Infectious Disease Modelling 8 (2023) 704–716

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

EpiMix: A novel method to estimate effective reproduction number

Shihui Jin [a], Borame Lee Dickens [a], Jue Tao Lim [b], Alex R. Cook [a, c, *]

[a] Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore
[b] Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
[c] Department of Statistics and Data Science, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 16 April 2023
Received in revised form 14 June 2023
Accepted 14 June 2023
Available online 20 June 2023
Handling Editor: Dr Daihai He

Keywords:
Epidemics
INLA
Regression
Reproduction number
SARS-CoV-2
Transmission dynamics

ABSTRACT

Transmission potential of a pathogen, often quantified by the time-varying reproduction number R_t, provides the current pace of infection for a disease and indicates whether an emerging epidemic is under control. In this study, we proposed a novel method, EpiMix, for R_t estimation, wherein we incorporated the impacts of exogenous factors and random effects under a Bayesian regression framework. Using Integrated Nested Laplace Approximation, EpiMix is able to efficiently generate reliable, deterministic R_t estimates. In the simulations and case studies performed, we further demonstrated the method's robustness in low-incidence scenarios, together with other merits, including its flexibility in selecting variables and tolerance of varying reporting rates. All these make EpiMix a potentially useful tool for real-time R_t estimation provided that the serial interval distribution, time series of case counts and external influencing factors are available.

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The recurrence of infectious diseases can cause great losses to both a country's economy and society (Allen et al., 2017). COVID-19, for example, has been an on-going global pandemic (WHO, 2020) for over three years to date after the first known case appeared in Wuhan, China in December 2019 (Jiang et al., 2020). Countries around the world have gone through several epidemic waves with the emergence of the Delta and Omicron variants (Kupferschmidt & Wadman, 2021; Viana et al., 2022), despite increasing vaccination coverage and various nonpharmaceutical interventions being implemented. The downside to disease management and containment measures has been the substantial healthcare, economic, social and environmental burdens that resulted (Chakraborty & Maity, 2020; Kano et al., 2021; Pfefferbaum & North, 2020), requiring the understanding of epidemic trajectories for the effective planning of resource distributions. The modelling and forecasting of ongoing virus transmission rates is therefore necessary to guide policymakers in establishing appropriate control measures and epidemic exit strategies.

Epidemic transmissibility can be measured using the basic reproduction number R_0 (Dietz, 1993), defined as the average number of secondary cases generated by an infected individual in an otherwise fully susceptible population. The time-varying

* Corresponding author. #10-01 Tahir Foundation Building, 12 Science Drive 2, 117549, Singapore.
E-mail address: ephcar@nus.edu.sg (A.R. Cook).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2023.06.002
2468-0427/© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

analogue, R_t, represents the population-level transmission potential at time t, tracking the evolution of disease transmissibility throughout the epidemic. The metric R_t is widely used for assessing epidemic control, with a threshold of 1 or greater indicating epidemic growth.

A variety of methods have been proposed to estimate R_t. A likelihood-based approach was utilized by Wallinga & Teunis, 2004 (Wallinga & Teunis, 2004) to reconstruct infection networks and derive R_t estimates for the original SARS epidemic, while EpiEstim by Cori et al. (Cori et al., 2013) calculates real-time R_t estimates with a Poisson model from a backward-looking perspective. Both methods are simple and generic, making few assumptions using incidence data and serial interval distributions but struggling in providing estimates when case counts are low. This limitation, however, was addressed by EpiFilter, a recursive Bayesian smoother presented by Parag (Parag, 2021), who introduced constraints on the variation between neighboring R_t s.

Another method for R_t estimation, proposed by Flaxman et al. in 2020, utilizes death data and obtains estimates of transmission potential from the product of the basic reproduction number and effects of diverse interventions, assuming this would have a direct and immediate impact on R_t s (Flaxman et al., 2020). Although reported deaths are a more transparent data source compared to daily case counts, time lags from infection to death are an important issue. Furthermore, when the number of deaths is small, stochasticity may make estimation challenging.

In previous work, the authors developed EpiRegress to provide real-time R_t estimation (Jin et al., 2022), wherein the time-varying reproduction number is taken to be the product of the effects of a diverse set of exogenous factors, including but not limited to mobility patterns and non-pharmaceutical interventions, and is linked to daily case counts through a negative binomial relationship. EpiRegress showed that the introduction of external data can effectively reduce uncertainty in the estimation of R_t in low-incidence scenarios. Fluctuations in R_t, however, may not always be fully captured due to the complex interplay of assorted social, economic, and political factors.

In this paper, we extend the work of Cori et al., Parag, Flaxman et al., and Jin et al. and propose a new method, EpiMix, to estimate R_t. Similar to EpiRegress, EpiMix is constructed using a Bayesian generalized linear regression framework, but with random noise incorporated for each individual time step. We use Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009, 2023) for parameter estimation. The algorithm's high computational speed enables us to obtain estimates within seconds, making EpiMix a practical and efficient tool for the nowcasting of real-time transmission potential.

We further demonstrate the robustness of EpiMix by leveraging case studies of COVID-19 incidence in Singapore and New Zealand, along with simulations, which involves different datasets for inference where changes in reporting rates occur. An R function has been made available, EpiMix (link: https://github.com/ShihuiJin/EpiMix), which can be utilized to estimate epidemic trajectories and the risk of resurgence.

2. Materials and methods

We begin by describing the formulation of the statistical model EpiMix which uses a Bayesian generalized linear regression framework incorporating the effects of exogenous factors and random effects. We then describe a simulation study in which we simulate data generated by known, synthetic R_t s, which are designed around empirical data from New South Wales, Australia. In the simulation study, we then fit EpiMix to the synthetic outbreaks to evaluate estimation errors. Finally, we apply EpiMix to estimate R_t for two low-incidence settings with unknown R_t s: Singapore and New Zealand.

2.1. Statistical model

Following Cori et al.'s work (Cori et al., 2013), we denote the serial interval distribution as w_s, which is the probability that one infected individual infects a susceptible person at time step (which we take to be days) $s = 1, 2, 3\cdots$ after being infected (w_s is a probability mass function of s that only take positive values at positive integer points and $\sum_{s=1}^{\infty} w_s = 1$). The total infection potential at time step t, Λ_t, can therefore be expressed as

$$\Lambda_t = \sum_{s=1}^{t-1} w_s I_{t-s}, \tag{1}$$

where I_t is the total number of new cases at time t.

