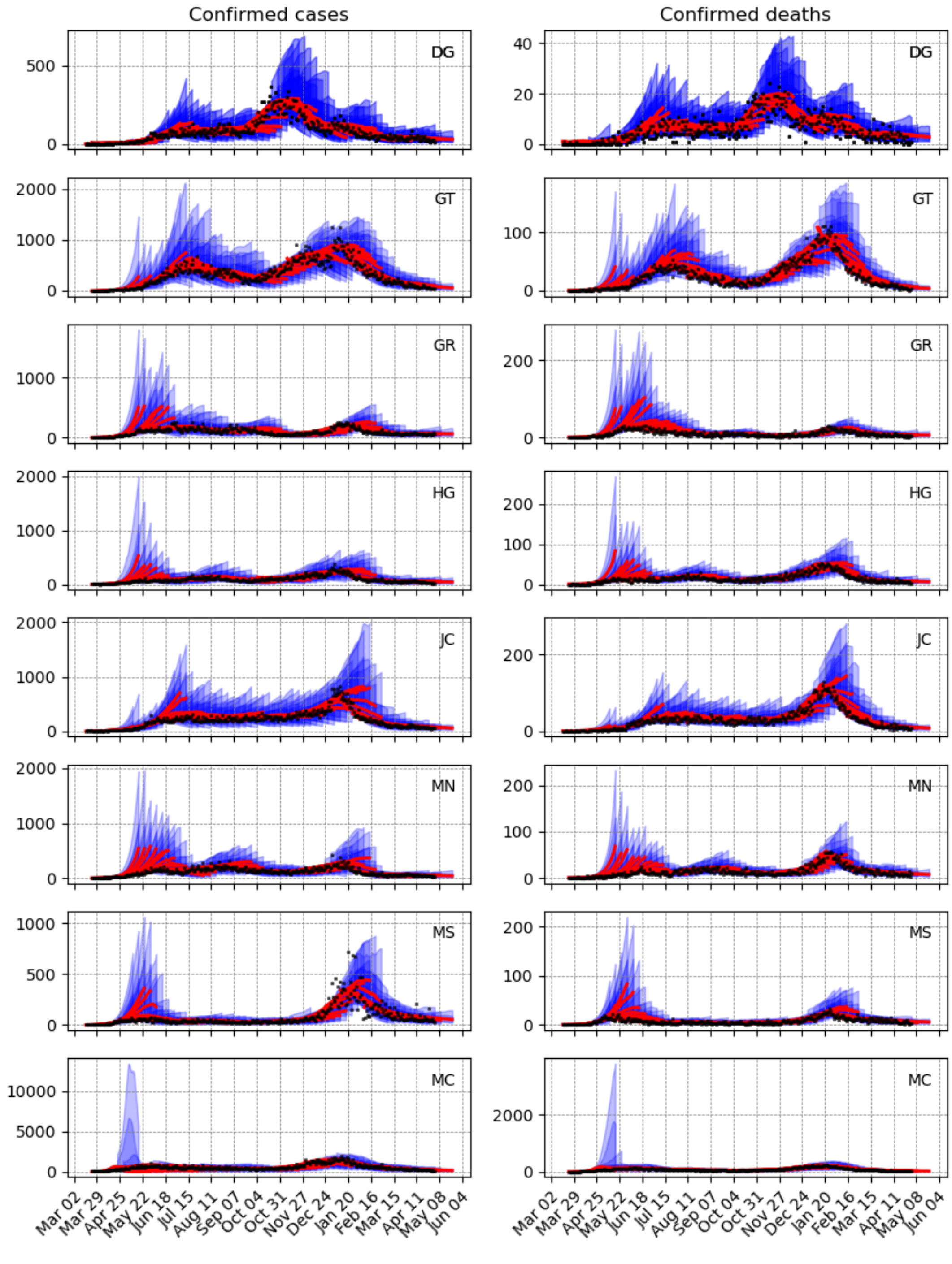


Figure S11: **Forecast results** for all Mexican states, see Figure S10 for details.

