Preprints.org

Article

Inference on a Multi-Patched Epidemic Model with Partial Mobility, Residency and Demography: The Case of 2020 COVID-19 Outbreak in Hermosillo, Mexico

Albert Orwa Akuno [*] , L. Leticia Ramírez-Ramírez , Jesús F. Espinoza

Article

Inference on a Multi-Patched Epidemic Model with Partial Mobility, Residency and Demography: The Case of 2020 COVID-19 Outbreak in Hermosillo, Mexico

Albert Orwa Akuno [1],* [ID], L. Leticia Ramírez-Ramírez [2] and Jesús F. Espinoza [3]

1   Centro de Investigación en Matemáticas A.C., México.; albert.akuno@cimat.mx
2   Centro de Investigación en Matemáticas A.C., México.; leticia.ramirez@cimat.mx
3   Universidad de Sonora, México.; jesusfrancisco.espinoza@unison.mx
*   Correspondence: albert.akuno@cimat.mx

Abstract: Most studies modelling population mobility and the spread of infectious diseases, particularly using meta-population-multi-patched models, tend to focus on theoretical properties and numerical simulations of such models. There is relatively scanty literature published on fit, inference and uncertainty quantification on epidemic models with population mobility. In this research, we have used three estimation techniques to solve an inverse problem and quantify its uncertainty on a human mobility-based multi-patched epidemic model, using mobile phone sensing data and COVID-19 confirmed positive cases in Hermosillo, Mexico. First, we have utilized a Brownian bridge model using mobile phone GPS data to estimate residence and mobility parameters of the epidemic model. In the second step, we have estimated the optimal model epidemiological parameters by deterministically inverting the model using a Darwinian inspired evolutionary algorithm (EA) known as the genetic algorithm (GA). The third part of the analysis involves performing inference and uncertainty quantification on the epidemic model using two Bayesian Monte Carlo sampling methods: t-walk and Hamiltonian Monte Carlo (HMC). The results show that the estimated model parameters and incidence adequately fit the observed daily COVID-19 incidence in Hermosillo. Moreover, the estimated parameters from HMC result into large credible intervals, improving their coverage for the observed and predicted daily incidences. We also observe improved predictions when using multi-patch model with mobility against the single-patch model.

Keywords: epidemics; human mobility; inference; deterministic inversion; Bayesian inference

---

1. Introduction

The mathematical modeling of infectious diseases has witnessed continuous growth as a research area in recent decades. This growth has been primarily motivated by the substantial impact that contagious diseases have on various crucial domains, including social, health, and economic development. [1–5].These domains are intricately linked to human population interaction, which serves as a key driving factor.

Indeed, works such as [6] and [7], raised awareness on the fact that the emergence and re-emergence of infectious diseases had historically been neglected as a significant threat to public health. Understanding the propagation dynamics of infectious diseases has thus become a central issue for epidemiologists and theoreticians [8–12]. Such knowledge would subsequently enable the formulation of effective infection mitigation policies. These policies, in turn, would contribute to maintaining a healthy human population, thereby significantly bolstering the economic development of their respective nations.

Since its advent and declaration as a pandemic in March 2020 by the WHO, COVID-19 has undeniably wrecked havoc and drastically impacted the lives and livelihoods of the global human population. In terms of significance, COVID-19 has been compared to the influenza pandemic of

1948-20, [8] but these are not yet equated in terms of severity. The 1918 influenza pandemic had an estimated death toll of 40 Million lives and infected 1/3 of the world population as compared to COVID-19 which so far has a death toll of 6.6 Million and confirmed cases of 633 Million (less than 10% of total human population) [13].However, the urgency for conducting further research on COVID-19 and other infectious diseases is underscored by the warnings issued by scientists in recent times regarding the imminent threat of another pandemic. These cautionary alerts emphasize the pressing need to proactively study and understand these diseases to be better prepared for any potential future outbreaks. [14]. These studies seek understanding the dynamics of the infection propagation, control and mitigation strategies. They encompass various aspects, including but not limited to model selection, statistical inference, uncertainty quantification and prediction using empirical epidemiological data. Such studies furnish epidemiologists and policymakers with valuable information in relation to the effectiveness of control and eradication measures. This knowledge facilitates the formulation of proactive response strategies in anticipation of a potential new pandemic crisis.

At the core of the transmission dynamics of infectious and communicable diseases like COVID-19 is human behaviour and interaction. Consequently, it behooves researchers and epidemiologists involved in the study of infectious diseases to accurately incorporate human behaviour as a critical component of the models used in modelling infectious diseases [15]. One significant human behavior that has consistently influenced person-to-person transmission of infectious diseases during outbreaks is mobility. [10,16–21]. The mobility and its travel patterns promote the creation of temporal social connections and networks that both local and global, play a crucial role in the spread of diseases. Human mobility, an undeniably crucial component of human existence and survival, not only plays a critical role in the spread and propagation of an epidemic [18,22–24], but it also crucially determines the spreading speed of infectious diseases, [23,25,26]. There is therefore no gainsaying that deep understanding of the role of human mobility in the spread of infectious diseases is central in the quest to have a clear view, for instance, of the impact of control strategies such as mobility restrictions [27–31].

Developments in epidemiological research have confirmed that traditional compartmental epidemic models are deficient in modeling strong heterogeneous disease spread, as they were formulated with the assumption of interaction of a well-mixed homogeneous population [15,32,33]. It is widely recognized that humans do not exhibit homogeneous mobility patterns, [15], which highlights the importance of incorporating heterogeneous population behavioral responses in epidemiological models as emphasized by [34]. The scientific literature offers a wide array of proposed models that aim to capture the heterogeneous mobility and dispersal patterns observed among humans. Many of these models consists of human groups of spatially separated populations which interact at some level (geo-meta-population-multi-patched framework), as evidenced by numerous references in the literature [e.g., 26,27,34–42]. Although modelling geo-spatial-temporal spread of infectious diseases is a complex task [43], models such as the ones in the mentioned references have crucially attempted to accurately capture human mobility in different settings. Nevertheless, there is still need to incorporate more realistic mobility patterns that reflect more complex mobility scenarios [17], which is an inevitable reality of human behaviour. In certain scenarios, it may be highly desirable to explore the relationship between the spread of infectious diseases and the proportion of time individuals spend in different locations during their travels [44]. This approach considers the impact of human movement on disease transmission and how it relates to the amount of time spent in locations, each having specific characteristics, such as the risk of infection level. These locations, that are homogeneous within them and possibly different between them, are usually referred in the literature as patches. Then, a single-patch model reduces to a model with a well-mixed homogeneous population where every susceptible individual can become infected with the same probability. The concept of multi-patches not only allows to consider locations with different characteristics but helps in capturing the spatial dynamics and interactions, produced by the human mobility, that influence the spread of infectious diseases.

Once an appropriate epidemic model has been formulated, a myriad of analysis relevant to epidemiologists, policy makers and public health officials can be conducted. These include but are not limited to theoretical analysis and numerical simulations [see e.g., 39,40,45–47], inference, uncertainty quantification and predictions [see e.g., 14,48–53]. Most studies modelling population mobility and the spread of infectious diseases, particularly using meta-population-multi-patched models, have tended to focus on theoretical and global properties and numerical simulations of such models. There is relatively scanty literature published on inference and uncertainty quantification on epidemic models with population mobility. To our knowledge, to this date, the only research work that has come close to performing inference, uncertainty quantification and predictions on a multi-patch epidemic model is [51] who developed and used a projected Stein Variational Gradient Descent (pSVGD) algorithm to model COVID-19 in the states of New Jersey and Texas in the US. However, their model did not incorporate population mobility as they modeled the dynamics and severity of COVID-19 inside and outside long-term care (LTC) facilities where the interaction was introduced through some contact matrix. Other works that have addressed the problem of statistical inference for epidemics on networks are [54] and [55]. These models consider families of networks of contact that remains constant through time. Filling the gap left by the conspicuous lack of literature on uncertainty quantification and Bayesian inverse problems on meta-population-multi-patched epidemic models, forms the central motivation of the present study. To achieve this objective, we conduct inference and uncertainty quantification using COVID-19 confirmed cases in Hermosillo, Mexico. We employ a multi-patched epidemic model that incorporates mobility, residency, and demography. This model was proposed and theoretically analyzed in [42].

Solving an inverse problem entails the utilization of both observed data and a model to deduce the parameter values that characterize the dynamics producing the observations [56–58]. However, employing models that adequately capture the heterogeneity and complexity of human behavior, such as mobility, in modeling the spread of infectious diseases, introduces numerous parameters [51,59] . This, in turn, presents a multitude of challenges when addressing inverse problems (fit) and conducting inference (uncertainty quantification). To confront the problems originated by the complex system and its high dimensional parameter space, we divide the problem into two parts. The first one is related to the mobility parameters that describe the proportion of individuals who move, and the time spent in each of the sub regions/patches. The second one, is associated to the numerical fit or statistical inference of the initial conditions and the parameters for the infectious agent and individual evolution of infected individuals.

The rest of this article is organized as follows: In Section 2, we present and describe the epidemic model on which the present analysis is centered. Section 3 discusses how the COVID-19 data used in this work was obtained. In Section 4, a discussion on the estimation of the mobility parameters and residence times of the epidemic model using mobile phone sensing data and the Brownian bridge model, is presented. We elucidate the formulation of the deterministic inversion model and the corresponding results in Section 5. Section 6 entails Bayesian model formulation, inference and the obtaining results. Finally, we give concluding remarks and possible extensions of the present work in Section 7.

2. The multi-patch model with mobility, residency and demography

The emergence and re-emergence of infectious diseases over the past several decades has led to the formulation, by researchers and epidemiologists, of numerous mathematical models that incorporate varied and distinct disease propagation dynamics. The central purpose of such models is to provide fundamental quantitative information and utilitarian guidelines in disease outbreak management and mitigation policy formulations. Provision of such useful information by such models is anchored on their ability to incorporate and capture realistic situations and disease transmission parameters. One such parameter, that has played a crucial role in the spread of infectious diseases, is human population mobility [16–18,21,39,40,60].

Several aspects of human history, such as social and economic, have been significantly impacted by the effects of the spread of infectious disease viruses among mobile sectors of human civilization [21]. Undoubtedly, the effects of the current COVID-19 pandemic is a demonstration of the devastating consequences of the interchange between human population mobility and transmission of infectious diseases. For instance, in December 2019, Wuhan, China was the only understood geographical boundary of COVID-19 [61–65]. However, it can be postulated that human mobility within and between cities in China and around the world aided the spread of the pathogen to a significant portion of the planet earth. In a span of three months, the pathogen had posed monumental health, social and economic stress to the point of being declared a pandemic by the WHO on March 11, 2020. Even though the exact nature of the contagion remained unclear, there was recognition of the undeniable interplay between the swift spread of the virus and human mobility through the universally known modes of travel [16,66–69]. Indeed, besides aiding the spread of COVID-19, such close relationship between population mobility and the spread of infectious diseases has the potential of shaping the emergence of future epidemics. Consequently, in order to formulate robust guidelines to manage both the current ongoing COVID-19 and possible future pandemics, there is dire need to construct appropriate epidemic models that accurately capture population mobility behaviour.

Human mobility, being a geo-spatio-temporal phenomena [17,43,70,71], can be modelled by meta-population-multi-patched models [4,34,39,40,42,44], as traditional homogeneous compartmental models are incapable of capturing such a strong heterogeneous human behaviour [15,17,34,51]. Once such models have been constructed, a wide spectrum of quantitative analysis that leads to deep understanding of the reciprocity of mobility and the spread of infectious diseases can be conducted. For instance, researchers and epidemiologists may focus on theoretical properties and numerical simulations of the model or parameter estimation, uncertainty quantification and predictions. In this research, we focus on the latter, using COVID-19 confirmed cases in Hermosillo, Mexico and a multi-patched SEIRS compartmental epidemic model with residency and demography. The model was proposed, and its theoretical properties analyzed in [42]. It is crucial to emphasize that in for this model human mobility encompasses more than just travel between two locations. It encompasses the multiple temporal interactions that individuals experience while traveling to different destinations.

The following ordinary differential equations (1) constitute the full form of the said multi-patch non-linear modelThis model extends the traditional compartmental SEIR model but in also divides the population into different compartments based on their disease status: Susceptible (S), Exposed (E), Infected (I), and Recovered (R). Most individuals start in the susceptible compartment (S) but when one of them comes into contact with an infectious individual, he moves from the susceptible compartment into the exposed compartment (E), meaning that he has been infected but is not yet infectious. This individual leaves the E compartment to enter the I compartment where he can transmit the disease to susceptible individuals. Over time, infected individuals either recover from the disease or may succumb to it. Once individuals recover, they move to the recovered compartment (R), gaining immunity to the disease. In the proposed model, the population dynamic also considers births (into the S compartment) and deaths (not produced by the disease).

The parameters of model (1) are described in Table 1. We refer the reader to [42] for a full description of the arguments leading up to this model formulation.

$$
\begin{cases}
\dot{S}_i &= \Lambda_i - \beta_i(1-\alpha_i)S_i \frac{(1-\alpha_i)I_i + \sum_{k=1}^n \tilde{p}_{ki}I_k}{(1-\alpha_i)N_i + \sum_{k=1}^n \tilde{p}_{ki}N_k} - \sum_{j=1}^n \left(\beta_j \tilde{p}_{ij} S_i \frac{(1-\alpha_j)I_j + \sum_{k=1}^n \tilde{p}_{kj}I_k}{(1-\alpha_j)N_j + \sum_{k=1}^n \tilde{p}_{kj}N_k} \right) \\
&\quad - \mu_i S_i + \tau_i R_i \\
\dot{E}_i &= \beta_i(1-\alpha_i)S_i \frac{(1-\alpha_i)I_i + \sum_{k=1}^n \tilde{p}_{ki}I_k}{(1-\alpha_i)N_i + \sum_{k=1}^n \tilde{p}_{ki}N_k} + \sum_{j=1}^n \left(\beta_j \tilde{p}_{ij} S_i \frac{(1-\alpha_j)I_j + \sum_{k=1}^n \tilde{p}_{kj}I_k}{(1-\alpha_j)N_j + \sum_{k=1}^n \tilde{p}_{kj}N_k} \right) \\
&\quad - (\kappa_i + \mu_i)E_i \\
\dot{I}_i &= \kappa_i E_i - (\gamma_i + \psi_i + \mu_i)I_i \\
\dot{R}_i &= \gamma_i I_i - (\tau_i + \mu_i)R_i
\end{cases}
\tag{1}
$$

where $\tilde{p}_{kj} = \alpha_k p_{kj}$, $i = 1, 2, \cdots, n$ and n is the number of patches.

Table 1. Parameters description.

Parameters	Description
α_i	The proportion of individuals that leave their residence patch i
p_{ij}	The proportion of time that an individual from patch i spends in patch j, given that is one individual that leaves its residence patch
Λ_i	Recruitment of Susceptible individuals in Patch i
β_j	Instantaneous risk of infection in Patch j
μ_i	Per capita natural death rate in Patch i
γ_i	Per capita recovery rate of individuals in Patch i
τ_i	Per capita loss of immunity rate
ψ_i	Per capita disease induced death rate of Patch i
κ_i	Per capita rate at which the exposed individuals in patch i becomes infectious

Model (1) is limited to infectious diseases with single strain mutation. That is, in the formulation of the model, [42] assume that the recovered individuals in each patch are conferred with partial immunity, which they lose at a rate τ_i. These individuals then reenter the susceptible compartment but it is only in relation to the same strain of the pathogen. However, many infectious diseases, including COVID-19, which is the center of this research, have several pathogenic strains whose dynamic properties should be studied for complete sensible management of emerging and re-emerging outbreaks with multiple variants. To this end, model (1) can be improved to include multiple COVID-19 disease strains.

Works such as [72–76] have considered models that incorporate several disease strains and mutations, with some of them focusing on this property for COVID-19 disease. Even though these works are mostly single-patch based, they provide valuable insights and useful starting point to extend model (1) to this important direction.

3. The data

Based on an institutional agreement between the University of Sonora (UNISON) and the Secretaría de Salud del Estado de Sonora (SSA), a collaborative effort was undertaken to provide accurate and reliable information to the population of Sonora during the sanitary emergency. As part of this initiative, researchers from the Mathematics Department at UNISON were granted access to georeferenced data pertaining to COVID-19 cases in Hermosillo, Mexico. The dataset encompassed the period from January 1, 2020, to September 6, 2020. It is important to emphasize that the usage of this data is strictly limited to academic activities in accordance with UNISON's agreements with public institutions, and utmost care has been taken to safeguard sensitive and private information.

Furthermore, on September 14th, 2020, UNISON entered into a specific agreement, referenced as 12615-5700000-000412, to provide statistical consulting services to LUMEX CONSULTORES, S.C. Esteemed research professors from the Mathematics Department, including co-author J.F.E., provided these consulting services. The dataset provided by LUMEX CONSULTORES, S.C. comprised a comprehensive collection of approximately 80 million records, each containing the GPS position and timestamp of nearly 300,000 devices. These devices were geographically located within Hermosillo city between September 21st, 2020, and November 15th, 2020. The usage of this data was explicitly authorized for non-profit and academic activities, with appropriate credits duly acknowledged. Notably, user privacy was carefully protected through the meticulous anonymization of the data, employing distinct alphanumeric IDs for each individual device.