The expected number of local cases reported at time step t is the product of R_t and the total infection potential at that time, Λ_t, i.e.,

$$\mathbb{E}I_t = R_t \Lambda_t = R_t \sum_{s=1}^{t-1} w_s I_{t-s}. \tag{2}$$

With the given case counts and corresponding serial interval, under the assumption that the reporting rate—the proportion of newly notified cases among all new infections—is always constant (though see later for a sensitivity analysis), the number of cases reported at time step t, I_t, is assumed to follow a Poisson distribution with mean $R_t \Lambda_t$. In short,

$$\left(I_t \middle| I_{1:(t-1)}, w_s, R_t\right) \sim \text{Poisson}(R_t \Lambda_t). \tag{3}$$

Meanwhile, by assuming that R_t is being driven by various exogenous factors at time step t, such as people's mobility patterns and policies implemented, we consider the following relationship between R_t and the corresponding covariates $\boldsymbol{X}_t$. as

$$\log R_t = \boldsymbol{X}_t . \boldsymbol{\beta} + \alpha + \varepsilon_t, \tag{4}$$

where α is the intercept, $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$ is the vector of regression coefficients and the ε_t s are independent and identically Gaussian distributed random variables with mean 0 and unknown variance σ^2. Each $\log \mu_t = \mathbb{E}(\log R_t)$ is thus a linear combination of the corresponding $\boldsymbol{X}$ covariates' values at time step t for the vector $\boldsymbol{X_t}.$, while R_t can be taken as a function of the covariate coefficients and random effect. Both α and each entry of $\boldsymbol{\beta}$ are assigned an uninformative Gaussian prior with mean 0 and variance 1000 while the prior for σ^2 is a Gamma distribution with shape and rate parameters as 1 and 0.00005 respectively.

The generalized linear model is outlined in Fig. 1. The parameters α, $\boldsymbol{\beta}$ and the R_t s are estimated concurrently using INLA (Rue et al., 2017, 2023), wherein the marginal posterior distribution of each individual parameter is derived through Laplace approximation with a grid strategy. We also compute 95% credible intervals (CrI) to reflect the uncertainty of the model parameters.

2.2. Datasets used

For covariates which impact R_t, we collected known extraneous drivers including mobility, epidemiological and policy-related data to build a covariate matrix $\boldsymbol{X}$, described later. In simulations to assess the model's performance, we used datasets for New South Wales for one year's time period from October 1, 2020 to September 30, 2021. For mobility data, Google's Community Mobility Reports (Google, 2020) were used which record differences in overall visitation patters across time. For policy-related data, the Oxford COVID-19 Government Response Tracker (OxCGRT) (Hale et al., 2021) was utilized, providing detail on travel restriction and non-pharmaceutical interventions. Detailed descriptions of the 20 variables included from these two sources are listed in the supplementary materials (SI Table S1).

Information of daily vaccine doses administered for New South Wales was obtained from the Australian website COVID LIVE (Macali, 2023), which compiled data from media releases and websites of state health departments. The proportion of cases of the more transmissible B.1.617.2 (Delta) variant (Liu & Rocklöv, 2021) was also included as a candidate factor from GISAID (Re3data.Org, 2012). Additionally, data for three weather covariates—daily maximum temperature, rainfall and solar exposure—were obtained from the Bureau of Meteorology, Australia (Bureau of Meteorology Australia, 2021). To better evaluate the impact of each individual covariate, columns in the $\boldsymbol{X}$ matrices were standardized to mean 0, variance 1 prior to analysis.

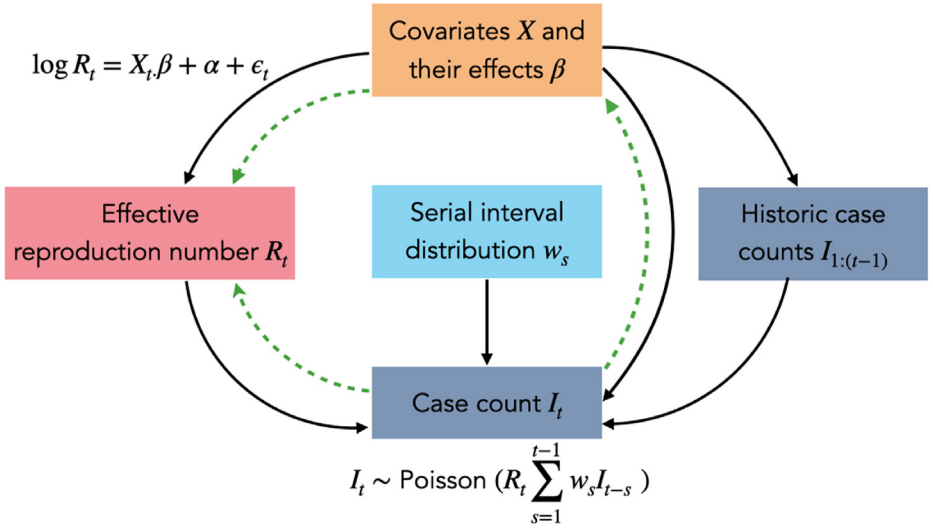

Fig. 1. Model schematic: relationship between different variables. Black arrowed lines represent the relationship in reality, while the green dashed, arrowed lines indicate how the parameters are estimated.

Analogous datasets were collected for two case studies for Singapore and New Zealand except for vaccination data which were extracted from covidvax.live (Redfern, 2021), an online platform that shows real-time statistics on vaccine doses registered worldwide. Daily COVID-19 case counts from January 01, 2021 until December 31, 2021 were obtained from government health websites (Ministry of Health New Zealand, 2021; Ministry of Health Singapore, 2021). These two countries were chosen as examples with stringent containment measures and relatively small outbreak sizes, exhibiting lower case counts (Fig. S1) which allows for EpiMix's performance to be evaluated for the more challenging situation of low case counts. For Singapore, cases reported in migrant worker dormitories in Singapore were not taken into account as they were largely segregated from the rest of the community at the time (Ministry of Manpower Singapore, 2023) and the high density accommodation settings caused different transmission patterns. We also assumed for the main analysis that imported cases were largely captured and isolated as border controls were highly restrictive for both countries over this period. In subsequent sensitivity analyses however, we tested this assumption.

3. Simulations

Since R_t s are not observable in reality, we used simulations to validate EpiMix's capacity in recovering the transmission potential. We also assessed the impact of violations of the original settings of the model, including omitted or extraneous covariates and changing the default inference window or the reporting rates. We also tested the approach in a low-incidence scenario to see how much the precision would be degraded by small denominators.

To achieve these goals, we began by creating pseudo datasets of one year's μ_t s (i.e., exponential of the expected logarithm of R_t s) from the mobility, epidemiological and policy data of New South Wales between October 1, 2020 and September 30, 2021. Based on the simulated datasets, we further generated R_t s and incidence curves which we used for R_t estimation and comparison. The involved methods and outcomes are as follows.