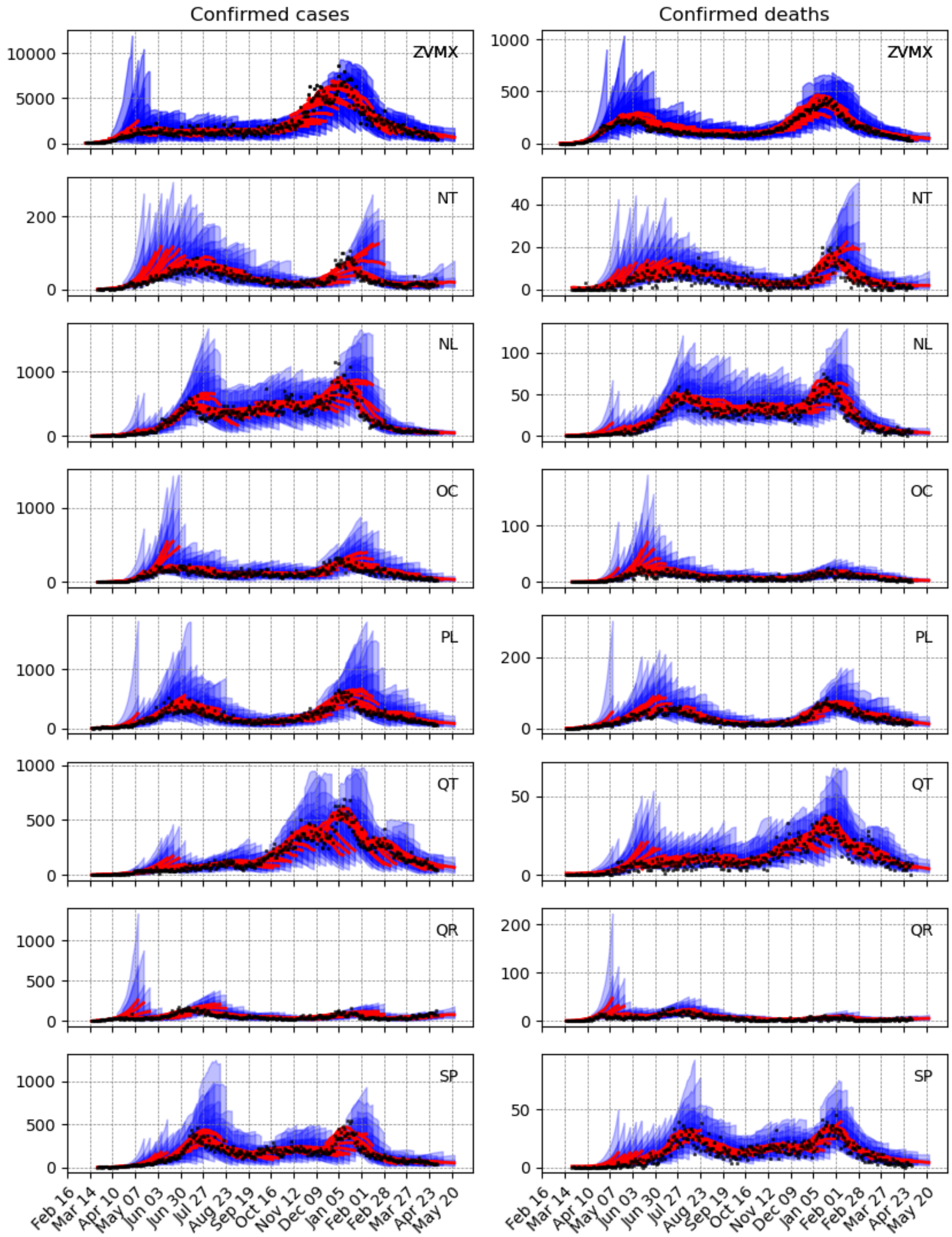


Figure S12: **Forecast results** for all Mexican states, see Figure S10 for details.

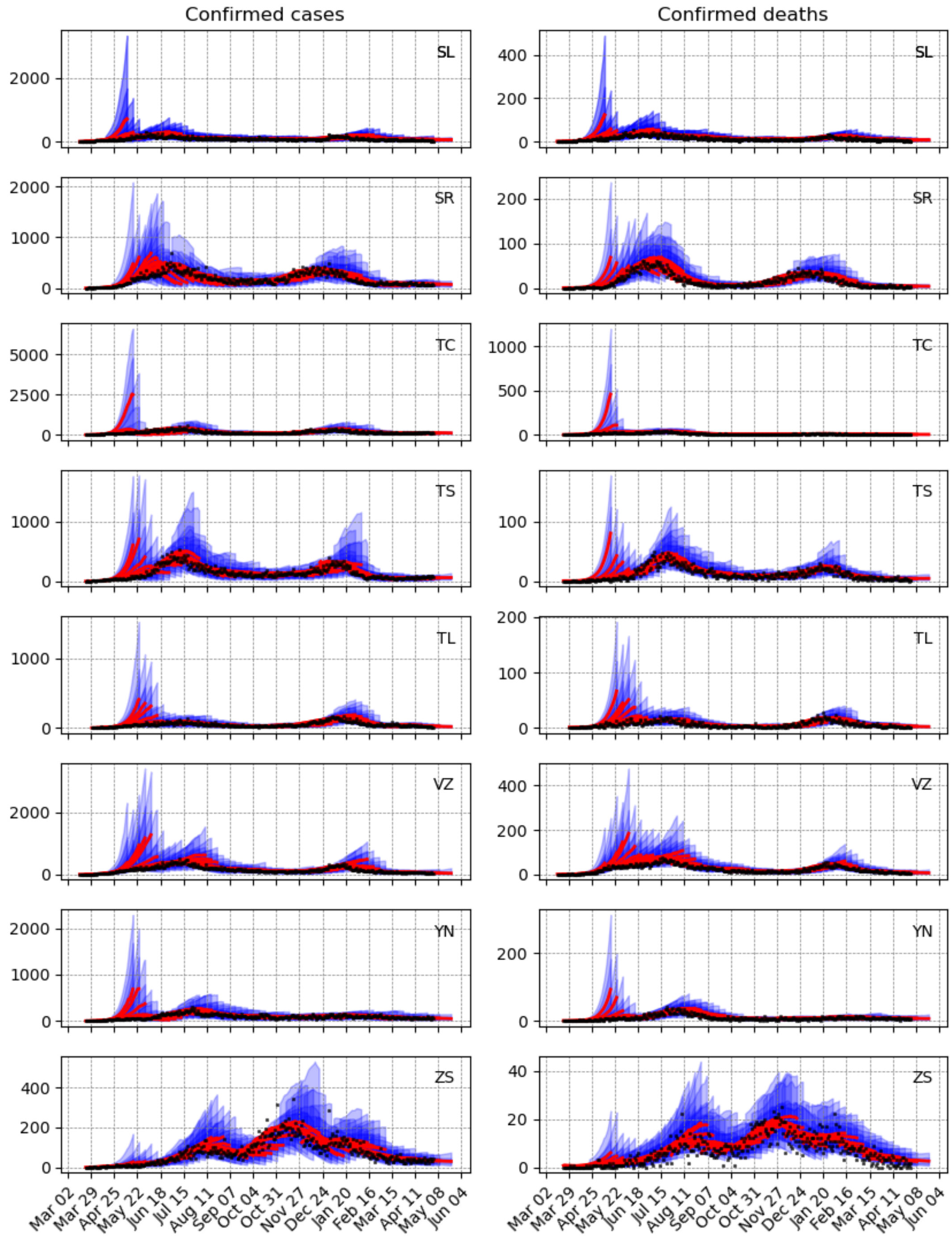


Figure S13: **Forecast results** for all Mexican states, see Figure S10 for details.