In a separate study, Ramirez et al. (2022) utilized mobile phone GPS data from September 21st, 2020, to November 15th, 2020, to estimate mobility parameters and residence times in Hermosillo's

582 urban AGEBs (Basic Census Geographical Units) [4]. In alignment with our research objectives of modeling the number of COVID-19 cases in four key zones within Hermosillo, which aggregate AGEBs as illustrated in Figure 1, we employed the geo-referenced and GPS information to obtain the number of COVID-19 cases and estimate mobility and residence patterns within and between these zones.

The two geo-referenced databases utilized in this study do not exhibit a temporal overlap, thereby limiting the feasibility of employing the complex model (1), which relies on the estimation of numerous epidemiological parameters. To address this problem, we have elected to use the weighted global COVID-19 data by zonal proportions.

The global confirmed positive COVID-19 cases in Hermosillo (Figure 3) are available in [77], a website managed by Consejo Nacional de Ciencias y Technología (CONACYT), since 2020-02-26. We use the weighted global cases by considering the zonal proportions of the global COVID-19 data, a selection we justify as in Figure 2, from which we can observe that the proportions of the available zonal COVID-19 data are approximately stable from June 2020. We therefore project that the data will have the same behaviour within the period for which the mobility parameters and mobility residence times were estimated. The global (Figure 3) and derived zonal COVID-19 cases, however, exhibits high variability. We therefore smoothen it using 7-days moving averages and use the smoothed version for the rest of the analysis.
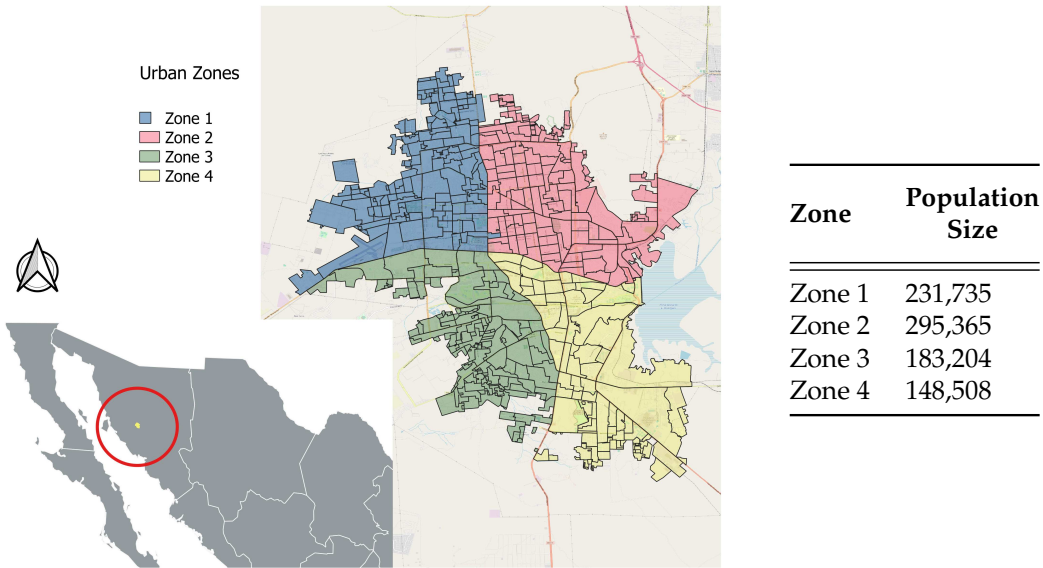

Zone	Population Size
Zone 1	231,735
Zone 2	295,365
Zone 3	183,204
Zone 4	148,508

**Figure 1.** Hermosillo city and Urban Zones

Most models that have been proposed to analyze and predict COVID-19 incidence, since the emergence of SARS-CoV-2, assume that all infected individuals are observed, which is inexact due to under-reporting of disease incidence. This assumption undermines the accuracy of the models and their predictions. Emerging diseases, such as COVID-19, Typhoid fever, Hepatitis B, Epstein-Barr virus and Zika, that present a large fraction of asymptomatic pathogen carriers, are often characterized by incidence under-reporting during disease surveillance [78–81]. Coupled with asymptomatic and subclinical carriers, a phase that is prevalent with COVID-19 disease [82–85], another source of under-counting of disease incidence is lack of systematic testing. It is therefore necessary to account for under-reporting when modeling and fitting COVID-19 disease incidence, as failure to do so will lead to underestimation of epidemiological characteristics of the disease, especially the transmission parameter [86].

In modelling COVID-19 observed incidence for eight American countries, [86] reported an under-reporting of Mexico COVID-19 cases by a factor of 15. This acute cases identification problem

corroborated the observation in [87], that Mexico has one of the lowest numbers of COVID-19 tests performed per reported cases. We believe that this under-reporting cascaded down to the local Mexico states and cities like Hermosillo. In order to account for this under-reporting, we inflated the smoothed observed COVID-19 incidence, which was obtained from the weighting procedure previously explained, by a factor of 15 to obtain the daily incidence data. We subsequently consider the inflated data (see Figure 4) as the true observed daily COVID-19 incidence in Hermosillo, Mexico, and use it for inference in this research.
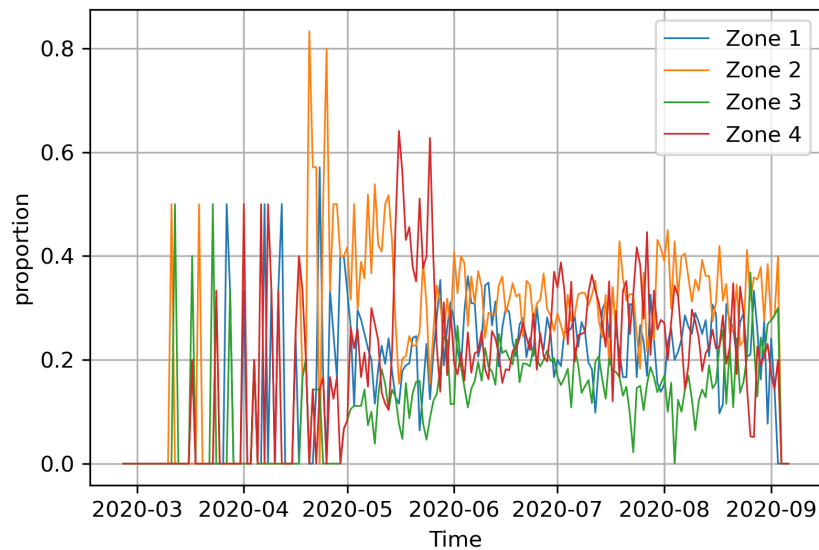
Figure 2. Proportions of zonal COVID-19 daily (YYYY-MM) cases.
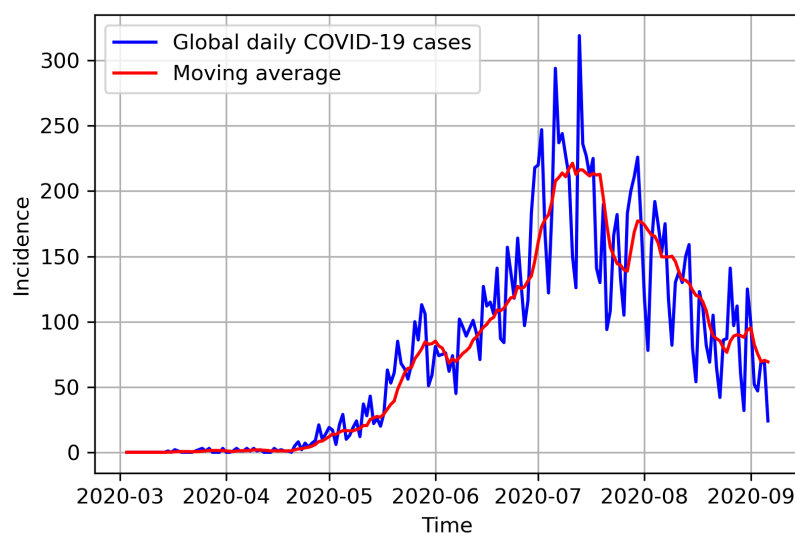
Figure 3. Hermosillo's COVID-19 daily (YYYY-MM) cases. Source: https://datos.covid-19.conacyt.mx
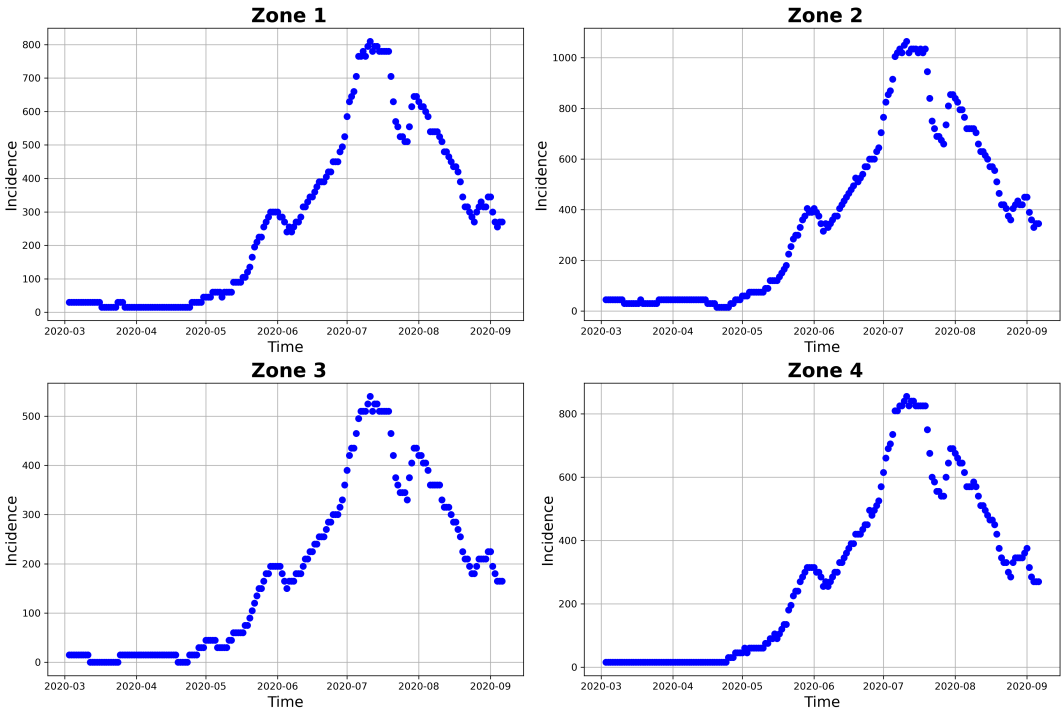
Figure 4. Observed Hermosillo COVID-19 daily (YYYY-MM) cases. The incidences (blue dots) are inflated by a factor of 15.

In addition to using real count data for model fitting and uncertainty quantification, we also utilized simulated incidence data generated from the model. Our goal was to assess the parameter and initial conditions identifiability of the inverse problem and statistical inference under the model. To this end, we used epidemiological parameters of COVID-19 gathered from various literature as presented in Table 2, and simulated from model (1), daily observed incidences for each zone in Hermosillo. For this simulation, we used epidemiological parameter values close to the ones used by [88], who modelled lockdown relaxation period of COVID-19 in Hermosillo. The parameters and initial condition values used in the simulation can be seen in Table 3. The latter were based on the total population of each zone. After the simulation of the zonal incidence from the model, we added a normal noise $e_t \sim \mathcal{N}(I_i(t), \sigma = 100)$, to incorporate variability in the model data. The final simulated incidence for each zone in Hermosillo is presented in Figure 5.
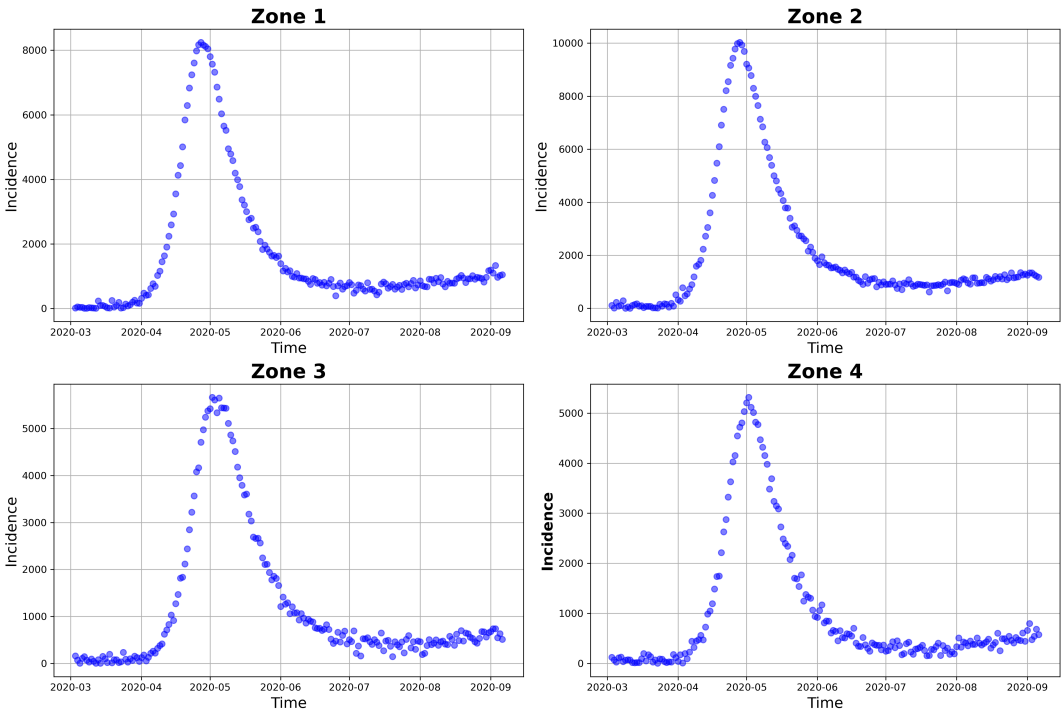
Table 2. Epidemiological parameter ranges/values and their associated literature

Parameters	Range/Value	Literature
β	0.0616 - 1.68	[88–90]
$1/\kappa$	2 - 14	[88,91–94]
$1/\gamma$	5 - 15.6	[49,88,90,95,96]
μ	$0.06/(1000 \times 365)$	[97]
ψ	$(0.0047+0.0030)/2$	[98]
$1/\tau$	90 - 300	[99–108]

Table 3. Model (1) parameter values used in the simulation of observed daily incidence in Figure 5.

Epidemiological		Initial conditions	
Parameters	Values	Parameters	Values
β_1	1.3000	$E_{1,0}$	10
β_2	1.4000	$I_{1,0}$	1
β_3	0.9500	$Y_{1,0}$	1
β_4	0.8000	$E_{2,0}$	20
κ_1	0.0833	$I_{2,0}$	2
κ_2	0.0714	$Y_{2,0}$	2
κ_3	0.0741	$E_{3,0}$	5
κ_4	0.1000	$I_{3,0}$	0
γ_1	0.1667	$Y_{3,0}$	0
γ_2	0.1429	$E_{4,0}$	5
γ_3	0.1818	$I_{4,0}$	0
γ_4	0.2000	$Y_{4,0}$	0

Figure 5. Simulated Hermosillo COVID-19 daily (YYYY-MM) cases. The daily model incidence (in blue) have been simulated using the parameters values in Table 3, together with μ and ψ values in Table 2 and $\tau = 1/180$.

4. Estimation of the mobility and Residence time parameters

The present study involves solving an inverse problem and performing uncertainty quantification on system (1), using 2020 COVID-19 confirmed cases in Hermosillo, Mexico. Such data may not be sufficient in estimating all the mobility and infection parameters, as well the initial conditions present in the system. Moreover, simultaneous estimation of several parameters oftentimes leads to parameter non-identifiability problems [109–115]. As such, we use the mobile phone sensing data used in [4] for

purposes of estimating the mobility vector α_i, $i = 1, 2, \cdots, n$ and the mobility residence time matrices $\mathbb{P} = (p_{ij})$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, n$ which we will incorporate into system (1). Using mobile consecutive *pings*, [4] estimated the mobility parameters vectors and mobility residence time matrices of 582 urban AGEBs in Hermosillo, Mexico, for three periods, each divided into two parts. In this research, we use the same GPS *pings* during the *Third Period – First Part* (2020-09-21 to 2020-10-11). Using this information, we estimate the mobility and residence time matrix of four zones in Hermosillo. Since we are interested in evaluating the forecasting (after November 2020) under the model, this period was selected as its start date was closer to the last period of considered COVID-19 cases. We note that we can also use mobility data from other periods and parts, and employ a similar procedure that we describe next, to estimate mobility and residence times for the four zones, within the period and part of interest.