3.1. Generation of simulated $\boldsymbol{\mu_t}$, $\boldsymbol{R_t}$ and $\boldsymbol{I_t}$ values

We derived the covariate matrix $\boldsymbol{X}$ from datasets of New South Wales aforementioned, standardizing each column and removing columns with constant values. We arbitrarily selected a corresponding coefficient vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_p)^T$. Together with a simulated intercept α, we calculated the log-expectation of the R_t s, represented as $\log \mu_t$, by

$$\log \mu_t = \boldsymbol{X}_{t\bullet} \times \boldsymbol{\beta} + \alpha = \alpha + \sum_{i=1}^{p} X_{ti}\beta_i, \tag{5}$$

where $t = 1, 2, 3\cdots365$ is the day involved. Based on the simulated μ_t s, we further obtained the $R_t s$ by

$$R_t = \mu_t \times \exp(\varepsilon_t), \tag{6}$$

where ε_t s are randomly generated values from a Gaussian distribution with mean 0 and standard deviation 0.3. Values for these parameters were chosen to create a plausible R_t function fluctuating around 1.

By assigning an initial value to I_0, the subsequent daily case count, I_t $(t > 1)$ was therefore generated as

$$I_t \sim \text{Poisson}\left(R_t \sum_{s=1}^{\min(t,50)} w_s I_{t-s}\right) \tag{7}$$

where $\{w_s\}$ is the vector for the serial interval distribution.

3.2. Evaluation of the estimation results

Let $\widehat{R}_t$ be the point estimate of the effective reproduction number at time t and R_t^{sim} be the corresponding simulated 'true' values. To assess the model fits, we considered mean absolute error (MAE) and mean percentage error (MAPE). More specifically, over the time interval $[t_1, t_2]$ when we wished to check the estimation errors, we calculated

$$MAE_R([t_1, t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t=t_2} \left|\widehat{R}_t - R_t^{sim}\right|, \tag{8}$$

and

$$MAPE_R([t_1, t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t=t_2} \left|\widehat{R}_t - R_t^{sim}\right| \Big/ R_t^{sim}. \tag{9}$$

We further used successful coverage rate (SCR) to assess the percentage of R_t^{sim} s that fell in the 95% credible interval (CrI) of the posterior predictive distribution of R_t, $[\widehat{R}_t^{0.025}, \widehat{R}_t^{0.975}]$, where $\widehat{R}_t^{\alpha}$ denotes the α-th quantile of that distribution. The SCR for R_t over the interval $[t_1, t_2]$ was then

$$SCR_R([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \mathbf{1}_{R_t^{sim} \in [\widehat{R}_t^{0.025}, \widehat{R}_t^{0.975}]}. \tag{10}$$

Similarly, for the point estimate of $\mu_t = \exp[\mathbb{E}(\log R_t)]$, $\widehat{\mu}_t$, the 95% CrI $[\widehat{\mu}_t^{0.025}, \widehat{\mu}_t^{0.975}]$ and the simulated true μ_t^{sim}, we had

$$MAE_\mu([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \left| \widehat{\mu}_t - \mu_t^{sim} \right|, \tag{11}$$

$$MAPE_I([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \left| \widehat{\mu}_t - \mu_t^{sim} \right| \Big/ \mu_t^{sim}, \tag{12}$$

and

$$SCR_\mu([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \mathbf{1}_{\mu_t^{sim} \in [\widehat{\mu}_t^{0.025}, \widehat{\mu}_t^{0.975}]}. \tag{13}$$

We also had, for the estimated expected number of case count on day t, $\widehat{I}_t$ and the corresponding I_t^{sim},

$$MAE_I([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \left| \widehat{I}_t - I_t^{sim} \right|, \tag{14}$$

and

$$MAPE_I([t_1,t_2]) = \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \left| \widehat{I}_t - I_t^{sim} \right| \Big/ I_t^{sim}. \tag{15}$$

3.3. Validation of EpiMix using simulated datasets

Using the method above and one randomly generated set of $(\alpha, \boldsymbol{\beta})$, we simulated 10 synthetic datasets of R_t s (Fig. S2) and for each time series of R_t s 10 corresponding sets of daily case counts. For every synthetic incidence curve, we performed estimation with EpiMix and compared estimates with simulated values (Fig. S3). On average, we found the mean absolute error for μ_t estimates to be 0.11 (11.9% of true values) and for R_t 0.18 (21.1%). Predicted case counts (mean of the Poisson distribution) had an average of 1.2 (0.16%) deviation from the observed values. The SCR for R_t estimates were all above 95%.

3.4. Adaptable variable selection

We simulated an incidence curve with a maximum case count smaller than 400 (Fig. S4), based on which we evaluated influence of changes in variable selection on estimations by EpiMix. We ran different sets of covariates to form six model candidates (Table S2). The candidates (1), (5) and (6) represent scenarios where a plethora of data are available for modelling whereas (2) – (4) were run to compare model performance with data scarcity. The versions of the $\boldsymbol{X}$ matrix we used to draw inference are as follows:

(1) Full: with all covariates (24 in total);
(2) Mobility-driven: only mobility related covariates;
(3) Reduced: all but covariates on health measures;
(4) Hybrid: subset of the full model with half of the covariates from each category (mobility, epidemiological, closure, and containment, economic measures and health measures);
(5) Weather-added: with 3 additional weather covariates apart from those in the full model;
(6) Mis-specified: with 3 additional weather covariates but lacking covariates regarding health measures.

When comparing between estimates and observations (Table 1), fits of R_t and μ_t were equally good for all but the mobility-driven model (2) and the hybrid subset model (4). More specifically, for all the models, around 95% of the observations fell in the estimated 95% CrIs for R_t and the MAE of R_t estimations was approximately 0.18, which covered 21% of the observed values. The MAEs for μ_t estimates were all around 0.11 for different models but only 65% and 83% of the estimates for the

Table 1
Fits of R_t and μ_t in terms of mean absolute error (MAE), mean absolute percentage error (MAPE) and successful coverage rate (SCR) when different models were used for estimation. The assessment window was from day 1 to day 365 and posterior means were used as point estimates.

Model Variant	Fit of $\boldsymbol{R_t}$: MAE (MAPE)	Fit of $\boldsymbol{R_t}$: SCR(%)	Fit of $\boldsymbol{\mu_t}$: MAE (MAPE)	Fit of $\boldsymbol{\mu_t}$: SCR (%)
Full	0.182 (21.1%)	97.8	0.110 (11.9%)	96.7
Mobility-driven	0.186 (21.2%)	97.3	0.116 (11.7%)	65.2
Reduced	0.182 (21.1%)	97.8	0.115 (12.3%)	95.3
Hybrid	0.183 (21.1%)	94.8	0.119 (12.1%)	82.7
Weather-added	0.180 (20.9%)	97.0	0.114 (12.3%)	96.4
Mis-specified	0.181 (20.9%)	97.0	0.103 (11.0%)	96.7

mobility-driven model and the hybrid subset model fell within the 95% CrIs, which is significantly lower than the expected proportion of 95%.