S6 Forecast performance for all Mexican states and Mexico City's metropolitan area

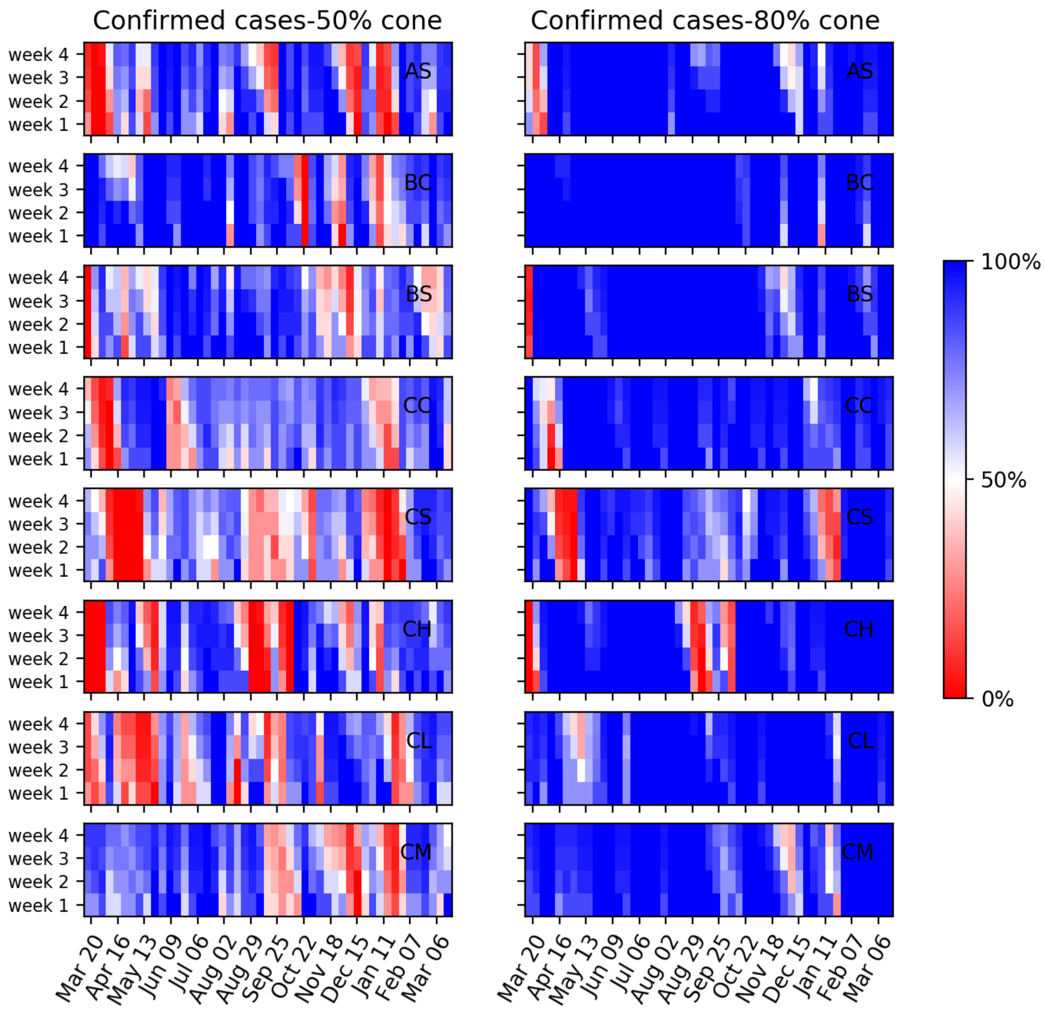


Figure S14: **Heatmap of the one to four week forecasts' performance of confirmed cases for all Mexican states and Mexico City's metropolitan area**. Left and right columns show the performance measure for the 50% (posterior interquartile range) and 80% (10% to 90% quantile range) prediction cones, respectively. Vertically colors are almost constant, showing low sensitivity concerning the prediction length. Prediction performance varies by state and in time, but we have good forecasting performance outside the exponential growth stages in the different pandemic waves, in most cases. See also Figures S15, S16 and S17 and also analogous forecasts' performance panels for confirmed deaths in Figures S18, S19, S20 and S21.