Solving the inverse problem and performing uncertainty quantification on system (1) for a total number of AGEBs (patches) poses a logical set-up challenge. Thus, using well-defined geographical demarcations of Hermosillo, we group the 582 AGEBs into four main zones. These zones, together with their corresponding total population, are shown in Figure 1. The mobility parameter vector α_i, and residence times p_{ij}, $i = 1, 2, 3, 4$, $j = 1, 2, 3, 4$ were estimated by searching for individuals who activated at least 11 pings and left their AGEBs within and between the zones. For the individual AGEBs, [4] estimates the probability density functions at position z for any individual originating from AGEB/patch r as

$$\hat{h}_r(z) = \frac{1}{n_r} \sum_{j=1}^{n_r} h_{rj}(z; \hat{\sigma}_{rj}, \delta), \tag{2}$$

where n_r is the patch population size, h_{rj} is the density of the expected time of the *j*-th individual originating from the *r*-th patch, $\hat{\sigma}_{rj}$ is the maximum likelihood estimate of the individual standard deviation of the mobility of the *j*-th resident from the *r*-th patch, and δ^2 is the variance of the geolocation error. After grouping and renaming the composing AGEBs of each of the four zones, we have

$$n_1^{[\ell]} \hat{h}_1^{[\ell]}(z) = \sum_{j=1}^{n_1} h_{1j}^{[\ell]}\left(z; \hat{\sigma}_{1j}^{[\ell]}, \delta\right), \cdots, n_{n_\ell}^{[\ell]} \hat{h}_{n_\ell}^{[\ell]}(z) = \sum_{j=1}^{n_1} h_{n_\ell j}^{[\ell]}\left(z; \hat{\sigma}_{n_\ell j}^{[\ell]}, \delta\right), \tag{3}$$

where the subscript index $[\ell]$ indicates that the said quantity is in zone ℓ and n_ℓ represents the total number of AGEBs in the ℓ-th zone.

From (3), we can obtain the estimated probability density function at location z for any individual originating from patch r in zone ℓ as

$$\hat{h}_\ell^*(z) = \frac{1}{N^{[\ell]}} \sum_{r=1}^{n_\ell} \sum_{j=1}^{n_r} h_{rj}^{[\ell]}\left(z; \hat{\sigma}_{rj}^{[\ell]}, \delta\right). \tag{4}$$

where $N^{[\ell]} = \sum_{j=1}^{n_r} n_j^{[\ell]}$. Then the expected occupation time in A of a resident of zone ℓ can thus be obtained, from (4), as

$$\hat{P}_\ell(A) = \int_A \hat{h}_\ell^*(z) dz = \frac{1}{N^{[\ell]}} \int_A \sum_{r=1}^{n_\ell} \sum_{j=1}^{n_r} h_{rj}^{[\ell]}\left(z; \hat{\sigma}_{rj}^{[\ell]}, \delta\right) dz. \tag{5}$$

As has been mentioned, for this research, we consider the mobility parameter and the mobility residence matrix for the *Third Period – First Part* (2020-09-21 to 2020-10-11), which we compute, using the above procedure, and obtain

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.9668, & 0.9265, & 0.9692, & 0.9680 \end{pmatrix}^{\top} \quad \text{and} \quad \mathbb{P} = \begin{pmatrix} 0.8164 & 0.1289 & 0.0372 & 0.0175 \\ 0.1222 & 0.8119 & 0.0215 & 0.0444 \\ 0.0722 & 0.0504 & 0.7293 & 0.1481 \\ 0.0313 & 0.1166 & 0.1278 & 0.7243 \end{pmatrix}.$$

5. Deterministic inversion

Due to their exponential growth in time, the state variables of the dynamical system (1) are extremely sensitive to the input parameters. This sensitivity cascades uncertainty into the pre-inferred parameter values, which, in turn, may result into blow-up or significant uncertainty in the model outputs. To have a robust and desirable model analysis, it is imperative to avoid such model blow-ups by first determining the optimal model parameters that fits the measured data. Moreover, prior knowledge of point parameter estimates is necessary for meaningful uncertainty quantification. To this end, we begin by deterministically inverting system (1) in this section, and then subsequently perform uncertainty quantification on the inferred point parameters in the next section. Following is an outline of the formulation of the inversion problem and the results.

5.1. Formulation of the model to minimize

Let $\boldsymbol{x} = (\boldsymbol{S}, \boldsymbol{E}, \boldsymbol{I}, \boldsymbol{R}) \in (L^2(0, T))^{4n}$ where $T > 0$, $\boldsymbol{S} = (S_1, \cdots, S_n)^{\top}$, $\boldsymbol{E} = (E_1, \cdots, E_n)^{\top}$, $\boldsymbol{I} = (I_1, \cdots, I_n)^{\top}$ and $\boldsymbol{R} = (R_1, \cdots, R_n)^{\top}$, denote the state variables in system (1) for n patches. Let m_i be the dimension of patch i parameters to be estimated. Then, we have the total number of patch i parameters to be estimated, denoted as $\theta_i = \{\theta_{i1}, \cdots, \theta_{im_i}\}$. Consequently, the total number of parameters, $m = \sum_{i=1}^{n} m_i$, for the global system is $\theta = \bigcup_{i=1}^{n} \theta_i \in \mathbb{R}_+^m$. System (1) can be viewed as a mapping $\Psi(\theta) = \boldsymbol{x}$ where $\Psi : \mathbb{R}_+^m \longrightarrow (L^2(0, T))^{4n}$, (i.e., Ψ maps the parameters to be estimated to the state variables) and it defines an initial value problem of the form;

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}, \theta), \quad \boldsymbol{x}(t_0) = \boldsymbol{x_0} \in \mathbb{R}^n, \quad t_0 \in \mathbb{R}. \tag{6}$$

The function f in the forward problem (6) is continuous and satisfies Lipschitz condition with respect to $\boldsymbol{x}$ [see 42]. Thus, given θ, problem (6) has a unique solution $\boldsymbol{x}$.

Besides the numerically simulated states $\boldsymbol{x}$, the inverse problem requires, as an input, directly measured state variables at a discrete set of points $t_1, t_2, \cdots, t_k$. In most circumstances, not all of the state variables of the system can be directly observed. For instance, in epidemiology, data on new confirmed cases of infected individuals is readily available compared to data on other epidemiological statuses of the population. For this reason, we use the new cases reported in periods of time (days, weeks, for example) from the model to formulate the inverse problem.

Let $\boldsymbol{y}_{\text{obs}} = \{y_{1,t}, \cdots, y_{n,t}\}_{t=1}^{M}$, $M \in \mathbb{N}$, denote the observed new infected cases in each unit/period of time $t \in 1, 2, \cdots, M$, for the n patches. Let $Z_i(t, \theta)$ be the incidence in Patch i, during $(t - 1, t]$, derived from the numerical solution of the forward problem (6) with parameter vector θ. The objective function of the inverse problem is thus defined as

$$F(\theta) = \sum_{i=1}^{n} \sum_{t=1}^{M} \frac{\{Z_i(t, \theta) - y_{i,t}\}^2}{Z_i(t, \theta)}. \tag{7}$$

Consequently, the inverse problem becomes: compute $\hat{\theta} \in \Theta$ such that

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min} \, F(\theta) \quad \text{s.t} \quad \Psi(\theta) = \boldsymbol{x}, \tag{8}$$

where Θ is the feasible region of the parameter θ.

Problem (8) can be optimized using methods such as Landweber in [116–119], faster methods such as Levenberg-Marquardt or gradient based numerical non-linear least squares minimization algorithms such as Sequential Least SQuares programming algorithm (SLSQP) in [e.g, 51,120]. Due to the complexity of the forward problem (6), the inverse problem (8) is complex, high dimensional and computationally intensive. As a consequence, problem (8) creates a complex landscape, making it difficult for gradient-based algorithms and other optimization techniques to find the global minima, thereby easily failing to reach the optimal solution. As an alternative, we solve problem (8) using Genetic Algorithm (GA), a type of the diverse machine learning Evolutionary Algorithms (EAs), that uses heuristic search and optimization methods. Besides their efficacy in the optimization of problems involving numerous parameters with large feasible regions (i.e, problems with increased dimensionality) and multiple local optima, GAs do not require gradient computations which may sometimes be computationally challenging or unavailable altogether [121,122].

5.2. Results

Since we are dealing with four zones, we minimize Equation (7) as in (8), for $n = 4$. Besides estimating the epidemiological parameters $\bigcup_{i=1}^{4} \{\beta_i, \kappa_i, \gamma_i\}$, we also estimate the initial conditions $\bigcup_{i=1}^{4} \{E_{i,0}, I_{i,0}\}$. Thus, from the deterministic inversion, we estimate a total of 20 parameters; $\theta = \bigcup_{i=1}^{4} \{\beta_i, \kappa_i, \gamma_i, E_{i,0}, I_{i,0}\}$.

Tables 4 and 5 shows the estimated parameters, while Figures 6 and 7 presents the fitted incidence for the simulated and observed data, respectively. For the simulated incidence, the fit is visibly good, even though there is variation in the estimated parameters from the real ones (see Table 3) and a more important difference on the initial conditions. These disparities can be as a result of the noise that was added to the simulated data, which represent an important modification to the original simulated incidence, especially in the early stages of the outbreak. As in complex models, we can also face the situation where there exists different sets of parameters that would give an equally good fit to the zonal model incidences.

As the estimated parameters from the GA give a good fit for the model incidence, we used them as a guide in specifying the initial sampling points for Stan and t-walk when performing Bayesian inference and uncertainty quantification in the next section. The values were particularly important for t-walk to begin sampling as without them, the algorithm does not converge and completely fails. In addition, as there are no literature that provides the spaces (upper and lower bounds) for the initial conditions of the incubation and prevalence states parameters for Hermosillo, we used the estimates of the said parameters from GA as a guide to set their boundaries in the Bayesian procedure.

Table 4. Model (1) parameter values estimated from deterministic inversion using Genetic Algorithm (GA). The values are used to obtain the fitted simulated model incidence in Figure 6.

Epidemiological		Initial conditions	
Parameters	GA	Parameters	GA
β_1	0.8322	$E_{1,0}$	0.0000
β_2	0.7473	$I_{1,0}$	0.0000
β_3	0.5259	$Y_{1,0} = E_{1,0} + I_{1,0} + R_{1,0}$	0.0000
β_4	1.4354	$E_{2,0}$	200.0000
κ_1	0.0829	$I_{2,0}$	126.4590
κ_2	0.5334	$Y_{2,0} = E_{2,0} + I_{2,0} + R_{2,0}$	326.4590
κ_3	0.1983	$E_{3,0}$	200.0000
κ_4	0.2115	$I_{3,0}$	0.0000
γ_1	0.0357	$Y_{3,0} = E_{3,0} + I_{3,0} + R_{3,0}$	200.0000
γ_2	1.0000	$E_{4,0}$	200.0000
γ_3	0.4248	$I_{4,0}$	8.4447
γ_4	1.0000	$Y_{4,0} = E_{4,0} + I_{4,0} + R_{4,0}$	208.4447

Table 5. Model (1) parameter values estimated from deterministic inversion using Genetic Algorithm (GA). The values are used to obtain the fitted model and observed daily incidence in Figure 7.

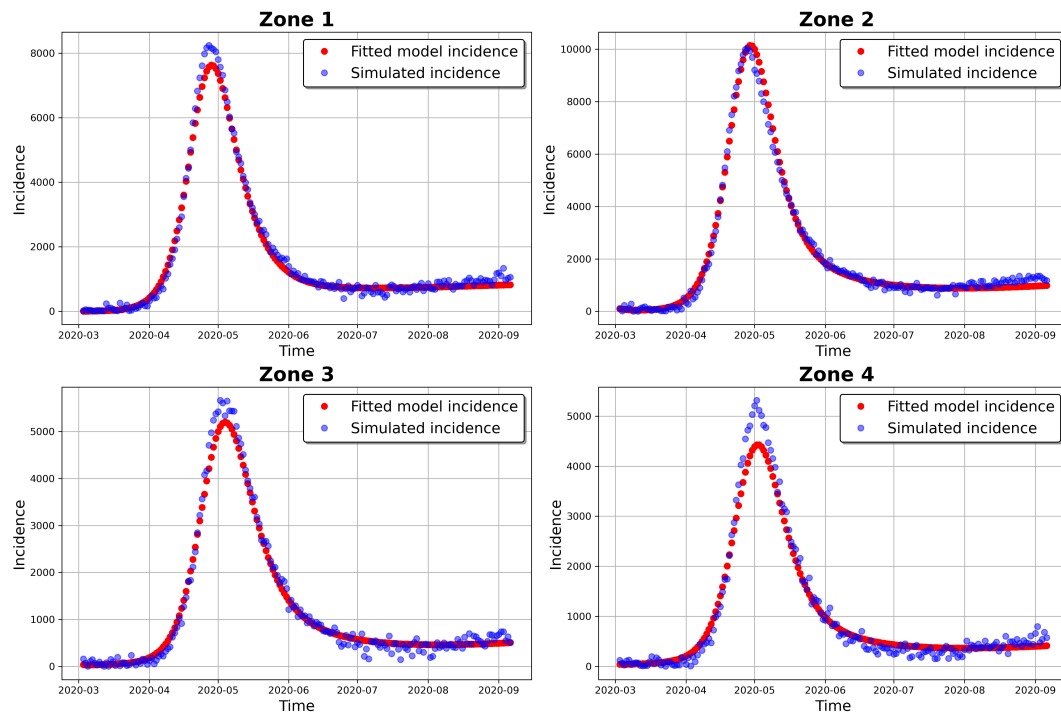| Epidemiological		Initial conditions	
Parameters	GA	Parameters	GA
β_1	1.0950	$E_{1,0}$	58.9925
β_2	1.0473	$I_{1,0}$	0.0000
β_3	0.4904	$Y_{1,0} = E_{1,0} + I_{1,0} + R_{1,0}$	58.9925
β_4	1.4951	$E_{2,0}$	78.1313
κ_1	0.3000	$I_{2,0}$	0.0000
κ_2	0.3000	$Y_{2,0} = E_{2,0} + I_{2,0} + R_{2,0}$	78.1313
κ_3	0.3000	$E_{3,0}$	22.9404
κ_4	0.3000	$I_{3,0}$	0.0000
γ_1	1.0000	$Y_{3,0} = E_{3,0} + I_{3,0} + R_{3,0}$	22.9404
γ_2	1.0000	$E_{4,0}$	4.4192
γ_3	1.0000	$I_{4,0}$	0.0000
γ_4	0.7853	$Y_{4,0} = E_{4,0} + I_{4,0} + R_{4,0}$	4.4192

Figure 6. Fitted model incidence and simulated observed daily incidence.
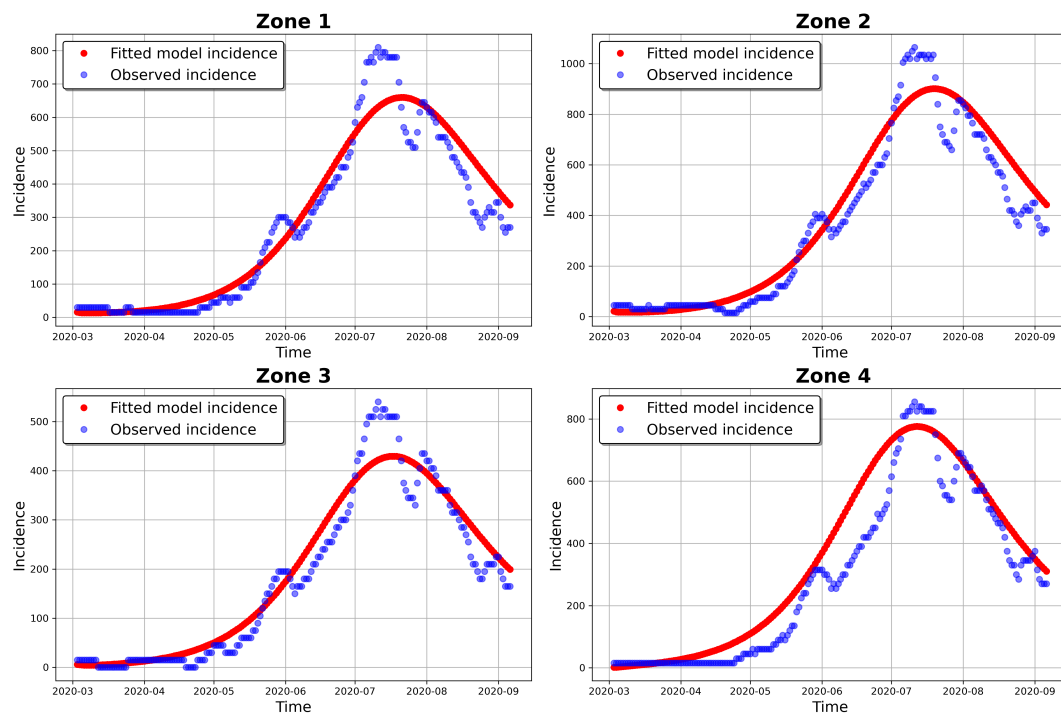

Figure 7. Fitted model incidence and observed daily incidence.

6. Bayesian inference

Deterministic inversion of epidemic models have gained prominence in learning the trajectories of the epidemic using measured epidemiological data. However, deterministic inversion ignores inherent uncertainties due to imperfect data, stochasticity, model structural discrepancies and even parameter uncertainties that usually plagues the epidemic model. In order to make robust, valid and reliable decisions regarding management, prediction and forecasts of current and future epidemics, good treatment and quantification of such uncertainties is imperative. In this section, we use Bayesian

techniques to quantify uncertainties that characterize the epidemic model (1), its parameters and the observed Hermosillo COVID-19 infection data and provide credible intervals for the model output.