Since multiple covariates in the $\boldsymbol{X}$ matrix are likely to be strongly correlated and the exclusion of various factors may not substantially influence estimation preciseness, we applied Principal Component Analysis (PCA) to the $\boldsymbol{X}$ matrices (Fig. S5) in versions (1)–(6) to obtain scenarios (7)–(12) by utilizing the formula

$$\boldsymbol{X}^{\mathrm{PCA}} = \boldsymbol{X} \times \boldsymbol{A}, \tag{16}$$

where $\boldsymbol{A}$ is the rotation matrix made up of the eigenvectors of $\boldsymbol{X}$.

We then selected the first 15 or 10 principal components for each model (i.e., the first 15 or 10 columns of $\boldsymbol{X}^{PCA}$) and performed estimations with EpiMix respectively (SI Table S3, S4). Fits were alike across the model variants, and the SCRs of R_t estimates remained high (>95%) despite decreases in the number of PCs selected. For μ_t s however, fewer simulated values fell in the 95% CrIs when the number of selected PCs was reduced. More specifically, for the full, weather-added and mis-specified models, in which the number of covariates far exceeded 15, SCR fell by around 5% when we used 15 PCs in comparison to no PCA, and when the number of PCs was reduced from 15 to 10.

To explore EpiMix's performance when we omitted from the fitted model some PCs which contributed to R_t, we additionally simulated a similar set of μ_t s from the first 15 columns of $\boldsymbol{X}^{PCA}$, i.e., the first 15 PCs of the original $\boldsymbol{X}$ covariates, and generated a new time series of R_t s as well as cases using the same noise (Fig. S6). We then estimated R_t s using the unadjusted full model and with 5, 10 and 15 PCs (Table 2). This showed that R_t did not significantly differ when a reduced set of the more important PCs were used. Estimations of μ_t s however had the greatest MAE (as well as MAPE) without PCA, though the SCR was the highest among the four model variants. To further investigate the exact relationship between number of uncorrelated variables included and estimation errors, we gradually increased the number of PCs in the model and performed estimations with EpiMix, finding a decreasing trend in the MAPE for μ_t estimates for the first few increments (Fig. 2).

3.5. The impact of inference window length

To assess the influence of inference window length on estimation accuracy, we chose a set of candidate window lengths, $T = \{90,120,150,\ldots,330,360\}$, and evaluated each candidate's performance. Using the same incidence curve as we did when testing variable selection (Fig. S5), for a fixed window of $\Delta t \in T$ days, we estimated the R_t s and μ_t s with EpiMix for the time intervals $[1,\Delta t]$, $[2,1+\Delta t]$, $[3,2+\Delta t]\ldots[366-\Delta t,365]$. We then calculated the MAE and MAPE for each estimation and obtain the mean error for each inference window length as the mean of MAE or MAPE for R_t and μ_t by averaging over the different time intervals involved respectively, i.e.,

$$Err_R(\Delta t) = \frac{1}{366-\Delta t}\sum_{i=1}^{366-\Delta t} MAE_R([i, i-1+\Delta t]), \tag{17}$$

Table 2
Fits of R_t and μ_t in terms of mean absolute error (MAE), mean absolute percentage error (MAPE) and successful coverage rate (SCR) when different model variants were used for estimation and when X^{PCA} were used to simulate case counts. The assessment window was from day 1 to day 365 and posterior means were used as point estimates.

Model Variant	Fit of $\boldsymbol{R_t}$: MAE (MAPE)	Fit of $\boldsymbol{R_t}$: SCR(%)	Fit of $\boldsymbol{\mu_t}$: MAE (MAPE)	Fit of $\boldsymbol{\mu_t}$: SCR (%)
Full model	0.140 (15.7%)	95.3	0.117 (12.6%)	95.1
Model with all the 15 PCs	0.140 (15.7%)	95.3	0.105 (11.5%)	89.3
Model with the first 10 PCs	0.142 (15.7%)	95.3	0.095 (10.5%)	88.2
Model with the first 5 PCs	0.139 (15.6%)	95.3	0.111 (11.4%)	74.0

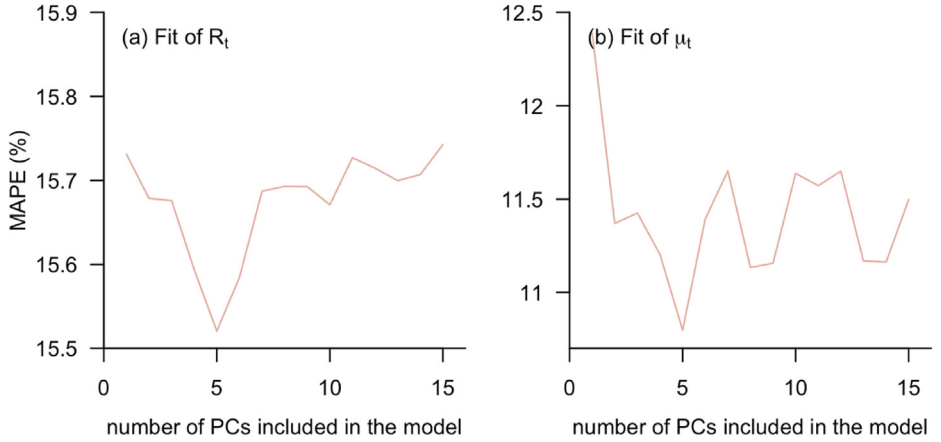

Fig. 2. Mean absolute percentage error (MAPE) for R_t and μ_t estimates. We began with one randomly selected covariates (in this case is the principal components) and afterwards included one more unselected in the model each time until all the covariates were selected.

$$Err_{\mu}(\Delta t) = \frac{1}{366-\Delta t} \sum_{i=1}^{366-\Delta t} MAE_R([i, i-1+\Delta t]), \tag{18}$$

$$Err_R^p(\Delta t) = \frac{1}{366-\Delta t} \sum_{i=1}^{366-\Delta t} MAPE_R([i, i-1+\Delta t]), \tag{19}$$

and

$$Err_{\mu}^p(\Delta t) = \frac{1}{366-\Delta t} \sum_{i=1}^{366-\Delta t} MAPE_{\mu}([i, i-1+\Delta t]). \tag{20}$$

Overall, variation in the mean error of μ_t was much larger than that of R_t, where the former declined over time at a greater rate with an average decrease of 0.9% for each 0.1% change in the latter. For μ_t, the error was highest at 22% when the inference length was 90 days, decreasing to 8% at a 360-day window, while the error for R_t first reduced, reached a minimal value of 18.1% at 180 days, and rose thereafter to an even higher value of 21% (Fig. 3).