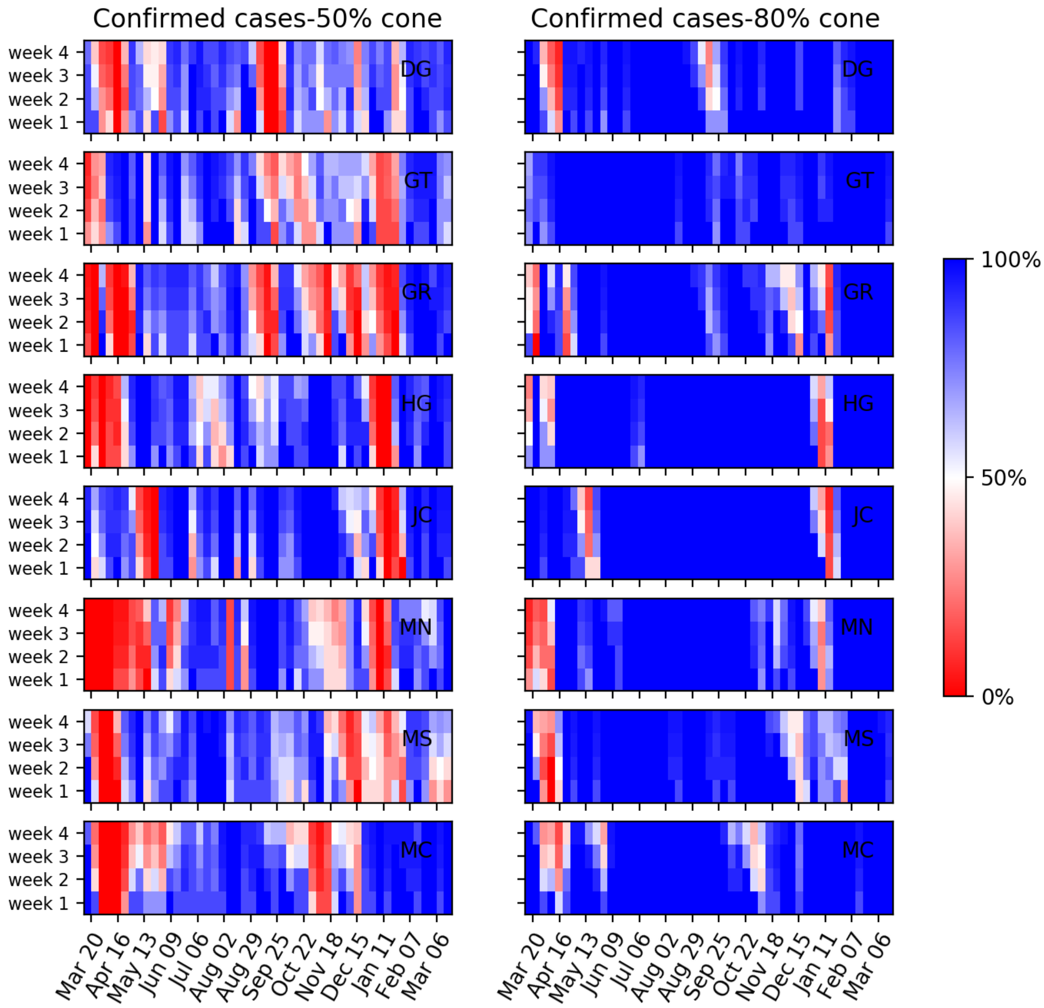


Figure S15: **Heatmap of the one to four week forecasts' performance of confirmed cases**, see Figure S14 for details.

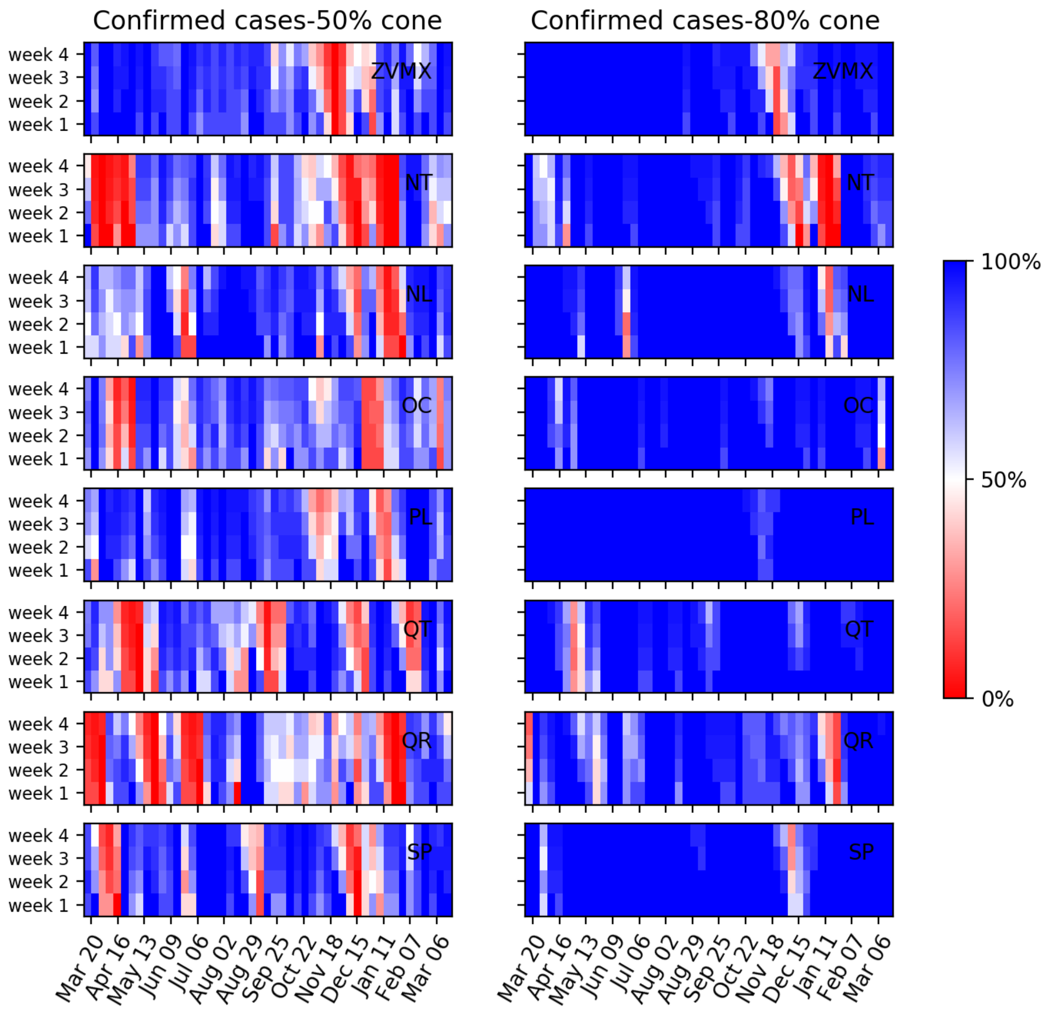


Figure S16: **Heatmap of the one to four week forecasts' performance of confirmed cases**, see Figure S14 for details.