Over the last few decades, Bayesian inference has become an important mechanism by providing an optimal probability platform for learning unknown parameters of a system given observed data. A Bayesian framework learns the trajectories of the unknown parameters by updating their prior distributions to the posterior distribution. Such prior updates and exploitation of the parameter space of the posterior distribution can be achieved using the standard Monte Carlo Markov Chain (MCMC) methods. Moreover, MCMC has somewhat led to approximate estimation of intractable posterior distributions, which often characterizes the modelling of real world phenomena. Estimation of such intractable posterior distribution has been a long-standing challenge in the realm of Bayesian inference, [123]. We use t-walk and Stan, which are MCMC based algorithms, to simulate the posterior distribution. Following is a detailed account on how the Bayesian framework is treated in this section.

6.1. Building the Posterior Distribution

In Bayesian realm, the observed data, y_{obs} is fixed while the state variable x and the parameter θ are considered as random variables. In this setting, the Bayes' theorem defines the posterior distribution as

$$\pi(\theta|y_{\text{obs}}) \propto p(y_{\text{obs}}|\theta)\pi(\theta), \tag{9}$$

where $\pi(\theta)$, called the prior distribution, codifies our belief about the unknown parameter θ before the data is observed and $p(y_{\text{obs}}|\theta)$, called the likelihood distribution, codifies all the information available regarding how the observed data y_{obs} was obtained. From the Bayesian perspective, (9) constitutes an inverse problem which may be solved by using the *Maximum a posteriori* (MAP), an argument that maximizes the posterior ($\arg\max_{\theta\in\Theta}\pi(\theta|y_{\text{obs}})$), or the posterior mean ($\mathbb{E}\left[\pi(\theta|y_{\text{obs}})\right]$) or posterior median, as the optimal value $\hat{\theta}$ of θ.

Since we are dealing with count data, which by their very nature are events that occur between a given period of time, a Poisson process would be a natural and meaningful starting point for modeling and performing inference on the observed cases. However, count data, especially epidemiological observations, usually exhibit more variations than is implied by the Poisson distribution, pointing to inherent over-dispersion in the data. Such variations may be due to sampling, aggregation and environmental variability or a combination of both factors making count data to have inherent over-dispersion. It is therefore not possible that count epidemiological data portrays equal mean-variance relationship, as is contemplated by the Poisson distribution. In fact, [124] posits that most commonly used approaches contemplates a quadratic mean-variance relationship. The mean-variance relationship in modeling count data, like epidemiological data, can adequately be described by the negative-binomial distribution as it has an additional parameter that permits the variance to be larger than the mean [125–127]. Moreover, the negative binomial distribution is a mixture of the Poisson and Gamma distribution [127,128], a pointer to the fact that the Poisson distribution is still involved in the modelling even when the negative binomial distribution is used.

We assume that the observed data follows the negative binomial distribution with one dispersion parameter ν and we write $y \sim NB(\nu, p)$ where p is the probability of success. The distribution has the form

$$p(y|\mu,\theta) = \binom{y+\nu-1}{\nu}\left(\frac{\mu}{\mu+\nu}\right)^{\nu}\left(\frac{\nu}{\mu+\nu}\right)^{y}, \tag{10}$$

where

$$\mu = \mathbb{E}(Y) = \frac{\nu p}{1-p} \quad \text{and} \quad \text{Var}(Y) + \frac{\mu^2}{\nu} = \mu + \rho\mu^2 \quad \text{with} \quad \rho = \frac{1}{\nu}.$$

It is noteworthy that $\rho\mu^2 > 0$ is the additional variance allowed by the Negative Binomial distribution with respect to the Poison distribution, and that the parameter ρ can be viewed as controlling the dispersion of the observed data.

We acknowledge that there are other forms of the Negative binomial distribution. For instance, [49] achieved success in modelling COVID-19 and hospital demand in Mexico using a Negative binomial distribution with two over-dispersion parameters. This distribution was proposed by [124]. However, in a model with numerous parameters to estimate as ours, it is our view that additional parameters would further increase the dimensionality of the already complex optimization problem.

We now construct the posterior distribution for the present multi-patched study. Using the already defined observed data and the variable θ in the previous section, we assume that the observed cases of each patch follow a negative-binomial distribution with over-dispersion parameter (the number of failures before first success) ν_i and success probabilities

$$p_i = \frac{\mu_i}{\mu_i + \nu_i},$$

where μ_i is the theoretical mean of the negative binomial distribution which can be taken as patch i incidence

$$\mu_i(j,\theta) = Z_i(j,\theta), \quad i = 1,2,\ldots,n \quad j = 1,2,\cdots,M. \tag{11}$$

With this setting in mind, we have that if $y_{i,j}$, represents the incidence in Patch i observed in day j with theoretical mean $\mu_i(j,\theta)$, as in (11), then we have that

$$y_{i,j}|\theta \sim NB\left(\nu_i, \frac{\mu_i(j,\theta)}{\mu_i(j,\theta)+\nu_i}\right).$$

Then the likelihood from Patch i, $i = 1,2,\cdots,n$, corresponds to

$$p_i(\boldsymbol{y}_i = (y_{i,1},\cdots,y_{i,M})|\theta) = \prod_{j=1}^{M} \frac{\Gamma(y_{i,j}+\nu_i)}{y_{ij}!\Gamma(\nu_i)}\left(\frac{\mu_i(j,\theta)}{\mu_i(j,\theta)+\nu_i}\right)^{\nu_i}\left(\frac{\nu_i}{\mu_i(j,\theta)+\nu_i}\right)^{y_{i,j}}$$

$$\propto \left(\prod_{j=1}^{M}\frac{\Gamma(y_{i,j}+\nu_i)}{\Gamma(\nu_i)}\right)\left(\frac{\mu_i(j,\theta)}{\mu_i(j,\theta)+\nu_i}\right)^{M\nu_i}\left(\frac{\nu_i}{\mu_i(j,\theta)+\nu_i}\right)^{\sum_{j=1}^{M}y_{i,j}}.$$

Hence, the likelihood of the combined data from the n patches is given as $p(\boldsymbol{y}_{\text{obs}}|\theta) = \prod_{i=1}^{n} p_i(\boldsymbol{y}_i|\theta)$.

To establish the prior distribution, we opt to define the joint prior distribution as the product of the marginal. That is, if we denote $\theta = \cup_{i=1,\ldots,n,\ell=1,\ldots,m_i}\theta_{i\ell}$, where $\theta_{i\ell}$ is the ℓ-parameter associated to Patch i, then

$$\pi(\theta) = \prod_{\ell=1}^{m_1} g(\theta_{1l})\prod_{\ell=1}^{m_2} g(\theta_{2l})\cdots\prod_{\ell=1}^{m_n} g(\theta_{nl}) = \prod_{i=1}^{n}\left[\prod_{\ell=1}^{m_i} g(\theta_{il})\right]. \tag{12}$$

where m_i is the number of parameters for Patch i.

The posterior distribution of the parameters from all the n patches can thus be obtained as the product of (9) and (12) as (9).

6.2. Results

We note, from the deterministic inversion section, that we are interested in estimating for each zone i, $i = 1,2,3,4$, the transmission, incubation and recovery rates β_i, κ_i and γ_i as well as the exposed, infected and incidence initial conditions $E_{i,0}$ and $I_{i,0}$. For the Bayesian inference, also called probabilistic inversion, there is an addition of four over-dispersion parameters ν_i, for each zone's observed incidence data. The dispersion parameters can be loosely defined as the number of trials

before we encounter the first successful infection in each zone. These parameters come with the negative binomial distribution, which we use as the distribution of the observed zonal incidences. We thus have, for the Bayesian inference in this section, a total of 24 parameters,

$$\theta = \bigcup_{i=1}^{4} \{\beta_i, \kappa_i, \gamma_i, E_{i,0}, I_{i,0}, \nu_i\}, \tag{13}$$

that we estimate.

An important problem that has been the topic of recent research, is the selection of adequate prior distributions, their corresponding hyperparameters and the delimitations of the parametric spaces of the parameters to be estimated [129–131]. In this research, for the epidemiological parameters, we choose prior distributions with 95% of their body falling within the parametric spaces (upper and lower bounds) provided from COVID-19 literature as in Table 2.

In this way, we use the following prior distributions for each of the parameters;

$$\beta_i \sim \text{LogNorm}(\mu_i, \sigma_i) = g(\beta_i),$$
$$\kappa_i \sim \text{Gamma}(4.5264, 19.1006) = g(\kappa_i),$$
$$\gamma_i \sim \text{Gamma}(1.9826, 3.6943) = g(\gamma_i),$$
$$E_{i,0} \sim \text{Unif}(a_i, b_i) = g(E_{i,0}),$$
$$I_{i,0} \sim \text{Unif}(0, 10) = g(I_{i,0}),$$
$$\nu_i \sim \text{Gamma}(0.7561, 0.2714) = g(\nu_i).$$

where the hyperparameters for the log Normal prior distributions of the contact parameters are $\mu_1 = -0.0529$, $\mu_2 = -0.0975$, $\mu_3 = -0.8562$, $\mu_4 = -0.2583$ and $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2803$, while those for the exposed states' initial conditions are $a_1 = 50$, $b_1 = 100$, $a_2 = 70$, $b_2 = 100$, $a_3 = 15$, $b_3 = 50$, and $a_4 = 0$, $b_4 = 20$. These prior probability distributions, which we have selected using the mentioned criteria, are versatile and allows for the assumption that some parameter values could occur with lower, equal or higher probability densities.

However, to avoid generalizations, it is imperative to perform a diagnostic analysis of the adequacy and correctness of the selected prior distributions before we begin sampling. These prior distributions should be in coherence with our expectations and correspond to the domain knowledge of each parameter as obtained from the deterministic inversion in the previous section. We desire priors that, in accordance with our deterministic inversion domain experience, permits every conceivable configuration of the data while excluding blatantly ludicrous scenarios. Figure 8 shows a graph of the selected prior distributions of the infection, incubation and recovery parameters, vis-à-vis the corresponding posterior distributions. Here we can observe that for most parameters we have allowed for low informative priors and the posterior distributions are within their respective prior probability support. This conforms with the suggestion by [132], that the model should not be overly confined by the priors, which should instead be too wide to encompass a wide variety of situations and data that are incredibly improbable. We thus conclude that our selected prior distributions are adequate and are capable of regularizing our estimates to avoid non-identifiability. We acknowledge that [131] also proposed an interesting criterion for selecting prior distributions together with their corresponding hyperparameters, based on parametric intervals. Applying the assumption of independence of the parameters, the joint prior distribution is $\tilde{\pi}(\theta) = \prod_{i=1}^{4} g(\beta_i)g(\kappa_i)g(\gamma_i)g(E_{i,0})g(I_{i,0})g(\nu_i)$.

The likelihood for the four zonal incidence is

$$p(\boldsymbol{y}_{\text{obs}}|\theta) \propto \prod_{i=1}^{4} \left[\left(\prod_{j=1}^{M_i} \frac{\Gamma(y_{i,j} + \nu_i)}{\Gamma(\nu_i)} \right) \left(\frac{\mu_i(j, \theta)}{\nu_i + \mu_i(j, \theta)} \right)^{M_i \nu_i} \left(\frac{\nu_i}{\nu_i + \mu_i(j, \theta)} \right)^{\sum_{j=1}^{M_i} y_{i,j}} \right],$$

where, M_i is the total number of days for which we observe the incidence for each zone $i = 1, 2, 3, 4$; $y_{i,j}$ is the j-th observed incidence for zone i, and $\mu_i(j, \theta)$ is the incidence of model (1). The posterior distribution is thus given as $p(\theta|\boldsymbol{y}_{\text{obs}}) = p(\boldsymbol{y}_{\text{obs}}|\theta) \tilde{\pi}(\theta)$.
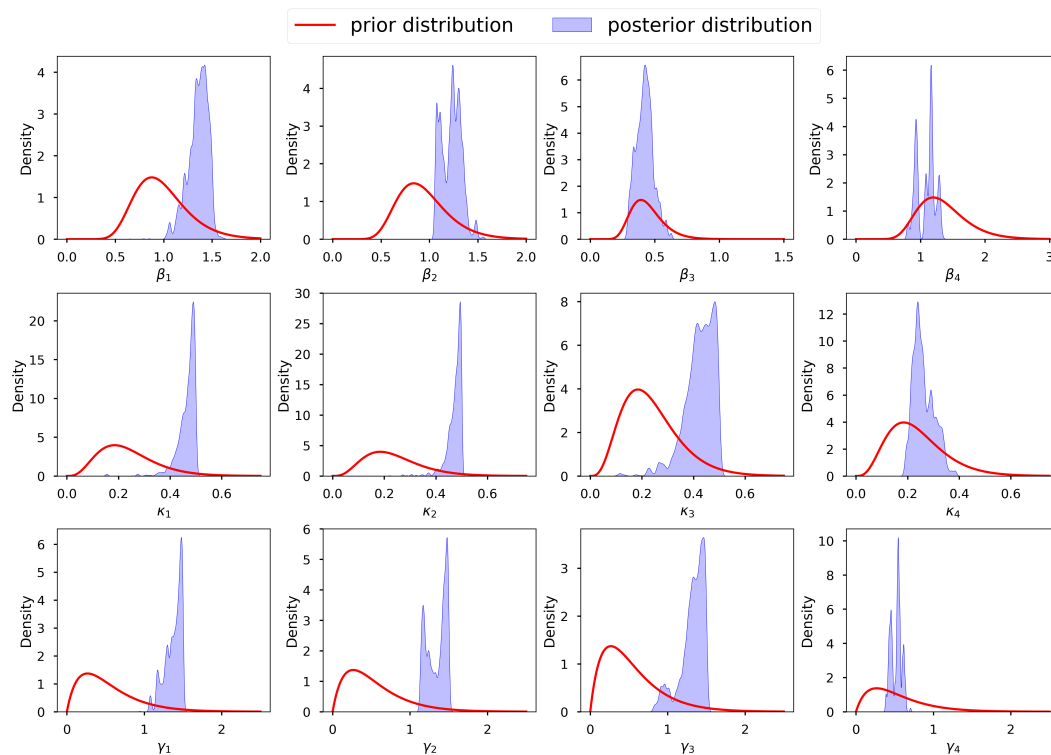

Figure 8. Prior distributions of the infection, incubation and recovery rates parameters and the corresponding posterior distributions obtained from t-walk.

The incidences output from the epidemic model (1) are very sensitive and exhibits a lot of uncertainty to slight changes, especially of the contact, incubation and recovery parameters. As a consequence, it was challenging to provide a wide range of values as support of the posterior distribution in Stan and t-walk. We addressed this challenge by beginning with the deterministic inversion of the epidemic model using GA, as is discussed in the previous section. The output of this inversion are given in Table 5. Consequently, we use the point parameter estimates obtained from the GA as the initial sampling points for t-walk and Stan. Moreover, as there is no literature for the parametric spaces of the incubation and prevalence states initial conditions, we delimit their support in the posterior distribution in the neighbourhood of their point parameter estimates obtained from the GA.

After performing the prior predictive checks, using the t-walk package [133], we obtain posterior samples for each of the 24 parameters (13). MCMC methods, from its definition, produces samples that can have high correlation, harboring some level of redundant information. One way of measuring the level of information redundancy in the posterior sample, is by using the effective sample size (ESS). Theoretically, ESS is the sample size we would obtain if we independently sampled the posterior distribution. The ESS and the proportion of ESS (pESS) of the simulated samples are given in the last two columns of Table A1. Going by how low (high) the values of ESS (pESS) are, we can tell that the simulated chains are highly correlated and thus, this MCMC method requires very high number of iterations. In other words, to obtain samples with insignificant auto correlations can require thinning at large lags, which can remarkably reduces the chains sizes. To obtain chains with reasonable sample sizes for analysis, after running this MCMC method for 1,000,000 iterations, we discarded 360,000 burnin samples and thinned the chains at lag 20.