To avoid the possible bias introduced by the selection of assessment window, i.e., the reference values to calculate the errors, we further used EpiMix to estimate R_t s and μ_t s by fixing the right end of inference window at day 365 and changing the left end from day 1 to day 276. We calculated MAEs and MAPEs of the estimates over a fixed 90-day interval $[275, 365]$ and found that for different inference window lengths, the MAPE for R_t estimates remained relatively constant at around 30%,

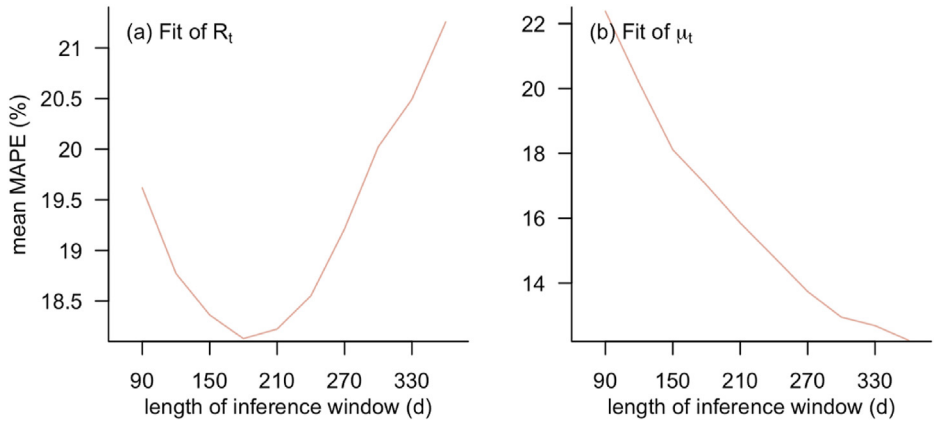

Fig. 3. Fits of R_t and μ_t when the length of inference window differed. Y-axis denotes the mean of mean absolute percentage errors (MAPE) of the estimates.

while there was a distinguishable increasing trend in MAPE for μ_t from 8% to 21% as the starting time for estimation was postponed and the inference window was shortened (Fig. S7).

3.6. Effects of reporting rates

Though we set them to be constant in previous simulations, reporting rates are likely to change throughout the inference window. To explore the effect of such changes on estimations, we set the reporting rate for the first day as baseline and considered the relative reporting rate, i.e.,

$$r_t = \frac{r_t^{real}}{r_1^{real}} \tag{21}$$

We considered scenarios with continuously increasing, decreasing, or fluctuating reporting rates by designing different functional forms that characterized a diverse set of behaviors. Specifically, if the change in r_t is monotonic (i.e., increasing or decreasing), for $t \geq t_1$, we let

$$r_t = \exp\left[a(\min(t, t_2) - t_1 + 1)^b\right], \tag{22}$$

where a, b are parameters to characterize the changes, t_1 and t_2 are the days when the rate begins and stops to change respectively. To be more specific, $b = 1$ if the change is small and $b = 1.35$ if it is large; $a > 0$ if the trend is increasing while if decreasing, $a < 0$; $|a| = 0.001$ if the change is gradual, $|a| = 0.1$ if the change is small and sharp and $|a| = 0.2$ if the change is large and sharp. For the case when r_t is fluctuating from day t_1 to t_2, however, we considered the model

$$r_t = \exp\{0.2\, sin[0.05(\min(t, t_2) - t_1 + 1)]\}. \tag{23}$$

For each scenario, we let t_1 be either one or some values in $\{1, 100, 200\}$ and t_2 in $\{203, 207, 299, 365\}$ (Fig. S8). The observed case count on day t was rounded to the nearest integer after calculating it as

$$I_t^{observed} = I_t \times r_t \tag{24}$$

Using the same set of R_t s and μ_t s as previous subsections, we simulated another incidence curve with a maximum daily case count exceeding 12,000 (Fig. S9). We performed estimations with EpiMix for the diverse scenarios (Fig. S8) and observed increasing errors, but the fits of both R_t and μ_t were not significantly worse if the inference window was chosen to be a year, from the first day to day 365 (Table 3).

Table 3
Fits of R_t and μ_t in terms of mean absolute error (MAE), mean absolute percentage error (MAPE) and successful coverage rate (SCR) when reporting rates were changed in different ways. The assessment window was from day 1 to day 365 and posterior means were used as point estimates.

Changes in reporting rate		Fit of $\boldsymbol{R_t}$: MAE (MAPE)	Fit of $\boldsymbol{R_t}$: SCR(%)	Fit of $\boldsymbol{\mu_t}$: MAE (MAPE)	Fit of $\boldsymbol{\mu_t}$: SCR (%)
Increase	**Small & gradual from D1**	0.183 (20.9%)	94.5	0.110 (11.8%)	97.0
	Small & gradual from D100	0.184 (21.1%)	95.9	0.110 (11.8%)	97.0
	Large & gradual from D1	0.200 (22.4%)	73.2	0.105 (11.4%)	99.2
	Large & gradual from D100	0.192 (21.7%)	86.6	0.106 (11.4%)	98.4
	Small & gradual from D100 to D299	0.182 (21.0%)	96.4	0.111 (11.9%)	96.7
	Small & gradual from D200 to D299	0.182 (21.0%)	96.4	0.111 (11.9%)	96.7
	Large & gradual from D100 to D299	0.190 (21.4%)	88.2	0.105 (11.3%)	98.4
	Large & gradual from D200 to D299	0.186 (21.9%)	99.2	0.114 (12.3%)	95.3
	Small & sharp from D200 to D203	0.188 (21.3%)	94.5	0.107 (11.5%)	97.8
	Large & sharp from D200 to D207	0.207 (22.9%)	82.2	0.098 (10.6%)	99.2
Decrease	**Small & gradual from D1**	0.183 (21.6%)	98.6	0.110 (11.9%)	96.7
	Small & gradual from D100	0.182 (21.3%)	97.8	0.111 (11.9%)	96.2
	Large & gradual from D1	0.235 (27.8%)	99.5	0.131 (14.0%)	93.7
	Large & gradual from D100	0.201 (23.8%)	99.2	0.118 (12.7%)	94.5
	Small & gradual from D100 to D299	0.183 (21.3%)	97.8	0.111 (12.0%)	96.2
	Small & gradual from D200 to D299	0.181 (21.2%)	97.8	0.111 (12.0%)	96.2
	Large & gradual from D100 to D299	0.202 (24.2%)	98.9	0.118 (12.7%)	94.8
	Large & gradual from D200 to D299	0.186 (21.9%)	99.2	0.114 (12.3%)	95.3
	Small & sharp from D200 to D203	0.185 (21.8%)	98.1	0.118 (12.7%)	93.7
	Large & sharp from D200 to D207	0.226 (26.7%)	96.7	0.151 (16.2%)	86.0
Fluctuation	**Biweekly cycle from D1**	0.221 (24.5%)	89.9	0.114 (12.3%)	94.8
	Monthly cycle from D1	0.200 (22.5%)	93.7	0.121 (13.1%)	92.3
	Quarterly cycle from D1	0.188 (21.5%)	96.7	0.114 (12.3%)	93.7
	Monthly cycle from D100	0.201 (22.5%)	93.7	0.108 (11.6%)	96.2