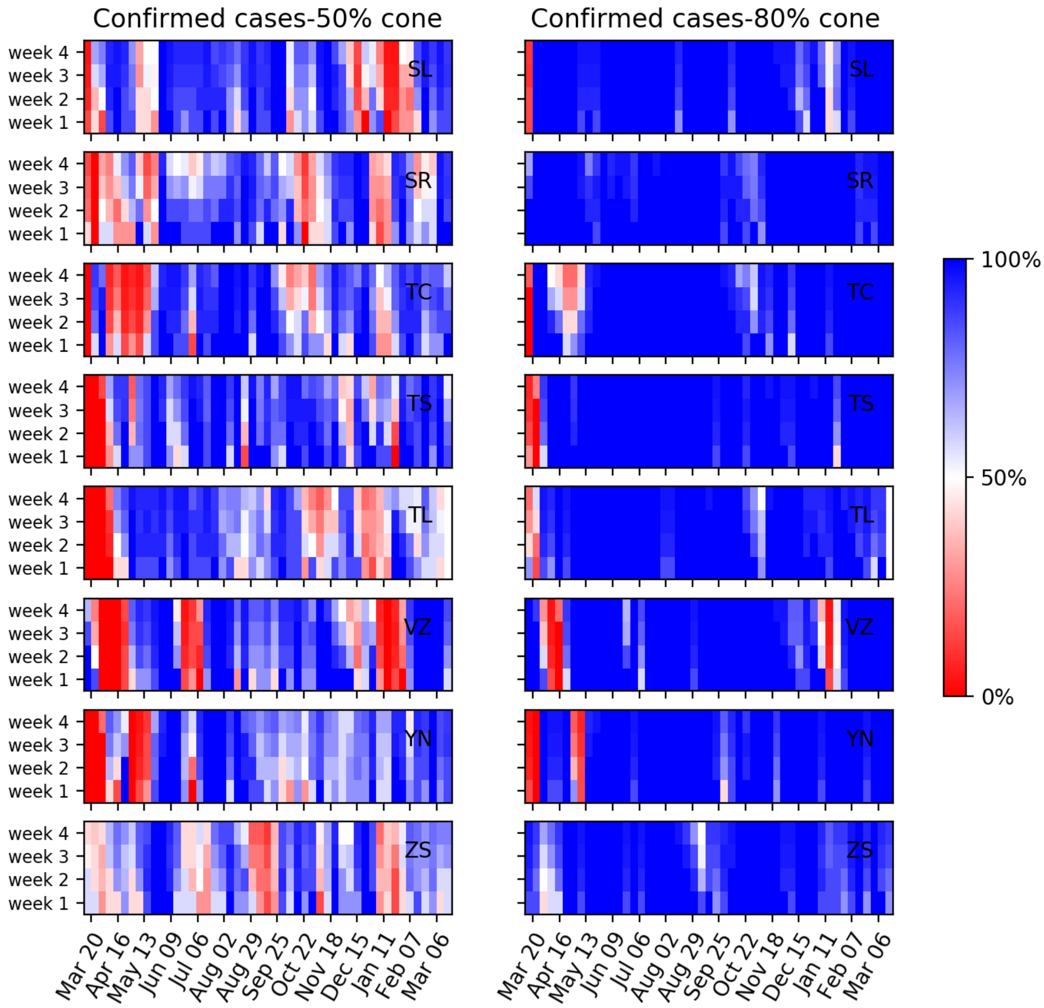


Figure S17: **Heatmap of the one to four week forecasts' performance of confirmed cases**, see Figure S14 for details.