The summary statistics of the generated samples are given in Table A1 (in the Appendix). As can be observed, the presented summary statistics are; the mean, variance, 95% credible interval (CI), median, ESS and pESS. From this table, we can conclude that in general, all the samples exhibited low variability and that their means are close to the estimates obtained from the GA in the previous section. Further, we can conclude that the means (and medians) are all within the 95% CI. The former and the latter observations can be confirmed from Figure 9, which shows the 95% CI of the samples, the mean and the estimates from the GA. Based on the scales of the values in the intervals, we can conclude that the credible intervals (CIs) are generally short, an indication of good estimates.
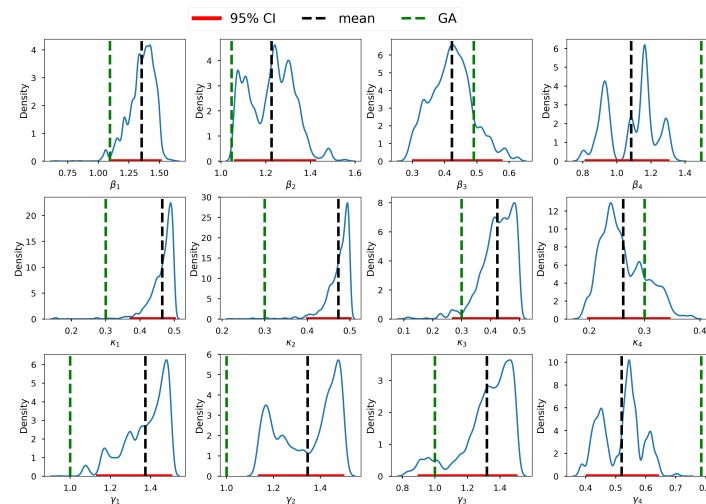
**(a)** Credible intervals, means and GA estimates of the infection, incubation and recovery rates.
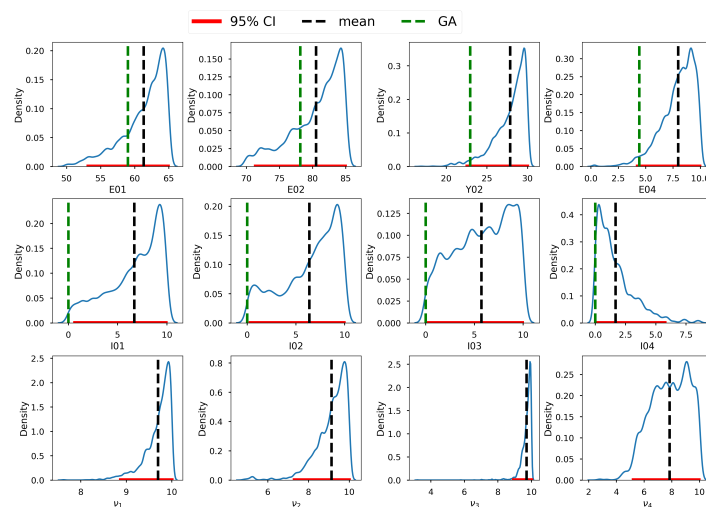
(b) Credible intervals, means and GA estimates of the exposed and prevalence initial conditions and the dispersion parameters.

Figure 9. Credible intervals, means and GA estimates of the incubation and recovery rates, exposed and prevalence initial conditions and dispersion parameters.

Figures 10 and 11 represents the posterior predictive checks. The figures graph the incidence simulations of model (1) from 2020-02-26 to 2020-10-17. This period includes the dates within which the incidence data was observed (2020-02-26 to 2020-09-06) and 41 days prediction dates (2020-09-07 to 2020-10-17). Figure 10 shows 30,000 model incidences (grey curves), from the first 30,000 MCMC iterations, the observed incidence and the *maximum a posteriori* (MAP) model incidence. Here, the MAP model incidence, which is apparently covered by the grey curves, is the model incidence output obtained from the parameter values that maximizes the posterior distribution. We can observe from

this figure, that 1) our fitted model produces simulations that are consistent with the observed incidence data and 2) the simulated trajectories do not portray diverse or variable changes. Indeed, the latter observation is consistent with our previous observation that there is low variability in the simulated samples.
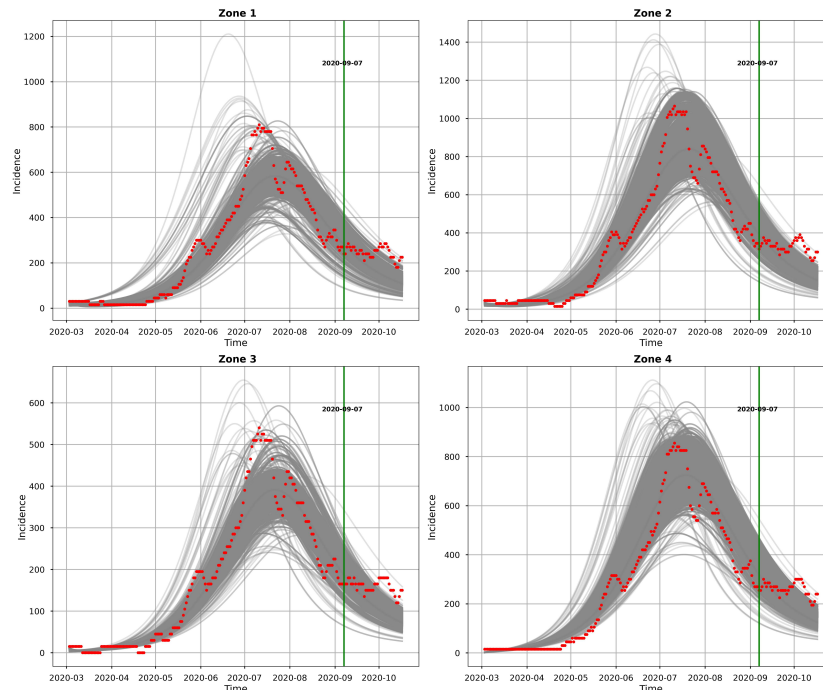
Figure 10. Observed (red points) and model (gray) incidences. The model incidences were obtained by solving the model using parameters from the first 30,000 iterations of MCMC after burnin.
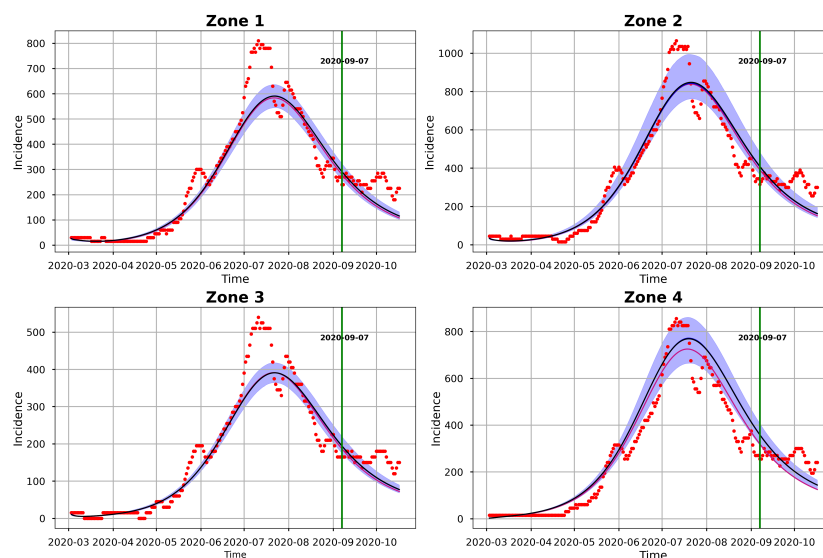
Figure 11. The mean (black), median (medium blue), MAP (medium violet red) and 95% CI (blue bands) of model incidence obtained from t-walk. The red points are observed daily incidences.

Figure 11 shows the 95% posterior confidence intervals (CIs), the mean, median and the MAP of the simulated model incidences. From this figure we observe that most of the observed training incidences for the four zones, fall outside the 95% CI of the simulated model incidence with zone 1 and zone 3 simulated model incidences providing the narrowest 95% CI. In contrast, zones 2 and 4 have larger 95% prediction CI, and they cover more of the observed incidence. Zone 1 model
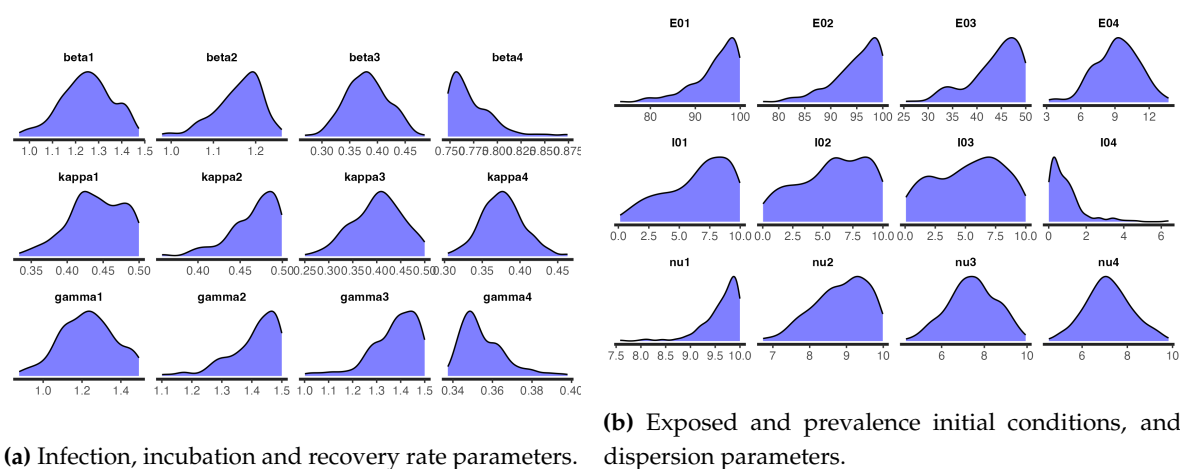
prediction interval provides the worst coverage of the observed prediction incidences in this zone. The proportions of the observed prediction incidence covered by the 95% CI of simulated model incidence from t-walk are presented in Table 8. As the CIs for the four zones are generally not too wide and covers a small proportion of the training observed incidence, and the prediction observed incidences for zones 3 and 4, we can conclude that the point estimates from t-walk seems adequate, but we tend to underestimate the uncertainty associated with the estimated parameters.

We can observe from both Figure 10 and Figure 11 that the model predicts that there will still be some infection levels beyond 2020-09-06, the date when we observed the last incidence. In fact, the figures reveal that in the four zones, some individuals will still be infected up until 2020-10-17, with the median and map incidences being within the 95% CI.

After checking the adequacy of the prior distributions with respect to the posterior distribution and sampling, using the t-walk package, we also performed sampling of the posterior distribution using the Stan package. Stan is a Hamiltonian Monte Carlo (HMC) based package designed for Bayesian statistical analysis. We refer the reader to [134] for more details about the software and [132,135] for an insight on how the package can be used to perform inference and uncertainty quantification of compartmental epidemic models.

For Stan, we used the same support for the prior distribution as used in t-walk, except for the incubation initial conditions. Using the same support for the incubation initial conditions in t-walk led to very low overall proposal acceptance rate. As such, we had to use a different and shorter support range for all the incubation initial conditions prior distributions in t-walk in order to have reasonable proposals acceptance rate for the algorithm. However, in Stan, the support range for the same parameters were larger as Stan is more robust. For the over-dispersion parameters, we used the transformed versions and sampled their inverses in Stan.

For the model at hand, we ran one chain with 1,000 iterations for each of the 24 parameters and took the first 300 iterations as burn-in samples. The number of iterations is not the same as those obtained from t-walk as Stan requires longer periods of time, than t-walk, to complete each iteration. However, from the analysis, we observe that the chains are importantly less correlated, and we can obtain bigger effective sample sizes. Figures 12 and 13 respectively shows the trace plots and the posterior densities of the samples of the estimated parameters. We obtained unit Rhat values from the Stan summary statistics output, an indication that all, except two, of the transitions converged. Additional Stan statistics, particularly, the diagnosis of the pairs plot of a subset of the estimated parameters, (Figure 14), confirms convergence of all, except two of the transitions of the chain during sampling.

(a) Infection, incubation and recovery rate parameters.

(b) Exposed and prevalence initial conditions, and dispersion parameters.

Figure 12. Posterior distributions of the infection, incubation and recovery rates, exposed and prevalence initial conditions and dispersion parameters as obtained from Stan.

(a) Infection, incubation and recovery rate parameters.

(b) Exposed and prevalence initial conditions and the dispersion parameters.
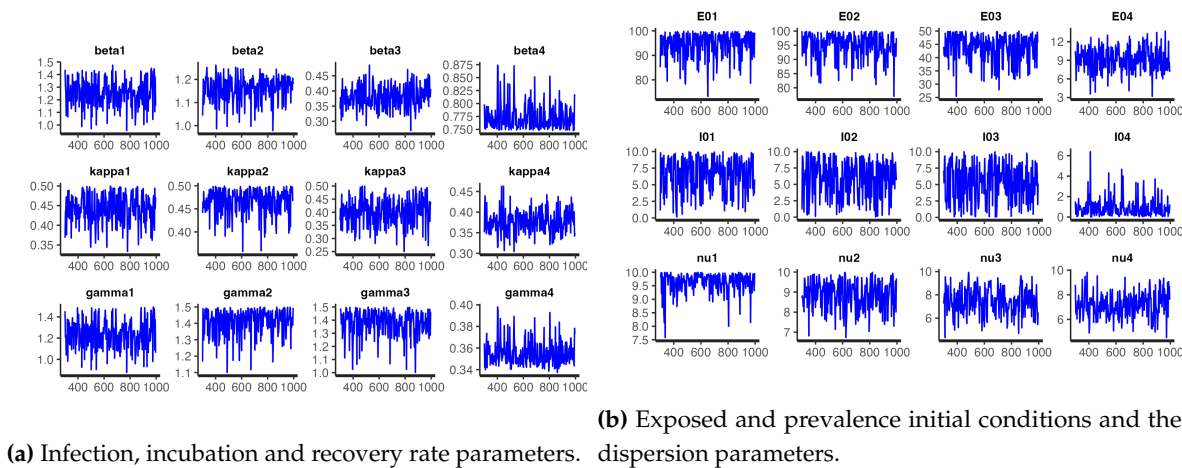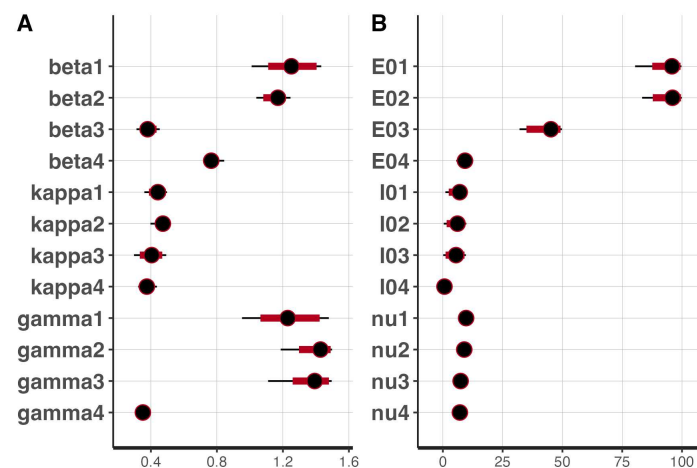
Figure 13. Traceplots of the samples of infection, incubation and recovery rates, exposed and prevalence initial conditions and over-dispersion parameters as obtained from Stan.
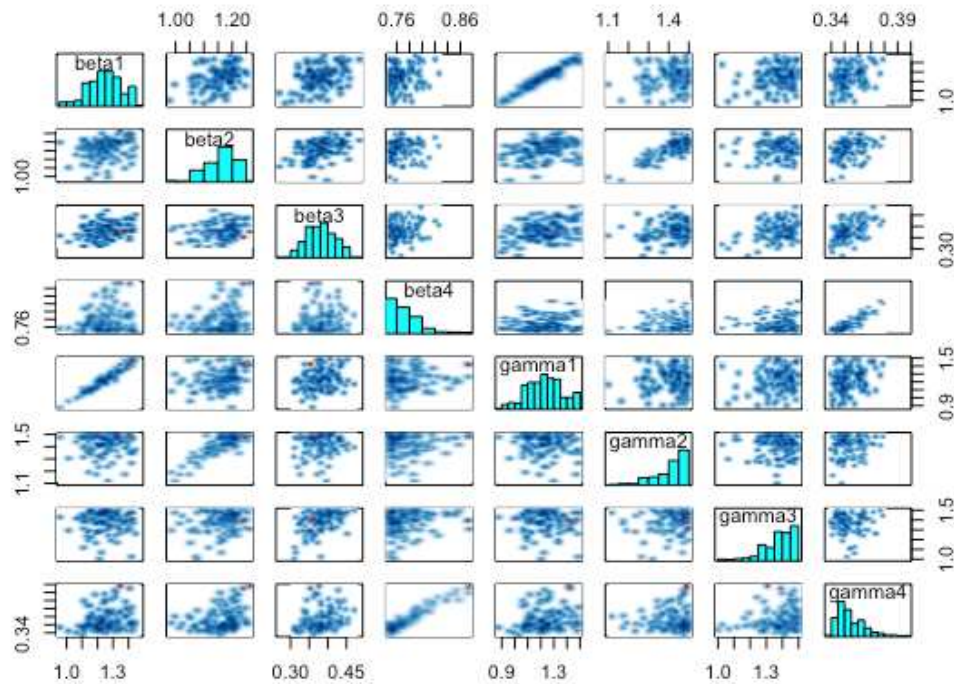
Figure 14. Pairs plot for the samples of infection and recovery rates as obtained from Stan.

Figure 15 shows the 95% CIs of the posterior distributions obtained from Stan. All the parameters, except β_1, γ_1, γ_2, and γ_3, had short CIs. It is noteworthy that just as in t-walk, the samples of the inverse of the over-dispersion parameters, ν_i from Stan had very narrow intervals.

Figure 15. 95% Credible intervals of the infection, incubation and recovery rates, the initial conditions and the over-dispersion parameters as obtained from Stan.

The pairs plot in Figure 14 shows the posterior distributions of the infection and recovery parameters (diagonal figures) and scatter plots (off-diagonal figures) showing the relationship between each of the samples. Since there were many (24) parameters to estimate, it was not appealing to have a joint pairs plot for all of them. In as much as this is the ideal procedure, doing this would lead to a figure with very tiny non-visualizable figure entries. We thus chose to pair the infection and recover rates (see Figure 14), as the scatter plots of these parameters exhibited interesting relationships. For instance, we observe from this figure, that there was a strong linear positive posterior correlation between β_1 and γ_1, β_2 and γ_2 and β_4 and γ_4. In contrast, the infection parameters for the four zones did not exhibit any important relationship among themselves. This was also the case with the recovery parameters.