Greater errors were observed with longer trends in reporting rate changes. For example, when reporting rates from day 1–365 were gradually reduced by 94%, we observed that the MAE for R_t was significantly higher at 0.24, changing to 0.20 when the reporting rates started to change at day 100 which was set at an 85% decrease in reporting rates. Should reporting rates remain fixed before day 100 and after day 299, resulting in a 72% decrease in reporting rates, the error was comparably 0.20 while a later start of reporting rate to decrease from day 200, which led to a 40% decline by day 299, caused a smaller MAE of 0.19. When the reductions in reporting rates for the four situations aforementioned were smaller at 30%, 23%, 18% and 10% respectively, the MAE for R_t estimates reduced to approximately 0.18. Furthermore, a sharp decrease from day 200 to day 203, resulting in a decrease of 33% for reporting rates, had a similar MAE of 0.19 to that of gradual change observations.

For periodically fluctuating reporting rates where the range of fluctuation is 0.82–1.22 times of the original value, the effect of an early change did not necessarily result in greater errors in R_t estimates, though for μ_t, a later start by 99 days brought down the error by 0.1 and increased SCR by 4%. An increase in the length of a cycle, such as from a biweekly to monthly length, resulted in reduced MAE by 0.02 and increased SCR by 4% but no such patterns were observed for μ_t (Table 3).

We further explored how non-constant reporting rates affected estimation errors with the extension of inference window. We focused on the scenario of sharp decreasing in reporting rates from day 200 to day 207, when we observed the large estimation errors for both R_t and μ_t. We fixed the left end of the inference window to be day 1 and gradually extended the right end from day 200 to day 365. We performed estimations for this scenario and the scenario with constant reporting rates respectively and compared the estimation errors between these two scenarios for different time points (Fig. 4, Fig. S10). For both R_t and μ_t estimates, the greatest difference took place in the first few days after the start of the change. For R_t estimates, the difference first increased dramatically, peaking at 2.4% when the inference end day was 216 while for μ_t estimates, a maximum gap of 1.5% was observed on the 6th day.

3.7. Performance of EpiMix in a low-incidence scenario

To test EpiMix's performance in low-incidence scenario, we simulated another set of case counts by starting with a smaller initial case count value $I_0 = 20$ using the same R_t and μ_t as the previous subsections (Table 4, Fig. S11). In this time series, the majority (92%) of daily case counts had no more than 50 cases and the maximum case count was 158. We estimated R_t s and μ_t s with EpiMix and compared the results with those by EpiFilter (Fig. S12), finding that the mean absolute error of point estimates was 0.2 (23% of the true value) for R_t by EpiMix and 0.12 (12% of the true value) for μ_t. By contrast, errors by EpiFilter were on average 0.42 (52% of the true value) and 0.23 (24% of the true value) respectively. In addition, the successful coverage rates for both R_t and μ_t estimates by EpiMix exceeded 95% but those for EpiFilter were 63% and 36% respectively.

4. Results

To demonstrate the method's applicability in reality, we applied EpiMix to estimate the time-varying reproduction number, R_t, in two Western Pacific countries, namely Singapore and New Zealand. Both countries had relatively low case counts for an extended period in 2021, making estimation of R_t more challenging. In the main analysis, we employed the full model with the longest possible inference window to estimate the R_ts, but our sensitivity analyses showed minimal changes would have been brought by different models (Tables S6–S8).

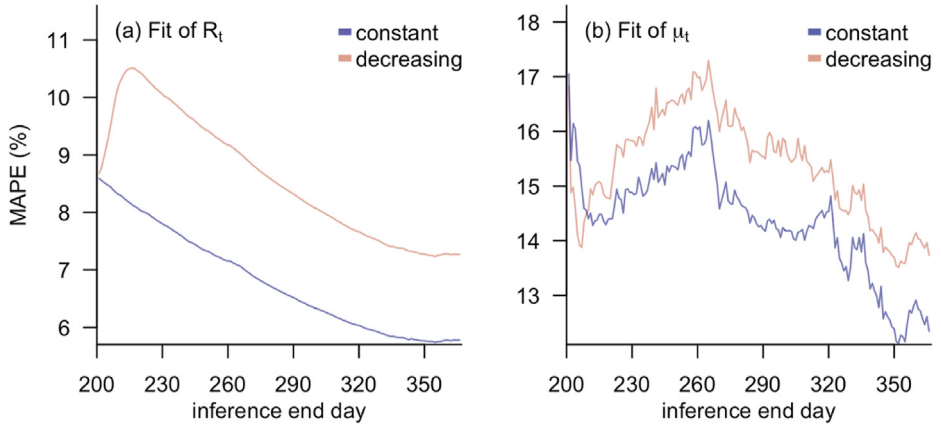

Fig. 4. Comparison of mean absolute percentage errors (MAPE) in the estimated R_t s and μ_t s when inference start time was fixed as day 1 and the end time was gradually postponed from day 200 to day 365, between the cases when reporting rate was constant and when there was a sharp, large increase in reporting rate from day 200–207.

Table 4
Fits of R_t and μ_t in terms of mean absolute error (MAE), mean absolute percentage error (MAPE) and successful coverage rate (SCR) when different methods (namely, EpiMix and EpiFilter) were used for estimation in the low-incidence scenario. Since EpiFilter does not provide μ_t estimates, we instead used R_t estimates to compare with the observed μ_t s. The assessment window was from day 1 to day 365 and posterior means were used as point estimates.

Method	Fit of $\boldsymbol{R_t}$: MAE (MAPE)	Fit of $\boldsymbol{R_t}$: SCR(%)	Fit of $\boldsymbol{\mu_t}$: MAE (MAPE)	Fit of $\boldsymbol{\mu_t}$: SCR (%)
EpiMix	0.200 (23.4%)	96.7	0.115 (12.3%)	97.3
EpiFilter	0.422 (51.9%)	35.9	0.226 (24.1%)	63.0

4.1. COVID-19 in Singapore

As reporting rates were likely to change with the introduction of the Home Recovery Programme (HRP) in October 2021, we considered two time periods: a 9-month period from January 1 to September 30, 2021, and a one-year period that ended on December 31, 2021, to check whether EpiMix was able to reasonably reproduce case counts when differences in reporting rates exist.