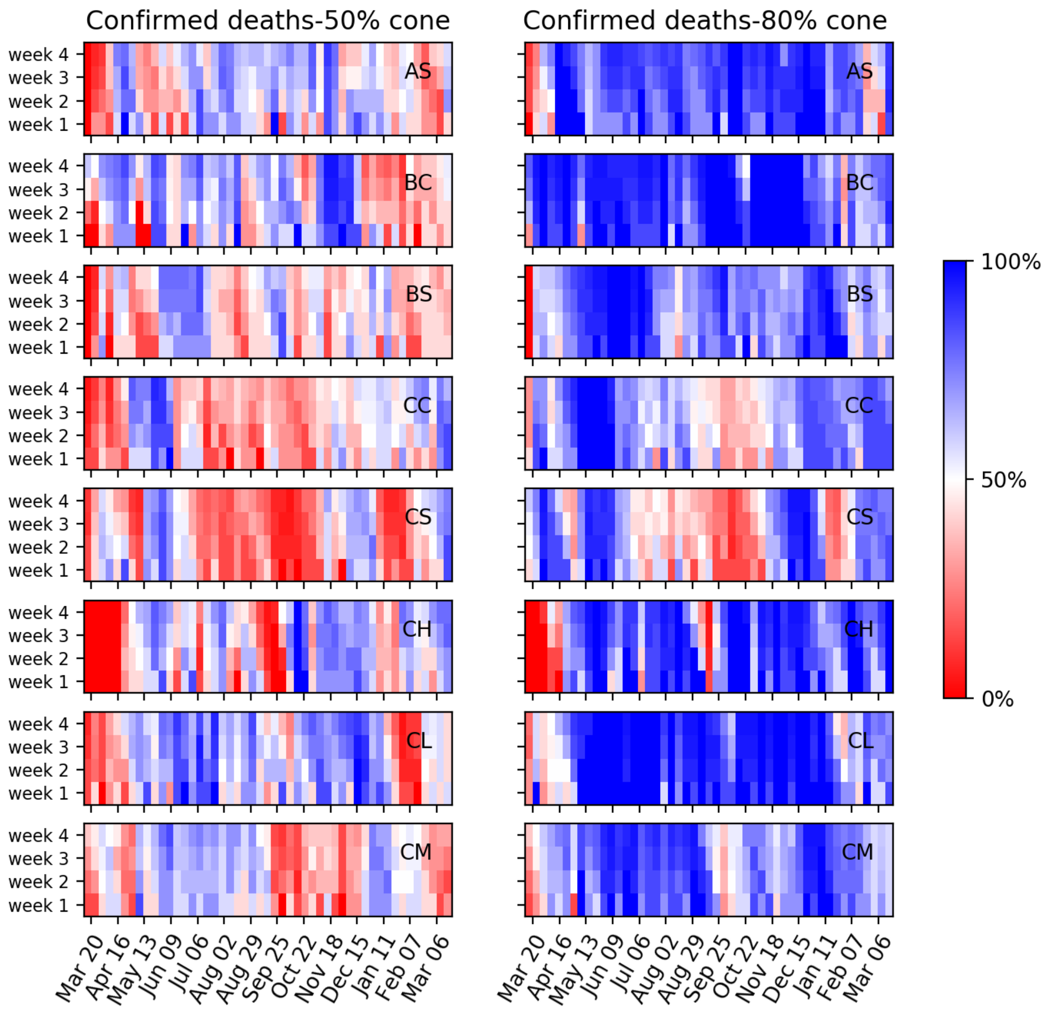


Figure S18: **Heatmap of the one to four week forecasts' performance of confirmed deaths for all Mexican states and Mexico City's metropolitan area**. Left and right columns show the performance measure for the 50% (posterior interquartile range) and 80% (10% to 90% quantile range) prediction cones, respectively. Vertically colors are almost constant, showing low sensitivity concerning the prediction length. Prediction performance varies by state and in time, but we have good forecasting performance outside the exponential growth stages in the different pandemic waves, in most cases. See also Figures S19, S20 and S21.

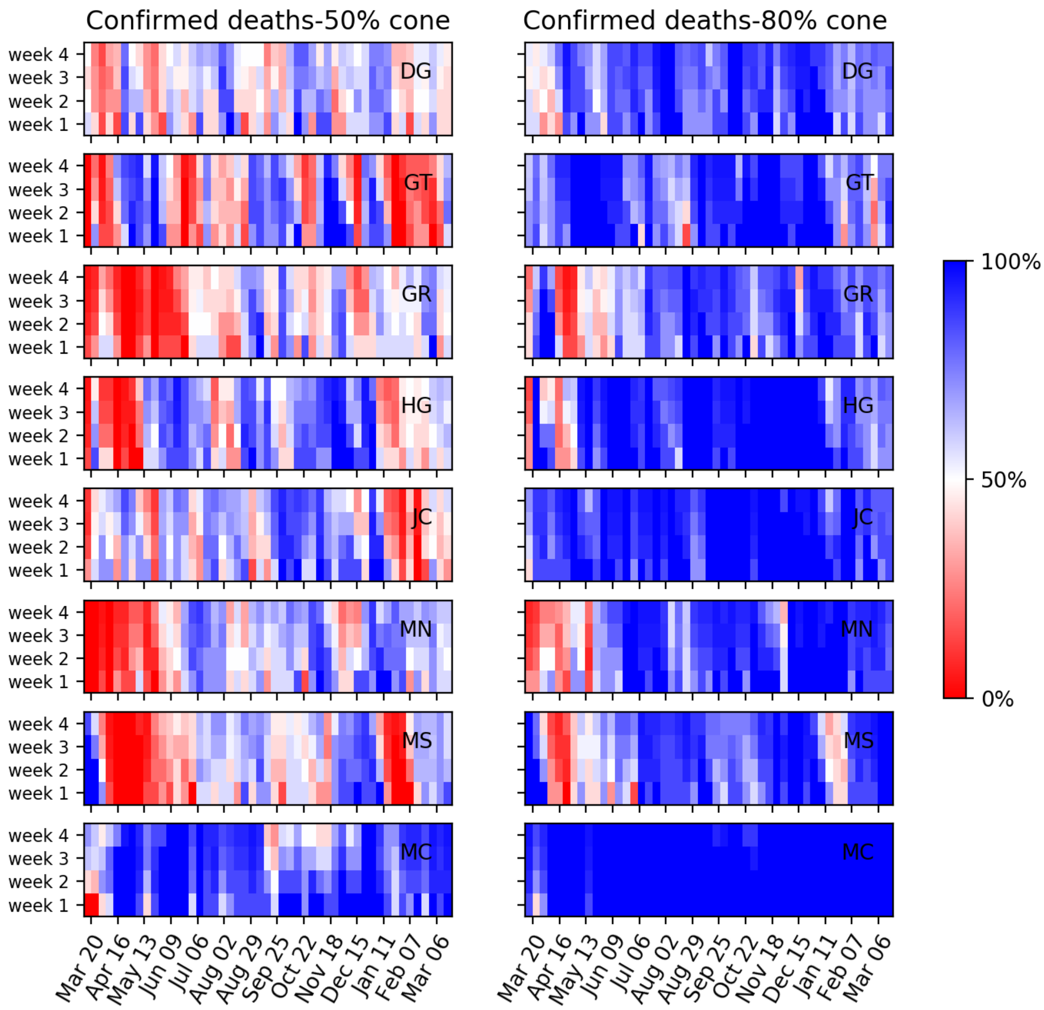


Figure S19: **Heatmap of the one to four week forecasts' performance of confirmed deaths**, see Figure S18 for details.