The summary statistics of the samples as obtained from Stan are shown in Table A2. These statistics are slightly different from the summary statistics from t-walk as displayed in Table A1. The differences in the statistics of the prevalence initial conditions and the over-dispersion parameters are obvious, as we have already mentioned that besides using different support ranges for the distributions of prevalence initial conditions, we transformed and sampled the inverses of the over-dispersion parameters ν_i, $i = 1, 2, 3, 4$. Otherwise, the slight differences in the statistics of the other parameters could have been occasioned by the difference in how the t-walk and Stan packages performs sampling. Notably, these differences would not lead to a different conclusion regarding the potency of COVID-19 infections in Hermosillo in 2020. The statistics in Table A2 reveals that the point estimates of the parameters are within the 95% CIs as is the case with the estimated parameters from t-walk, as can be seen in Table A1.

The differences in the summary statistics of the samples obtained from both packages is further reflected in Figures 11 and 16, which are the posterior predictive check figures from t-walk and Stan respectively. We can observe that the mean, median and the observed incidences for each zone are all covered by the respective 95% CIs. In comparison, we observe from the t-walk output in Figure 11, that whereas the 95% CI of the epidemic model incidences did not cover some observed daily incidences, the same interval from Stan covered all the training observed zonal incidences. We thus conclude that the estimated parameters from Stan can capture a more realistic uncertainty level, compared to those estimated by t-walk. Both Figures 11 and 16 reveals that COVID-19 infections in the four zones, and consequently, the whole of Hermosillo, was at its peak in mid-July 2020, with a prediction of continued infections beyond 2020-09-06, which was the last date of observed training daily incidences.
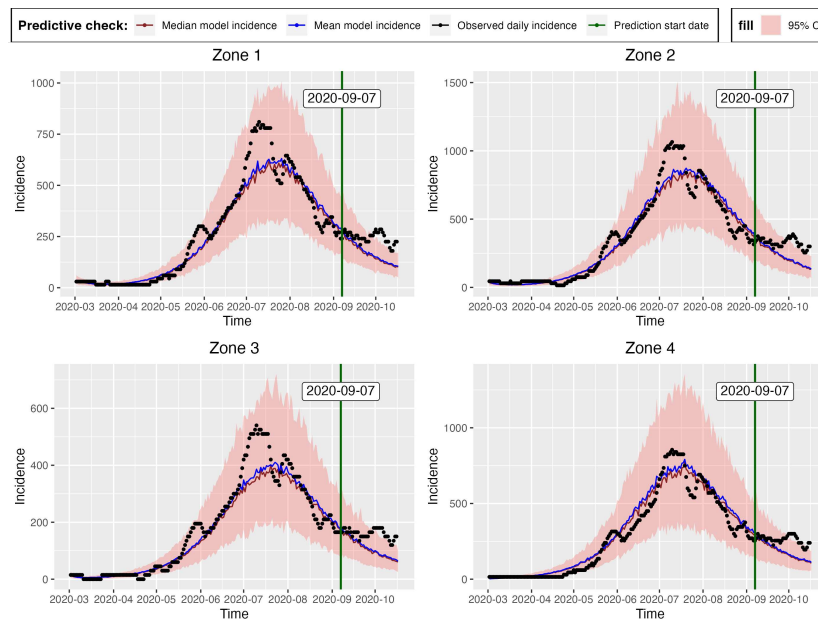
Figure 16. Observed daily incidence, Posterior incidence predictions and 95% uncertainty bands obtained from Stan.

Table 2 shows the ranges of the parameters of the model as obtained from various COVID-19 literature. Our results, both from the GA, t-walk and Stan packages (see Tables 5, A1 and A2), indicate that the point estimates of the infection and incubation parameters; β_i and κ_i, $i = 1, 2, 3, 4$, that best fit the incidences for the four zones, were within the ranges from literature. This is however not the case for the recovery parameters, γ_i, whose estimates were obtained as generally averaging close to one day recovery period for the four zones. This result is not surprising, as our incidence data for the four zones were inflated by a large under-reporting factor. This inflation, besides accounting for a possible lack of systematic testing of COVID-19 in Hermosillo, also accounted for the existence of subclinical and asymptomatic COVID-19 patients in the four zones under study. Going by how well the model fits the data, (see Figures 7, 10, 11 and 16), we believe that accounting for incidence under-reporting by a factor of 15, as reported by [86] is realistic.

We further evaluate the prediction capability of the multi-patch model by checking the proportion of the observed zonal prediction incidences that are within the 95% CIs of the zone model incidences. Table 8 shows these proportions obtained from Stan (see Figure 16) and t-walk (see Figure 11). From this table, we can see that Stan produces 95% prediction bands that cover more than 50% of the observed prediction incidences for all the zones. On the other hand, t-walk provides 95% CIs large enough to cover less than 50% of observed prediction incidences for all the four zones, with zone 1 providing the narrowest simulated model incidence CIs that covers only 22.22% of the observed prediction incidence. In effect, we conclude that Stan outperforms t-walk as regards to the percentages of covered observed daily prediction incidence for all the zones.

6.3. Comparison to a single Patch model

To assess the improvement of using a more complex model against the usual single population model, in this section we fit the single-patch model (global SEIRS model) and compare its predictive performance against the multi-patch model with mobility. The systems of differential equations forming the single-patch model are presented in equation (14).

$$\begin{cases} \dot{S} = \Lambda - \beta SI/N - \mu S + \tau R \\ \dot{E} = \beta SI/N - (\kappa + \mu)E \\ \dot{I} = \kappa E - (\gamma + \phi + \mu)I \\ \dot{R} = \gamma I - (\mu + \tau)R. \end{cases} \tag{14}$$

We run the single-patch model using similar values of birth, natural death and disease induced mortality rates used in the multi-patched model. We estimated the infection (β), incubation(κ) and recovery(γ) rates and the initial conditions E_0 and I_0 for the single patch model. The prior distributions of these parameters and the likelihood of the global observed incidence remains the same as the ones used in the multi-patched model. The initial population was taken as the sum of the zones population sizes, shown in Figure 1. That is, $N_0 = \sum_{i=1}^{4} N_i(0) = 858,812$.

We use the observed global incidence in the period 2020-02-26 to 2020-09-06 to train the global single-patch model and then use the incidence in the period 2020-10-17 to assess the prediction efficiency of the model. Table 6 shows the summary statistics of the estimated parameters. For comparison purposes with the multi-patch model, we sum all the zones' predicted mean and median incidences and compute the efficiency measures. The efficiency measures used in this case are the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(Y_t - \hat{Y}_t\right)^2} \quad \text{and} \quad \text{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|,$$

where n is the total number of the predicted incidences, Y_t is the actual observed incidence and $\hat{Y}_t$ is the predicted incidence, which in this case is the mean and median incidences for both the multi-patch and global single-patched models.

Table 6. Statistics of the single patch model parameters estimated using probabilistic (Stan) inversion.

Parameters	Probabilistic inversion (Stan) statistics								
	Mean	SD	$q_{0.025}$	$q_{0.250}$	$q_{0.5}$	$q_{0.750}$	$q_{0.975}$	ESS	pESS
β	1.0139	0.0972	0.8168	0.9503	1.0191	1.0843	1.1701	244	0.0122
κ	0.8690	0.0847	0.6911	0.8140	0.8798	0.9436	0.9917	301	0.0151
γ	0.8569	0.0862	0.6797	0.8015	0.8640	0.9191	0.9936	244	0.0122
E_0	1.2060	0.8259	0.0497	0.5090	1.0952	1.7977	2.8544	296	0.0148
I_0	1.6101	0.7787	0.1173	1.0340	1.6587	2.2338	2.9343	237	0.0119
ν	4.0792	0.5313	3.0715	3.7108	4.0322	4.4569	5.1152	264	0.0132

Figure 17 shows the global single-patch predicted incidences together with the 95% credible incidence intervals, and Table 7 shows the efficiency measures of the predictions both from the multi and single-patched models. From this table, and indeed, comparing Figures 16 and 17, we see that the multi-patched model outperforms the global single patch model, as it gives predicted incidences with lower RMSE and MAPE values. Moreover, Figure 16 reveals that a bigger percentage of the observed prediction and model predicted incidences from the multi-patched model are all within the 95% credible incidence interval. We thus conclude that in modelling COVID-19 in Hermosillo, using the multi-patched model with mobility represents an important improvement compared to the single-patch model.
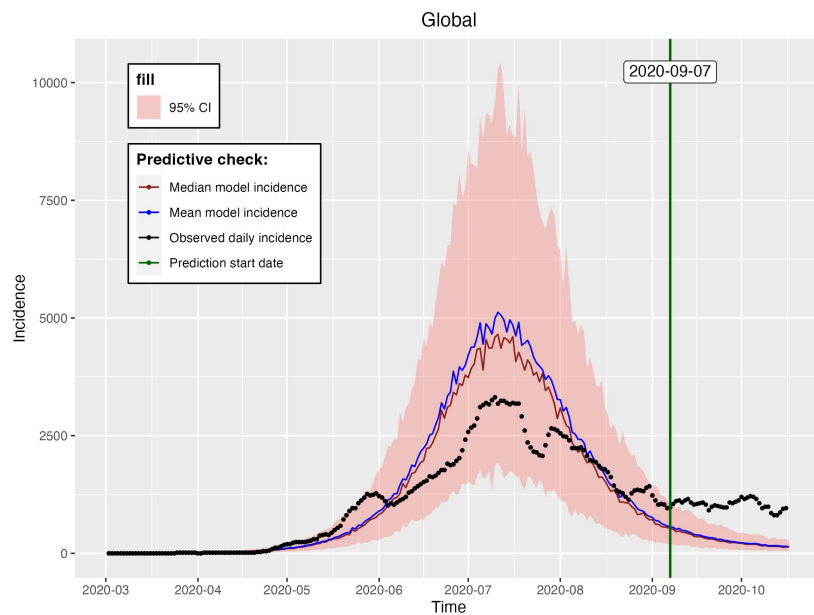
Figure 17. Global single patched observed and predicted mean and median incidences together with the 95% CI.

Table 7. Efficiency measures for the predicted mean and median incidences from the multi-patch and single-patch epidemic models.

Efficiency measure	Multi-patch		Single-patch	
	Mean	Median	Mean	Median
RMSE	395.61	418.61	773.98	796.06
MAPE	34.40	37.06	73.86	76.10

Table 8. Proportion of the observed zones prediction incidence covered by the CIs of the model incidence both from t-walk and Stan.

Zone	% of observed prediction incidence covered by CIs	
	Stan	t-walk
Zone 1	55.56	22.22
Zone 2	55.56	44.44
Zone 3	58.33	30.56
Zone 4	63.89	38.89

7. Conclusions

In this work, we have utilized three techniques to estimate the parameters and perform inference and uncertainty quantification on a multi-patched epidemic model with mobility, residency and demography. The analysis has been conducted using mobile phone GPS data and confirmed COVID-19 cases in the four zones of Hermosillo, Mexico. The model incorporates two sets of parameters; 20 mobility and residence times parameters and 20 epidemiological parameters (including the incubation and prevalence initial conditions) which have all been estimated.

The complexity of the epidemic model makes it susceptible to the problem of parameter non-identifiability, a complication we have addressed by using the two distinctive data sets and different estimation techniques to estimate the two sets of parameters separately. In the first step, we

have used Brownian Bridge model and mobile phone GPS data to estimate the first set of 20 mobility and residence times for the four zones in Hermosillo. The next set of the model epidemiological parameters have been estimated using confirmed positive COVID-19 cases and two estimation techniques; deterministic inversion using GA and probabilistic inversion, using two different MCMC methods implemented in the t-walk and Stan statistical packages.

The incidence outputs of the epidemic model are very sensitive to slight changes of its parameters, especially the infection, incubation and recovery rates. This necessitated the use of the deterministic inversion technique as the first step of estimation, as it provided us with an idea of the possible parametric values, and model incidences that best fit the observed incidences. This first step of estimation thus presented us with the scope of the parametric space within which to delimit the support of the posterior distribution and the initial sampling values used in the probabilistic inversion stage, using t-walk and Stan.

In this article, we are not concerned with a detailed comparison of t-walk and Stan, as there could be many factors that affect their performance, that undoubtedly vary among different models. Some of these factors may include the chosen parametrization of the model, the prior distribution and initial values from where the sampling starts. We however mention that some literature have established that it may be difficult to use MCMC-based inference tools to model systems of ODEs, since the relevant likelihood function may contain several local minima that causes them to fail the standard regulatory conditions. In these cases, Stan offers high statistical efficiency and, in complex scenarios, may produce more accurate results compared to t-walk. For the presented epidemic model, Stan performs better than t-walk as the posterior distributions are characterized by highly correlated parameter spaces. We have observed this fact in the present study based on the large ESS proportions obtained from Stan compared to those obtained from t-walk.

The results from the estimation in this study shows that the infection and incubation rates for the four zones in Hermosillo were within the ranges postulated in various COVID-19 literature. However, the recovery parameters fell outside the estimated ranges reported in other COVID-19 studies. This deviation has however been justified as possibly emanating from the inflation of the incidences in the four zones by a factor of 15 to account for under-reporting. . Further, from the estimated parameters, the model presents a good fit to COVID-19 infections in the four zones, and consequently, the whole city of Hermosillo, and under the model it also has its peak in mid-July 2020.

The epidemic model and the results in this research can be useful to public health officials as it not only provides better predictions, but it also explicitly introduces mobility parameters that can be modified to assess the effectiveness of different levels of mobility restrictions. These restrictions can be global or applied to only some patches. We believe the model and methods then can turn into a user-friendly tool to guide public health policies for management and mitigation of emerging and re-emerging infectious diseases.

The epidemic model in this article is confined within the limits of modelling a single strain of the COVID-19 virus. However, like many infectious diseases, multiple strains of COVID-19, resulting from its gene mutation, have emerged. For instance, four mutants of the pathogen were detected within the first one and a half years of the pandemic. It would therefore be relevant to extend the present epidemic model to include the dynamics of several COVID-19 pathogenic strains. This extension would lead to better understanding and management of future infectious outbreaks.

In addition, besides urban mobility, the epidemic model used in this research did not account for government intervention measures , such as vaccination and regional or national mobility. As the vaccination campaign in Mexico started in December 2020, this factor is absent, but it would be desirable to introduce these, and other mitigation strategies in future models, for modelling subsequent outbreaks of COVID-19 and other infectious diseases.

It is worth noting that the assumptions for the proposed model include those for the ODEs to describe the dynamics within each patch. While ODEs can be effective in approximating dynamics in

many scenarios, there are cases where, even with homogeneous mixing, the mathematical system may fail to accurately represent the behavior of a small population.

One reason for this is the assumption of continuous variables and smooth transitions inherent in ODEs. In small populations, discrete effects and stochastic events can have a significant impact on the dynamics. These discrete effects, such as random fluctuations or individual interactions, introduce variability that is not captured by ODEs, leading to the model inaccurately reflecting the true behavior of the system.

To address this, one option for small spatial scales and patch sizes, is to leverage computational techniques such as agent-based modeling. This approach captures the important variability that arises from individual interactions [136–141]. Additionally, to reduce the variability of simulations, relevant contact patterns produced by street or building layouts can be incorporated. At a more detailed level, modeling individual interactions as a network of contacts still allows for fitting or performing statistical inference using computer-intensive methods [55,142–144]. Because of this, scaling up these methods on slightly larger patches can pose serious challenges that requiere careful considerations.

Author Contributions: Conceptualization, A.O.A, L.L.R-R.; methodology,A.O.A, L.L.R-R.; software, A.O.A, L.L.R-R. and J.F.E; validation, A.O.A, L.L.R-R. and J.F.E.; formal analysis, A.O.A, L.L.R-R.; investigation, A.O.A, L.L.R-R.; resources, A.O.A, L.L.R-R. and J.F.E.; data curation, A.O.A, L.L.R-R. and J.F.E.; writing—original draft preparation, A.O.A, L.L.R-R.; writing—review and editing, A.O.A, L.L.R-R.; visualization, A.O.A, L.L.R-R.; supervision, L.L.R-R.; project administration, L.L.R-R.. All authors have read and agreed to the published version of the manuscript.