Estimates were similar for both time periods especially when the case counts were greater than 10 (Fig. 5). The mean absolute difference between the expected values for the posterior predictive distributions of community case counts and the observed values was 1.60 when the inference window was one year. This reduced to 1.13 when the inference window was the first nine months of the year. In the latter case, the R_t point estimates for Singapore have a mean of 1 (IQR: 0.84–1.35) and peaked on July 14, 2021 at 7.95. When changing the inference window to a year, we estimated that R_t peaked on the same day though at a slightly lower level of 7.08. For the period between mid-March and mid-April 2021, however, disparities were

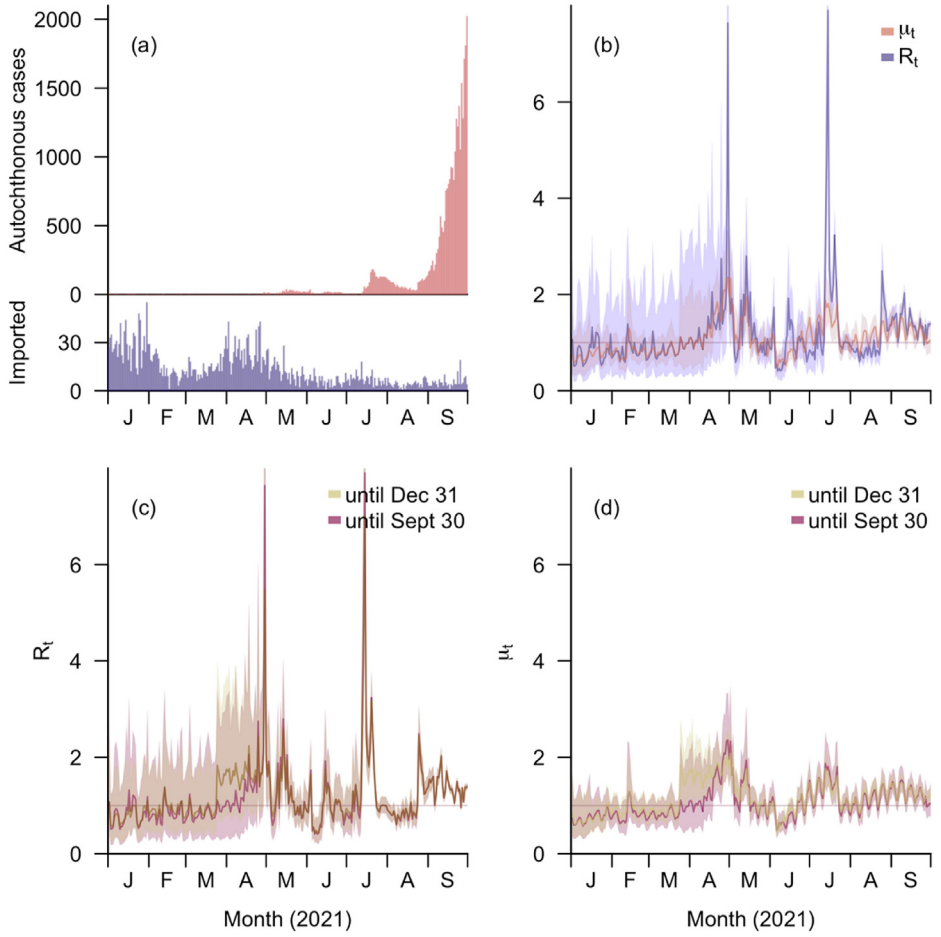

Fig. 5. (a) Imported and autochthonous case counts in Singapore from January 1 to September 30, 2021 (before the start of HRP); (b) The corresponding R_t and μ_t estimates by EpiMix for that time interval; (c) Comparison between R_t estimates by EpiMix using on two different inference windows—one from January 1 to September 30, 2021 and the other from January 1 to December 31, 2021; (d) Comparison between μ_t estimates by EpiMix using the two different inference windows aforementioned.

observed in both R_t and μ_t estimates, but overall, the mean absolute differences between R_t and μ_t point estimates were 0.11 and 0.12 respectively.

The sensitivity analysis on the impact of imported cases showed that fits of both R_t and μ_t over the time interval January to September 2021 were similar when the infectiousness of imported cases was introduced to the model as a contributing factor of predicted case counts additional to R_t and local infectiousness (Fig. S14). The 95% CrIs overlapped for 66% of the period where the differences in the upper bound for the 95% CrIs of R_t were smaller than 0.1 for 66% of the days while the percentage for the corresponding lower bound was 87%; for μ_t, the differences were 57% and 73% respectively. The posterior means of R_t and μ_t from May 2021 were almost identical with a median difference of 0.004 (IQR: 0–0.01) and a maximum of 0.13 in July. Imported cases on average caused both R_t and μ_t estimates to lower by 0.25 from mid-February to the end of March when there were almost no autochthonous cases and the total case counts were dominated by imported cases.

4.2. COVID-19 in New Zealand

In New Zealand, no more than 230 cases were reported on any individual day in 2021 (Fig. 6). When there were no reported cases for 50 consecutive days, the infection potential became 0 and the Poisson model would fail due to the 0 mean. Therefore, we started estimating from August 19, 2021, 2 days after the first COVID-19 case since April was reported.

Model performance was fair with an average absolute difference of 2.36 between actual case counts and the expectation from the posterior predictive distributions. The R_t point estimates decreased rapidly from 4 at the beginning of the wave (August 19) to 1 by the end of August 2021, while the range of the 95% CrIs decreased simultaneously from 4 to 0.4. The μ_t point estimates dropped from 3 on August 19 to 0.6 on September 1, 2021 and the range for 95% CrIs also fell from 2 to 0.3 over the same period of time. From September to December, the estimates fluctuated around 1 with a standard deviation of 0.17 for μ_t and 0.31 for R_t. The range for μ_t point estimates (0.55–1.28) was also smaller than that for R_t (0.44–2.53) but their interquartile ranges were relatively similar (μ_t: 0.86–1.10; R_t: 0.81–1.17).

5. Discussion

In this paper, we developed a novel approach, termed EpiMix, which provides robust estimates of the time-varying effective reproduction number, R_t. Based on the Poisson transmission model specified in Cori et al.'s work (Cori et al., 2013), we added pooled information of exogeneous factors, denoted as the $\boldsymbol{X}$ covariate matrix, in the estimation process to reduce the uncertainty of the R_t estimates, particularly when case data are too scarce to be the only source of information. This is achieved by the generalized linear regression utilized in the model, in which the logarithm of R_t is constrained by its mean, log μ_t, a projection of the vector $\boldsymbol{X}_t$., the covariates' values at time step t. Meanwhile, a longer inference window means more information is accumulated, which in any case would be advised to facilitate better estimation. Furthermore, flat, uninformative priors are used in this Bayesian framework to ensure the estimates are data-driven and hence potentially to better recover the transmission potential than currently popular approaches (Parag, 2021). A simulation study (Table 1) was conducted to validate the results.