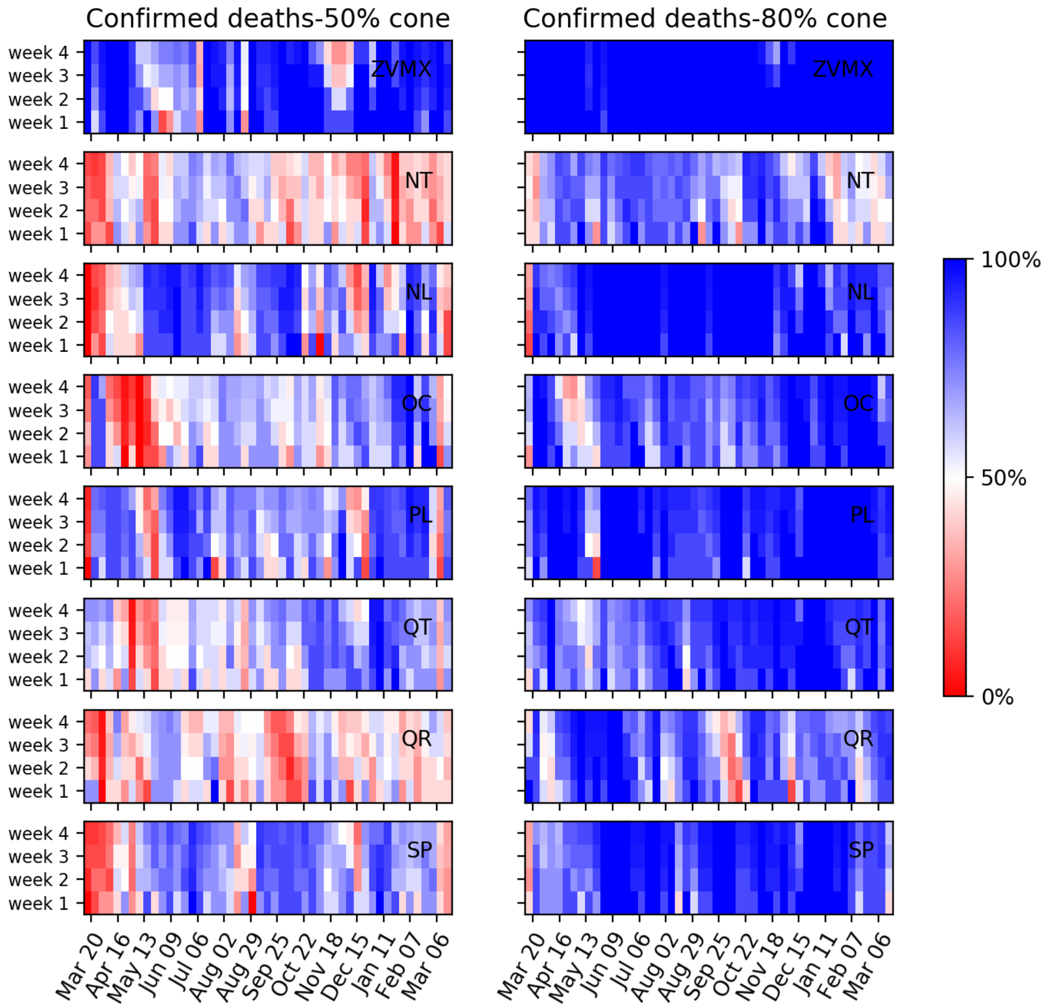


Figure S20: **Heatmap of the one to four week forecasts' performance of confirmed deaths**, see Figure S18 for details.