Appendix A. Posterior statistics from t-walk and Stan

Appendix A.1. Posterior statistics from t-walk

Table A1. Statistics of model parameters estimated using probabilistic (twalk) inversion

Parameters	Probabilistic inversion (t-walk) statistics								
	Mean	SD	$q_{0.025}$	$q_{0.250}$	$q_{0.5}$	$q_{0.750}$	$q_{0.975}$	ESS	pESS
β_1	1.3552	0.1039	1.1166	1.2951	1.3724	1.4307	1.5033	25	0.0008
β_2	1.2258	0.1014	1.0668	1.1321	1.2366	1.3036	1.4168	22	0.0007
β_3	0.4223	0.0663	0.3046	0.3734	0.4225	0.4629	0.5734	28	0.0009
β_4	1.0839	0.1404	0.8218	0.9383	1.1300	1.1700	1.2970	18	0.0006
κ_1	0.4649	0.0368	0.3754	0.4510	0.4757	0.4888	0.4989	233	0.0073
κ_2	0.4730	0.0294	0.4013	0.4627	0.4811	0.4925	0.4994	112	0.0035
κ_3	0.4236	0.0575	0.2735	0.3940	0.4320	0.4683	0.4964	122	0.0038
κ_4	0.2614	0.0401	0.1991	0.2318	0.2526	0.2897	0.3436	21	0.0006
γ_1	1.3740	0.1082	1.1382	1.3000	1.4047	1.4649	1.4977	22	0.0007
γ_2	1.3461	0.1243	1.1416	1.2259	1.3920	1.4620	1.4945	20	0.0006
γ_3	1.3196	0.1551	0.9050	1.2604	1.3547	1.4367	1.4959	32	0.0010
γ_4	0.5198	0.0673	0.4061	0.4546	0.5358	0.5669	0.6372	18	0.0006
$E_{1,0}$	61.3153	3.1875	53.1884	59.6405	62.2854	63.8354	64.8765	92	0.0029
$I_{1,0}$	6.6851	2.6790	0.6896	4.9450	7.3324	8.9846	9.8718	175	0.0055
$E_{2,0}$	80.5102	3.8104	71.4076	78.1892	81.6369	83.5947	84.8731	88	0.0027
$I_{2,0}$	6.3754	2.8778	0.3521	4.4653	7.1644	8.8401	9.9016	75	0.0023
$E_{3,0}$	27.8516	1.9766	22.5930	26.8572	28.4460	29.3748	29.9438	160	0.0050
$I_{3,0}$	5.7194	2.8518	0.3452	3.4194	6.0118	8.2424	9.8920	288	0.0090
$E_{4,0}$	7.9665	1.5100	4.2772	7.1346	8.2710	9.1433	9.9091	87	0.0027
$I_{4,0}$	1.6821	1.5480	0.0551	0.4926	1.2054	2.3330	5.7315	70	0.0022
ν_1	9.6957	0.2948	8.8788	9.5804	9.7863	9.9084	9.9933	226	0.0071
ν_2	9.1320	0.7812	7.3177	8.7525	9.3166	9.7064	9.9798	141	0.0044
ν_3	9.7135	0.3543	8.9596	9.6075	9.8080	9.9251	9.9939	147	0.0046
ν_4	7.8458	1.3651	5.2517	6.8248	7.9591	9.0183	9.9349	35	0.0010

Appendix A.2. Posterior statistics from Stan

Table A2. Statistics of model parameters estimated using probabilistic (Stan) inversion

Parameters	Probabilistic inversion (Stan) statistics								
	Mean	SD	$q_{0.025}$	$q_{0.250}$	$q_{0.5}$	$q_{0.750}$	$q_{0.975}$	ESS	pESS
β_1	1.2476	0.1080	1.0107	1.1807	1.2498	1.3193	1.4317	255	0.3643
β_2	1.1591	0.0546	1.0432	1.1294	1.1698	1.2000	1.2460	213	0.3043
β_3	0.3811	0.0380	0.3129	0.3543	0.3805	0.4057	0.4536	244	0.3486
β_4	0.7728	0.0245	0.7489	0.7548	0.7656	0.7855	0.8440	173	0.2471
κ_1	0.4409	0.0377	0.3606	0.4166	0.4427	0.4729	0.4976	188	0.2686
κ_2	0.4660	0.0282	0.3973	0.4501	0.4728	0.4881	0.4985	248	0.3543
κ_3	0.4019	0.0517	0.2978	0.3641	0.4053	0.4385	0.4930	244	0.3486
κ_4	0.3766	0.0280	0.3228	0.3578	0.3758	0.3933	0.4365	196	0.2800
γ_1	1.2275	0.1359	0.9530	1.1274	1.2277	1.3274	1.4783	241	0.3443
γ_2	1.4091	0.0795	1.2259	1.3692	1.4280	1.4729	1.4980	209	0.2986
γ_3	1.3743	0.0983	1.1104	1.3161	1.3922	1.4521	1.4958	204	0.2914
γ_4	0.3550	0.0113	0.3408	0.3472	0.3514	0.3612	0.3846	188	0.2686
$E_{1,0}$	94.4414	5.1047	80.4055	92.2135	95.8150	98.3510	99.7569	234	0.3343
$I_{1,0}$	6.5046	2.5618	1.0641	4.6339	7.0732	8.6031	9.8693	202	0.2886
$E_{2,0}$	94.8808	4.5037	83.3513	92.6637	95.9344	98.4198	99.8949	143	0.2043
$I_{2,0}$	5.8008	2.8349	0.3588	3.6777	6.1113	8.2312	9.8222	217	0.3100
$E_{3,0}$	43.6993	5.1480	32.0696	40.9228	45.1198	47.7972	49.8012	224	0.3200
$I_{3,0}$	5.2505	2.8159	0.1604	2.5964	5.4800	7.5498	9.6372	217	0.3100
$E_{4,0}$	9.1961	1.9174	5.6108	7.9395	9.2555	10.4385	12.6814	262	0.3743
$I_{4,0}$	0.8923	0.9075	0.0397	0.2712	0.6660	1.1865	3.4709	202	0.2886
ν_1	9.6200	0.3765	8.6530	9.4789	9.7231	9.8887	9.9887	167	0.2386
ν_2	8.8543	0.7276	7.3756	8.3520	8.9653	9.4043	9.8811	222	0.3171
ν_3	7.4251	1.1378	5.1812	6.6442	7.4181	8.2871	9.4924	146	0.2086
ν_4	7.1626	1.0035	5.2185	6.4972	7.1173	7.7591	9.2361	151	0.2158

References

1. Jašková, D.; Havierniková, K. The human resources as an important factor of regional development. *International Journal of Business and Society* **2020**, *21*, 1464–1478.

2. Krypa, N. Social economic development and the human resources management. *Academic Journal of Interdisciplinary Studies* **2017**, *6*, 73.

3. Peterson, E.W.F. The role of population in economic growth. *Sage Open* **2017**, *7*, 2158244017736094.

4. Ramírez-Ramírez, L.L.; Montoya, J.A.; Espinoza, J.F.; Mehta, C.; Akuno, A.O.; Bui-Thanh, T. Use of mobile phone sensing data to estimate residence and mobility times in urban patches during the COVID-19 epidemic: The case of the 2020 outbreak in Hermosillo, Mexico. *Preprint in Research Square* **2022**. Available at https://doi.org/10.21203/rs.3.rs-2219155/v1.

5. VLAD, C.A.; Ungureanu, G.; Militaru, M. Human resources contribution to economic growth. *REVISTA ECONOMICĂ* **2012**, p. 850.

6. Oaks Jr, S.C.; Shope, R.E.; Lederberg, J.; et al. *Emerging infections: microbial threats to health in the United States*; National Academies Press, 1992.

7. Monto, A.S. The Global Threat of New and Reemerging Infectious Diseases: Reconciling US National Security and Public Health Policy, 2003.

8. Adiga, A.; Dubhashi, D.; Lewis, B.; Marathe, M.; Venkatramanan, S.; Vullikanti, A. Mathematical models for covid-19 pandemic: A comparative analysis. *Journal of the Indian Institute of Science* **2020**, *100*, 793–807.

9. Schmidtchen, M.; Tse, O.; Wackerle, S. A multiscale approach for spatially inhomogeneous disease dynamics. *arXiv preprint arXiv:1602.05927* **2016**.

10. Castillo-Chavez, C.; Bichara, D.; Morin, B.R. Perspectives on the role of mobility, behavior, and time scales in the spread of diseases. *Proceedings of the National Academy of Sciences* **2016**, *113*, 14582,14588, [https://www.pnas.org/content/113/51/14582.full.pdf]. https://doi.org/10.1073/pnas.1604994113.

11. Rvachev, L.A.; Longini Jr, I.M. A mathematical model for the global spread of influenza. *Mathematical biosciences* **1985**, *75*, 3–22.

12. Baroyan, O.; Rvachev, L.; Basilevsky, U.; Ermakov, V.; Frank, K.; Rvachev, M.; Shashkov, V. Computer modelling of influenza epidemics for the whole country (USSR). *Advances in Applied Probability* **1971**, *3*, 224–226.

13. Anifandis, G.; Tempest, H.G.; Oliva, R.; Swanson, G.M.; Simopoulou, M.; Easley, C.A.; Primig, M.; Messini, C.I.; Turek, P.J.; Sutovsky, P.; et al. COVID-19 and human reproduction: A pandemic that packs a serious punch. *Systems Biology in Reproductive Medicine* **2021**, *67*, 3–23.

14. Prieto, K. Current forecast of COVID-19 in Mexico: A Bayesian and machine learning approaches. *PloS one* **2022**, *17*, e0259958.

15. Weston, D.; Hauck, K.; Amlôt, R. Infection prevention behaviour and infectious disease modelling: a review of the literature and recommendations for the future. *BMC public health* **2018**, *18*, 1–16.

16. Luo, M.; Qin, S.; Tan, B.; Cai, M.; Yue, Y.; Xiong, Q. Population Mobility and the Transmission Risk of the COVID-19 in Wuhan, China. *ISPRS International Journal of Geo-Information* **2021**, *10*, 395.

17. Changruenngam, S.; Bicout, D.J.; Modchang, C. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports* **2020**, *10*, 1–13.

18. Findlater, A.; Bogoch, I.I. Human mobility and the global spread of infectious diseases: a focus on air travel. *Trends in parasitology* **2018**, *34*, 772–783.

19. Danon, L.; Ford, A.P.; House, T.; Jewell, C.P.; Keeling, M.J.; Roberts, G.O.; Ross, J.V.; Vernon, M.C. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases* **2011**, *2011*.

20. Apostolopoulos, Y.; Sönmez, S.F. *Population mobility and infectious disease*; Springer, 2007.

21. Gushulak, B.D.; MacPherson, D.W. Population mobility and infectious diseases: the diminishing impact of classical infectious diseases and new approaches for the 21st century. *Clinical infectious diseases* **2000**, *31*, 776–780.

22. Dalziel, B.D.; Pourbohloul, B.; Ellner, S.P. Human mobility patterns predict divergent epidemic dynamics among cities. *Proceedings of the Royal Society B: Biological Sciences* **2013**, *280*, 20130763.

23. Ni, S.; Weng, W. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. *Physical review E* **2009**, *79*, 016111.

24. Tang, M.; Liu, Z.; Li, B. Epidemic spreading by objective traveling. *EPL (Europhysics Letters)* **2009**, *87*, 18005.

25. Yang, Z.; Cui, A.X.; Zhou, T. Impact of heterogeneous human activities on epidemic spreading. *Physica A: Statistical Mechanics and its Applications* **2011**, *390*, 4543–4548.

26. Meloni, S.; Perra, N.; Arenas, A.; Gómez, S.; Moreno, Y.; Vespignani, A. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports* **2011**, *1*, 1–7.

27. Wang, B.; Cao, L.; Suzuki, H.; Aihara, K. Safety-information-driven human mobility patterns with metapopulation epidemic dynamics. *Scientific reports* **2012**, *2*, 1–8.

28. Epstein, J.M.; Goedecke, D.M.; Yu, F.; Morris, R.J.; Wagener, D.K.; Bobashev, G.V. Controlling pandemic flu: the value of international air travel restrictions. *PloS one* **2007**, *2*, e401.

29. Colizza, V.; Barrat, A.; Barthelemy, M.; Valleron, A.J.; Vespignani, A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine* **2007**, *4*, e13.

30. Bajardi, P.; Poletto, C.; Ramasco, J.J.; Tizzoni, M.; Colizza, V.; Vespignani, A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PloS one* **2011**, *6*, e16591.

31. Ferguson, N.M.; Cummings, D.A.; Fraser, C.; Cajka, J.C.; Cooley, P.C.; Burke, D.S. Strategies for mitigating an influenza pandemic. *Nature* **2006**, *442*, 448–452.

32. Keeling, M.; Rohani, P. Model. Infectious Diseases Humans Animals, 2008.

33. Frıas-Martınez, E.; Williamson, G.; Frıas-Martınez, V. Agent-based modelling of epidemic spreading using social networks and human mobility patterns. In Proceedings of the Proceedings of the 2011 IEEE third international conference on privacy, security, risk and trust and IEEE third international conference on social computing. Citeseer, 2011, pp. 57–64.

34. Cota, W.; Soriano-Paños, D.; Arenas, A.; Ferreira, S.C.; Gómez-Gardeñes, J. Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility. *New Journal of Physics* **2021**, *23*, 073019.

35. Balcan, D.; Gonçalves, B.; Hu, H.; Ramasco, J.J.; Colizza, V.; Vespignani, A. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of computational science* **2010**, *1*, 132–145.

36. Driessche, P.v. Spatial structure: patch models. In *Mathematical epidemiology*; Springer, 2008; pp. 179–189.

37. Herrera-Valdez, M.A.; Cruz-Aponte, M.; Castillo-Chavez, C. Multiple outbreaks for the same pandemic: Local transportation and social distancing explain the different" waves" of A-H1N1pdm cases observed in México during 2009. *Mathematical Biosciences & Engineering* **2011**, *8*, 21.

38. Khan, K.; Arino, J.; Hu, W.; Raposo, P.; Sears, J.; Calderon, F.; Heidebrecht, C.; Macdonald, M.; Liauw, J.; Chan, A.; et al. Spread of a novel influenza A (H1N1) virus via global airline transportation. *New England journal of medicine* **2009**, *361*, 212–214.

39. Bichara, D.; Kang, Y.; Castillo-Chavez, C.; Horan, R.; Perrings, C. SIS and SIR Epidemic Models Under Virtual Dispersal. *Bulletin of Mathematical Biology* **2015**, *77*, 2004,2034. https://doi.org/10.1007/s11538-015-0113-5.

40. Bichara, D.; Abderrahman, I. Multi-patch and multi-group epidemic models: a new framework. *Journal of Mathematical Biology* **2018**, *77*, 107,134. https://doi.org/10.1007/s00285-017-1191-9.

41. Riley, S. Large-scale spatial-transmission models of infectious disease. *Science* **2007**, *316*, 1298–1301.

42. Akuno, A.O.; Ramirez-Ramirez, L.L.; Mehta, C.; Krishnanuni, C.; Bui-Thanh, T.; Montoya, J.A. Multi-patch epidemic models with partial mobility, residency, and demography. *Preprint in Research Square* **2022**. Available at https://doi.org/10.21203/rs.3.rs-1954140/v1.

43. Arino, J.; van den Driessche, P. A multi-city epidemic model. *Mathematical Population Studies* **2003**, *10*, 175–193, [https://doi.org/10.1080/08898480306720]. https://doi.org/10.1080/08898480306720.

44. Brauer, F. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* **2017**, *2*, 113,127. https://doi.org/https://doi.org/10.1016/j.idm.2017.02.001.

45. Moualeu, D.; Bowong, S.; Tsanou, B.; Temgoua, A. A patchy model for the transmission dynamics of tuberculosis in sub-Saharan Africa. *International journal of dynamics and control* **2018**, *6*, 122–139.

46. Salmani, M.; van den Driessche, P. A model for disease transmission in a patchy environment. *Discrete & Continuous Dynamical Systems-B* **2006**, *6*, 185.

47. Phaijoo, G.R.; Gurung, D.B.; et al. Mathematical study of dengue disease transmission in multi-patch environment. *Applied Mathematics* **2016**, *7*, 1521.

48. Swallow, B.; Birrell, P.; Blake, J.; Burgman, M.; Challenor, P.; Coffeng, L.E.; Dawid, P.; De Angelis, D.; Goldstein, M.; Hemming, V.; et al. Challenges in estimation, uncertainty quantification and elicitation for pandemic modelling. *Epidemics* **2022**, *38*, 100547.

49. Capistran, M.A.; Capella, A.; Christen, J.A. Forecasting hospital demand in metropolitan areas during the current COVID-19 pandemic and estimates of lockdown-induced 2nd waves. *PloS one* **2021**, *16*, e0245669.

50. Wang, Y.; Wang, P.; Zhang, S.; Pan, H. Uncertainty Modeling of a Modified SEIR Epidemic Model for COVID-19. *Biology* **2022**, *11*, 1157.

51. Chen, P.; Wu, K.; Ghattas, O. Bayesian inference of heterogeneous epidemic models: Application to COVID-19 spread accounting for long-term care facilities. *Computer Methods in Applied Mechanics and Engineering* **2021**, *385*, 114020.

52. Huppert, A.; Katriel, G. Mathematical modelling and prediction in infectious disease epidemiology. *Clinical microbiology and infection* **2013**, *19*, 999–1005.

53. Márquez Urbina, J.U.; González Farías, G.; Ramírez Ramírez, L.L.; Rodríguez González, D.I. A multi-source global-local model for epidemic management. *Plos one* **2022**, *17*, e0261650.

54. Groendyke, C.; Welch, D.; Hunter, D.R. Bayesian inference for contact networks given epidemic data. *Scandinavian Journal of Statistics* **2011**, *38*, 600–616.

55. Dutta, R.; Mira, A.; Onnela, J.P. Bayesian inference of spreading processes on networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *474*, 20180129.

56. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of inverse problems*; Vol. 375, Springer Science & Business Media, 1996.

57. Tarantola, A. *Inverse problem theory and methods for model parameter estimation*; SIAM, 2005.

58. Vogel, C.R. *Computational methods for inverse problems (Frontiers in Applied Mathematics)*; SIAM, 2002.

59. Alexanderian, A. Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review. *Inverse Problems* **2021**, *37*, 043001.