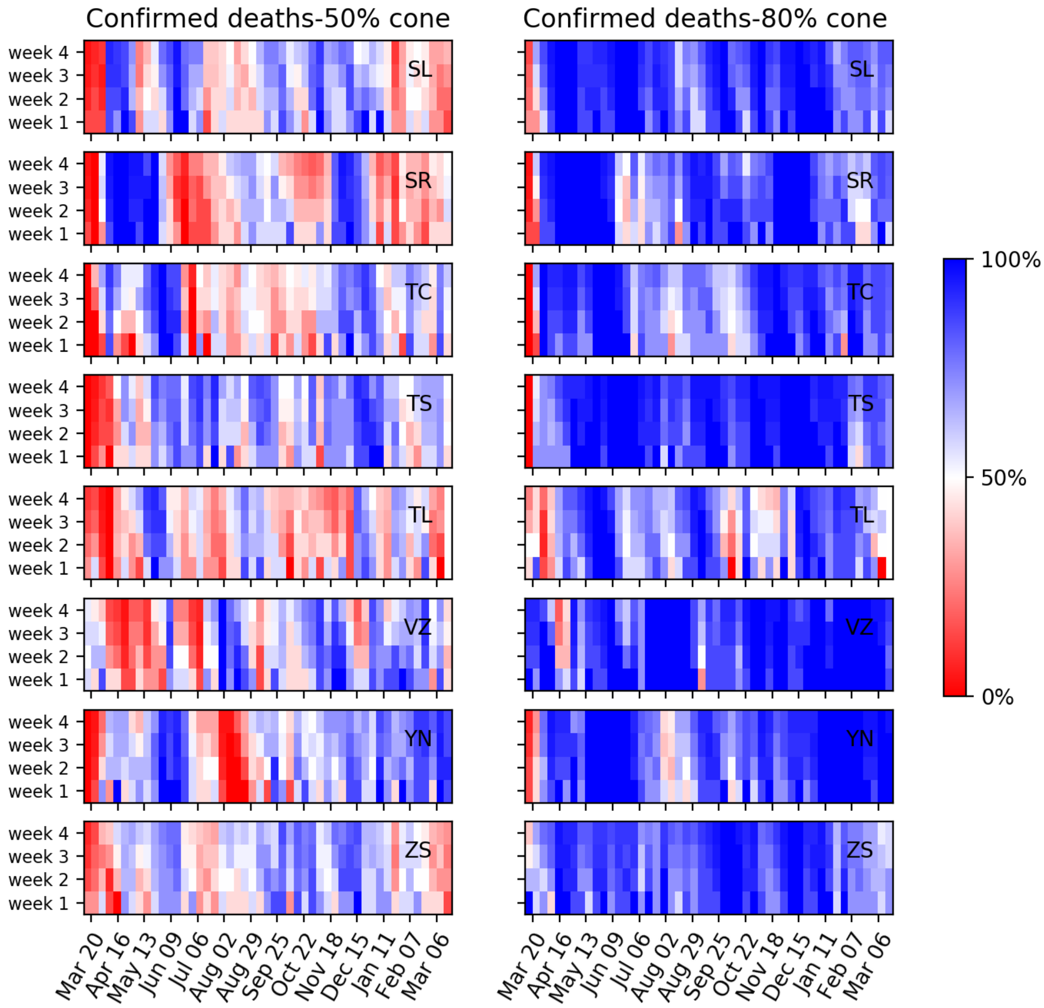


Figure S21: **Heatmap of the one to four week forecasts' performance of confirmed deaths**, see Figure S18 for details.

S7 MCMC convergence

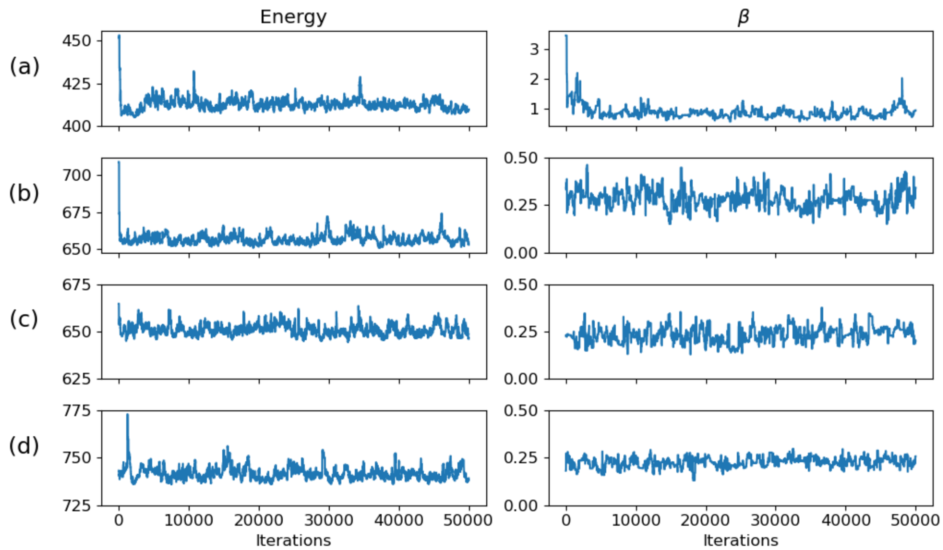


Figure S22: Marginal posterior for the Mexico City Metropolitan Area forecast. After 50,000 MCMC samples, trace plot of the minus logarithm of the posterior distribution (Energy) and the marginal posterior distribution for the contact rate (β) in the periods (a) March 8 - April 5, 2020, (b) May 3 - May 31, 2020, (c) June 7 - July 5, 2020, and (d) December 20, 2020 - January 17, 2021.

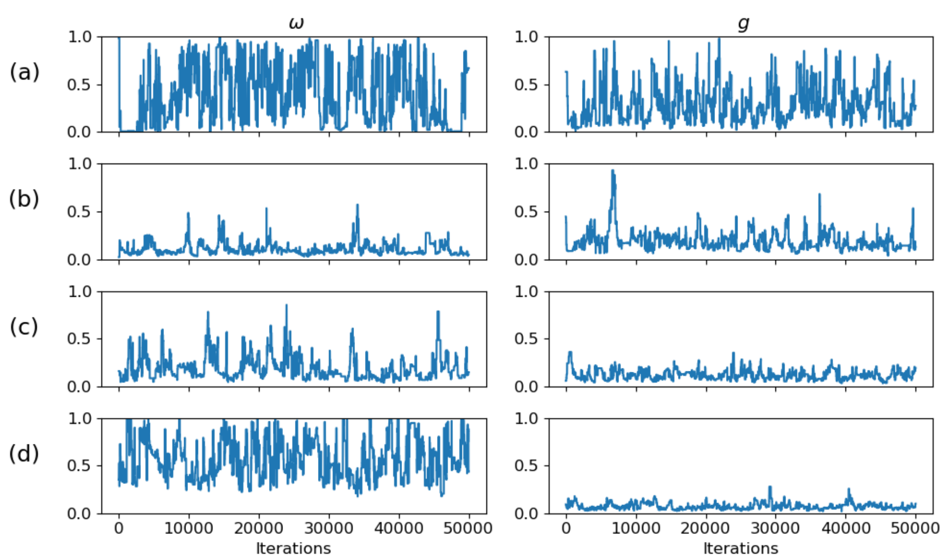


Figure S23: Marginal posterior for the Mexico City Metropolitan Area forecast. After 50,000 MCMC samples, trace plot of the marginal posterior distribution for the effective population proportion (ω) and the fraction observed infected individuals dying in the periods (g) in the periods (a) March 8 - April 5, 2020, (b) May 3 - May 31, 2020, (c) June 7 - July 5, 2020 and (d) December 20, 2020 - January 17, 2021.