60. Wesolowski, A.; Buckee, C.O.; Engø-Monsen, K.; Metcalf, C.J.E. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of infectious diseases* **2016**, *214*, S414–S420.

61. Wei, J.T.; Liu, Y.X.; Zhu, Y.C.; Qian, J.; Ye, R.Z.; Li, C.Y.; Ji, X.K.; Li, H.K.; Qi, C.; Wang, Y.; et al. Impacts of transportation and meteorological factors on the transmission of COVID-19. *International journal of hygiene and environmental health* **2020**, *230*, 113610.

62. Silva, P.C.; Batista, P.V.; Lima, H.S.; Alves, M.A.; Guimarães, F.G.; Silva, R.C. COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals* **2020**, *139*, 110088.

63. Cartenì, A.; Di Francesco, L.; Martino, M. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment* **2020**, *741*, 140489.

64. Ogen, Y. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. *Science of the Total Environment* **2020**, *726*, 138605.

65. He, G.; Pan, Y.; Tanaka, T. The short-term impacts of COVID-19 lockdown on urban air pollution in China. *Nature Sustainability* **2020**, *3*, 1005–1011.

66. Iacus, S.M.; Santamaria, C.; Sermi, F.; Spyratos, S.; Tarchi, D.; Vespe, M. Human mobility and COVID-19 initial dynamics. *Nonlinear Dynamics* **2020**, *101*, 1901–1919.

67. Alessandretti, L. What human mobility data tell us about COVID-19 spread. *Nature Reviews Physics* **2022**, *4*, 12–13.

68. Rahman, M.M.; Paul, K.C.; Hossain, M.A.; Ali, G.M.N.; Rahman, M.S.; Thill, J.C. Machine learning on the COVID-19 pandemic, human mobility and air quality: A review. *Ieee Access* **2021**, *9*, 72420–72450.

69. Hu, T.; Wang, S.; She, B.; Zhang, M.; Huang, X.; Cui, Y.; Khuri, J.; Hu, Y.; Fu, X.; Wang, X.; et al. Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. *International Journal of Digital Earth* **2021**, *14*, 1126–1147.

70. Lin, C.H.; Wen, T.H. How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission. *Tropical Medicine and Infectious Disease* **2022**, *7*, 164.

71. Perez, L.; Dragicevic, S. An agent-based approach for modeling dynamics of contagious disease spread. *International journal of health geographics* **2009**, *8*, 1–17.

72. Lazebnik, T.; Blumrosen, G. Advanced multi-mutation with intervention policies pandemic model. *IEEE Access* **2022**, *10*, 22769–22781.

73. Fudolig, M.; Howard, R. The local stability of a modified multi-strain SIR model for emerging viral strains. *PloS one* **2020**, *15*, e0243408.

74. Yusuf, A.; Qureshi, S.; Inc, M.; Aliyu, A.I.; Baleanu, D.; Shaikh, A.A. Two-strain epidemic model involving fractional derivative with Mittag-Leffler kernel. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **2018**, *28*, 123121.

75. Lazebnik, T.; Bunimovich-Mendrazitsky, S. Generic approach for mathematical model of multi-strain pandemics. *PloS one* **2022**, *17*, e0260683.

76. Khyar, O.; Allali, K. Global dynamics of a multi-strain SEIR epidemic model with general incidence rates: application to COVID-19 pandemic. *Nonlinear dynamics* **2020**, *102*, 489–509.

77. Conacyt. Covid-19 México, 2020. Accessed in, 2022.

78. Del Valle, S.Y.; McMahon, B.H.; Asher, J.; Hatchett, R.; Lega, J.C.; Brown, H.E.; Leany, M.E.; Pantazis, Y.; Roberts, D.J.; Moore, S.; et al. Summary results of the 2014-2015 DARPA Chikungunya challenge. *BMC infectious diseases* **2018**, *18*, 1–14.

79. Lai, C.C.; Liu, Y.H.; Wang, C.Y.; Wang, Y.H.; Hsueh, S.C.; Yen, M.Y.; Ko, W.C.; Hsueh, P.R. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *Journal of Microbiology, Immunology and Infection* **2020**, *53*, 404–412.

80. Kalajdzievska, D.; Li, M.Y. Modeling the effects of carriers on transmission dynamics of infectious diseases. *Math. Biosci. Eng* **2011**, *8*, 711–722.

81. Duffy, M.R.; Chen, T.H.; Hancock, W.T.; Powers, A.M.; Kool, J.L.; Lanciotti, R.S.; Pretrick, M.; Marfel, M.; Holzbauer, S.; Dubray, C.; et al. Zika virus outbreak on Yap Island, federated states of Micronesia. *New England Journal of Medicine* **2009**, *360*, 2536–2543.

82. Doll, M.E.; Pryor, R.; Mackey, D.; Doern, C.D.; Bryson, A.; Bailey, P.; Cooper, K.; Godbout, E.; Stevens, M.P.; Bearman, G. Utility of retesting for diagnosis of SARS-CoV-2/COVID-19 in hospitalized patients: Impact of the interval between tests. *Infection Control & Hospital Epidemiology* **2020**, *41*, 859–861.

83. Esbin, M.N.; Whitney, O.N.; Chong, S.; Maurer, A.; Darzacq, X.; Tjian, R. Overcoming the bottleneck to widespread testing: a rapid review of nucleic acid testing approaches for COVID-19 detection. *Rna* **2020**, *26*, 771–783.

84. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **2020**, *368*, 489–493.

85. Team, E. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC weekly* **2020**, *2*, 113.

86. Trejo, I.; Hengartner, N.W. A modified Susceptible-Infected-Recovered model for observed under-reported incidence data. *PloS one* **2022**, *17*, e0263047.

87. Hasell, J.; Mathieu, E.; Beltekian, D.; Macdonald, B.; Giattino, C.; Ortiz-Ospina, E.; Roser, M.; Ritchie, H. A cross-country database of COVID-19 testing. *Scientific data* **2020**, *7*, 345.

88. Tocto-Erazo, M.R.; Espíndola-Zepeda, J.A.; Montoya-Laos, J.A.; Acuña-Zegarra, M.A.; Olmos-Liceaga, D.; Reyes-Castro, P.A.; Figueroa-Preciado, G. Lockdown, relaxation, and acme period in COVID-19: A study of disease dynamics in Hermosillo, Sonora, Mexico. *PloS one* **2020**, *15*, e0242957.

89. Lin, Q.; Zhao, S.; Gao, D.; Lou, Y.; Yang, S.; Musa, S.S.; Wang, M.H.; Cai, Y.; Wang, W.; Yang, L.; et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International journal of infectious diseases* **2020**, *93*, 211–216.

90. Liu, X.; Hewings, G.; Wang, S.; Qin, M.; Xiang, X.; Zheng, S.; Li, X. Modelling the situation of COVID-19 and effects of different containment strategies in China with dynamic differential equations and parameters estimation. *MedRxiv* **2020**, pp. 2020–03.

91. Chatterjee, K.; Chatterjee, K.; Kumar, A.; Shankar, S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India* **2020**, *76*, 147–155.

92. Organization, W.H.; et al. Coronavirus disease 2019 (COVID-19): situation report, 73, 2020.

93. Ganyani, T.; Kremer, C.; Chen, D.; Torneri, A.; Faes, C.; Wallinga, J.; Hens, N. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **2020**, *25*, 2000257.

94. Linton, N.M.; Kobayashi, T.; Yang, Y.; Hayashi, K.; Akhmetzhanov, A.R.; Jung, S.m.; Yuan, B.; Kinoshita, R.; Nishiura, H. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine* **2020**, *9*, 538.

95. Wu, J.T.; Leung, K.; Leung, G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* **2020**, *395*, 689–697.

96. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine* **2020**, *172*, 577–582.

97. Knoema. World data atlas, Mexico, Sonora mortality. https://knoema.com/atlas/Mexico/Sonora/Crude-Death-Rate, 2021.

98. Torres-Ibarra, L.; Basto-Abreu, A.; Carnalla, M.; Torres-Alvarez, R.; Reyes-Sanchez, F.; Hernández-Ávila, J.E.; Palacio-Mejia, L.S.; Alpuche-Aranda, C.; Shamah-Levy, T.; Rivera, J.A.; et al. SARS-CoV-2 infection fatality rate after the first epidemic wave in Mexico. *International journal of epidemiology* **2022**, *51*, 429–439.

99. Organization, W.H.; et al. COVID-19 natural immunity: scientific brief, 10 May 2021. Technical report, World Health Organization, 2021.

100. Pilz, S.; Chakeri, A.; Ioannidis, J.P.; Richter, L.; Theiler-Schwetz, V.; Trummer, C.; Krause, R.; Allerberger, F. SARS-CoV-2 re-infection risk in Austria. *European journal of clinical investigation* **2021**, *51*, e13520.

101. Hansen, C.H.; Michlmayr, D.; Gubbels, S.M.; Mølbak, K.; Ethelberg, S. Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study. *The lancet* **2021**, *397*, 1204–1212.

102. Kim, P.; Gordon, S.M.; Sheehan, M.M.; Rothberg, M.B. Duration of SARS-CoV-2 natural immunity and protection against the Delta variant: a retrospective cohort study. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **2021**.

103. Spicer, K.B.; Glick, C.; Cavanaugh, A.M.; Thoroughman, D. Protective immunity after natural infection with severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2)–Kentucky, USA, 2020. *International Journal of Infectious Diseases* **2022**, *114*, 21–28.

104. Sheehan, M.M.; Reddy, A.J.; Rothberg, M.B. Reinfection rates among patients who previously tested positive for coronavirus disease 2019: a retrospective cohort study. *Clinical Infectious Diseases* **2021**, *73*, 1882–1886.

105. Mahase, E. Covid-19: Past infection provides 83% protection for five months but may not stop transmission, study finds, 2021.

106. Abu-Raddad, L.J.; Chemaitelly, H.; Coyle, P.; Malek, J.A.; Ahmed, A.A.; Mohamoud, Y.A.; Younuskunju, S.; Ayoub, H.H.; Al Kanaani, Z.; Al Kuwari, E.; et al. SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy. *EClinicalMedicine* **2021**, *35*, 100861.

107. Harvey, R.A.; Rassen, J.A.; Kabelac, C.A.; Turenne, W.; Leonard, S.; Klesh, R.; Meyer, W.A.; Kaufman, H.W.; Anderson, S.; Cohen, O.; et al. Association of SARS-CoV-2 seropositive antibody test with risk of future infection. *JAMA internal medicine* **2021**, *181*, 672–679.

108. Pilz, S.; Theiler-Schwetz, V.; Trummer, C.; Krause, R.; Ioannidis, J.P. SARS-CoV-2 reinfections: Overview of efficacy and duration of natural and hybrid immunity. *Environmental research* **2022**, *209*, 112911.

109. Capaldi, A.; Behrend, S.; Berman, B.; Smith, J.; Wright, J.; Lloyd, A.L. Parameter estimation and uncertainty quantication for an epidemic model. *Mathematical biosciences and engineering* **2012**, p. 553.

110. Anh, D.T.; Bonnet, M.P.; Vachaud, G.; Van Minh, C.; Prieur, N.; Duc, L.V.; et al. Biochemical modeling of the Nhue River (Hanoi, Vietnam): Practical identifiability analysis and parameters estimation. *Ecological modelling* **2006**, *193*, 182–204.

111. Bellman, R.; Åström, K.J. On structural identifiability. *Mathematical biosciences* **1970**, *7*, 329–339.

112. Evans, N.D.; White, L.J.; Chapman, M.J.; Godfrey, K.R.; Chappell, M.J. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical biosciences* **2005**, *194*, 175–197.

113. Xia, X.; Moog, C.H. Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE transactions on automatic control* **2003**, *48*, 330–336.

114. White, L.; Evans, N.; Lam, T.; Schukken, Y.; Medley, G.; Godfrey, K.; Chappell, M. The structural identifiability and parameter estimation of a multispecies model for the transmission of mastitis in dairy cows. *Mathematical biosciences* **2001**, *174*, 77–90.

115. Reid, J. Structural identifiability in linear time-invariant systems. *IEEE Transactions on Automatic Control* **1977**, *22*, 242–246.

116. Prieto, K.; Dorn, O. Sparsity and level set regularization for diffuse optical tomography using a transport model in 2D. *Inverse Problems* **2016**, *33*, 014001.

117. Prieto, K.; Ibarguen-Mondragon, E. Parameter estimation, sensitivity and control strategies analysis in the spread of influenza in Mexico. In Proceedings of the Journal of Physics: Conference Series. IOP Publishing, 2019, Vol. 1408, pp. 012–020.

118. Smirnova, A.; DeCamp, L.; Liu, H. Inverse problems and ebola virus disease using an age of infection model. In *Mathematical and Statistical Modeling for Emerging and Re-emerging Infectious Diseases*; Springer, 2016; pp. 103–121.

119. Capistrán, M.A.; Moreles, M.A.; Lara, B. Parameter estimation of some epidemic models. The case of recurrent epidemics caused by respiratory syncytial virus. *Bulletin of mathematical biology* **2009**, *71*, 1890–1901.

120. Kraft, D. A software package for sequential quadratic programming. *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt* **1988**.

121. Spataru, D. Using a Genetic Algorithm for Parameter Estimation in a Modified SEIR Model of COVID-19 Spread in Ontario. PhD thesis, University of Guelph, 2021.

122. Whitley, D. A genetic algorithm tutorial. *Statistics and computing* **1994**, *4*, 65–85.

123. Liu, Q.; Wang, D. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems* **2016**, *29*.

124. Lindén, A.; Mäntyniemi, S. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* **2011**, *92*, 1414–1421.

125. Bliznashki, S. A Bayesian logistic growth model for the spread of COVID-19 in New York. *medRxiv* **2020**.

126. Arguedas, Y.N.; Santana-Cibrian, M.; Velasco-Hernández, J.X. Transmission dynamics of acute respiratory diseases in a population structured by age. *Mathematical Biosciences and Engineering* **2019**, *16*, 7477–7493.

127. Neyens, T.; Faes, C.; Molenberghs, G. A generalized Poisson-gamma model for spatially overdispersed data. *Spatial and spatio-temporal epidemiology* **2012**, *3*, 185–194.

128. Coly, S.; Yao, A.F.; Abrial, D.; Charras-Garrido, M. Distributions to model overdispersed count data. *Journal de la Société Française de Statistique* **2016**, *157*, 39–63.

129. Gutiérrez-Pulido, H.; Aguirre-Torres, V.; Christen, J.A. A practical method for obtaining prior distributions in reliability. *IEEE Transactions on Reliability* **2005**, *54*, 262–269.

130. Gutiérrez-Pulido, H.; Aguirre-Torres, V.; Christen, J.A. A Bayesian approach for the determination of warranty length. *Journal of quality Technology* **2006**, *38*, 180–189.

131. Gutiérrez Pulido, H.; Gutiérrez González, P. Fundamentos y aplicaciones de la estadística Bayesiana. *México: Universidad de Guadalajara* **2013**.

132. Grinsztajn, L.; Semenova, E.; Margossian, C.C.; Riou, J. Bayesian workflow for disease transmission modeling in Stan. *Statistics in medicine* **2021**, *40*, 6209–6234.

133. Christen, J.A.; Fox, C. A General Purpose Sampling Algorithm for Continuous Distributions (the t-walk). *Bayesian Analysis* **2010**, *5*, 263–282.

134. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *Journal of statistical software* **2017**, *76*.

135. Chatzilena, A.; van Leeuwen, E.; Ratmann, O.; Baguelin, M.; Demiris, N. Contemporary statistical inference for infectious disease models using Stan. *Epidemics* **2019**, *29*, 100367.

136. Banks, D.L.; Hooten, M.B. Statistical challenges in agent-based modeling. *The American Statistician* **2021**, *75*, 235–242.

137. Ju, N.; Heng, J.; Jacob, P.E. Sequential Monte Carlo algorithms for agent-based models of disease transmission. *arXiv preprint arXiv:2101.12156* **2021**.

138. Shiono, T. Estimation of agent-based models using Bayesian deep learning approach of BayesFlow. *Journal of Economic Dynamics and Control* **2021**, *125*, 104082.

139. Manzo, G. *Agent-based Models and Causal Inference*; John Wiley & Sons, 2022.

140. Dyer, J.; Cannon, P.; Farmer, J.D.; Schmon, S. Black-box Bayesian inference for economic agent-based models. *arXiv preprint arXiv:2202.00625* **2022**.

141. Dyer, J.; Cannon, P.; Farmer, J.D.; Schmon, S.M. Calibrating agent-based models to microdata with graph neural networks. *arXiv preprint arXiv:2206.07570* **2022**.

142. Britton, T. Epidemic models on social networks—With inference. *Statistica Neerlandica* **2020**, *74*, 222–241.

143. Bu, F.; Aiello, A.E.; Xu, J.; Volfovsky, A. Likelihood-based inference for partially observed epidemics on dynamic networks. *Journal of the American Statistical Association* **2022**, *117*, 510–526.

144. Murphy, C.; Laurence, E.; Allard, A. Deep learning of contagion dynamics on complex networks. *Nature Communications* **2021**, *12*, 4720.