Variations in the $\boldsymbol{X}$ covariate values in the model are directly captured by the changes in μ_t estimates, which are deterministic exponential functions of the linear combinations of the covariates' effects. This approach shares similarities with EpiRegress and the method proposed by Flaxman et al. (Flaxman et al., 2020; Jin et al., 2022). Periodicity was observed in estimates of μ_t in both the simulations and case studies, probably because mobility patterns, an important driver of the linear predictors (Table 1), tended to have a weekly cycle. Compared with μ_t estimates, those of R_t were less stable and more

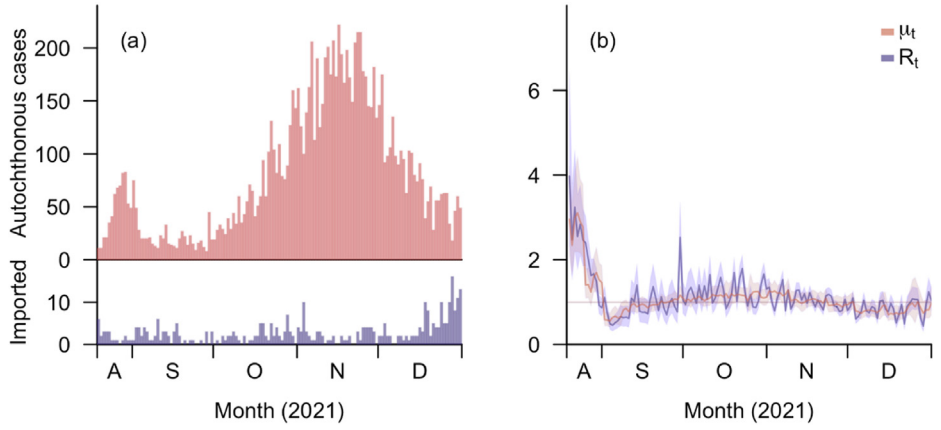

Fig. 6. (a) Imported and autochthonous case counts in New Zealand from August 19 to December 31, 2021; (b) The corresponding R_t and μ_t estimates by EpiMix for that time interval.

sensitive to changes in the daily case counts, as the innovative incorporation of the random noises ε_t s in EpiMix allowed them to better approximate the true values. The differences between the logarithms of R_t and μ_t estimates, ε_t s, however, could be interpreted as the effects wrought by changes in unobserved factors, such as the susceptible population (Cori et al., 2013).

Another significant advantage of EpiMix is its computational efficiency, providing rapid estimations of R_t s for one year in less than 5 s in the previous simulation studies. This is attributed to the INLA approach, which approximates the posterior marginal distributions of the hyperparameters (Blangiardo et al., 2013) instead of doing posterior sampling which more conventional Markov chain Monte Carlo (McMC) methods (Gelman et al., 2021) utilize. EpiMix therefore circumvents common issues in the McMC estimation process such as slow convergence and poor mixing (Paul et al., 2010). INLA can also produce deterministic estimation results (Beguin et al., 2012), enabling the reproducibility of the R_t estimates by EpiMix if the same datasets and parameters are used. More importantly, these merits are not at the expense of accuracy. Estimations results by INLA and McMC have proven to be alike (De Smedt et al., 2015; Rue et al., 2017).

Furthermore, as is shown in simulation results, EpiMix allows for the flexible selection of covariates for estimation. Since effects of individual covariates might not be identifiable due to the strong correlations, the covariate matrix in the model only serves as additional information that reduces the uncertainty caused by the scarcity of case information. The introduction of the noise item reduces the possibility that the R_t estimates are driven by extreme linear combinations of the entries in the matrix, allowing them to be more flexible and adaptable to the frequently complex epidemiological reality.

The simulation results (Tables 1–2, SI Table S3–S4) also indicate that the performance of the models (in terms of errors and successful coverage rates of the estimates) was not sensitive to the specific variables included. The reduced model in which we excluded the health-related measures, for example, had the same error (0.18) of posterior means for R_t estimates as the full model, while the error for μ_t point estimates slightly increased by 0.005. The successful coverage rates of the 95% CrIs were also over 95% for both μ_t and R_t. Moreover, for the mis-specified model where additional weather-related variables were included compared to the reduced model, the errors for the point estimates of both μ_t and R_t were even slightly smaller than those for either the full model or the reduced model.

This robustness is partly because of the mutual dependence between different covariates, since variations in the estimates are largely explained by those of the covariates. In the cases where too few covariates were included, however, the errors were relatively larger and the coverage rates of the estimated 95% credible intervals decreased significantly, especially for those of μ_t s, which are more sensitive to the variables considered in the model. The strong correlations, however, make it challenging for EpiMix to assess each individual covariate's impact on transmissibility.

Although the model was built under the assumption of static reporting rates over the inference window (Cori et al., 2013), EpiMix still managed to give reliable estimates when some small variations in the rate were introduced in the simulations (Table 3). Still, EpiMix may fail to produce desirable results if the change is great or sudden, especially when the inference end time is only days after the rate stops to change. The limitation, however, is intrinsic and can hardly be alleviated unless exact changes in reporting rates or the actual case counts are recovered, as Flaxman et al. did by substituting number of deaths (Flaxman et al., 2020) for case counts in the model.

We should be cautious to apply EpiMix if there are no cases reported for many successive time steps, i.e., a period longer than the maximum interval with non-zero infection probability in the serial interval distribution. Though under a regression framework, EpiMix may still provide estimates as long as the case counts are not always zero, the extreme values and large credible intervals make the results unreliable. This problem is inevitable when a Poisson relationship is assumed between case counts and the effective reproduction numbers, since if the number of consecutive time steps with no cases before a time step t is long enough to make the infectiousness zero, the probability for no cases on day t is always 1 for any arbitrary value of R_t while that for a positive case count can never be positive unless the corresponding R_t equals to infinity. In this case, many other prevailing methods, including EpiFilter and EpiEstim, are not applicable either and we need to seek for a new estimation approach.

Despite the limitations mentioned above, EpiMix, we believe, remains a powerful tool for estimating a pathogen's transmission potential and predicting its spread in community. The hybrid approach, which creatively combines regression with the principles of the widely-used EpiEstim method, yields precise estimates of the instantaneous reproduction number straight-forwardly and quickly with little additional assumptions. It is simple to comprehend and ready to use. More importantly, apart from COVID-19, EpiMix can be adopted to investigate outbreaks of any other infectious diseases if elements required for estimation, including case data, serial interval distribution and time series of factors that potentially influence the epidemic curve, are accessible.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was suppoted by Singapore's Ministry of Education (through a Tier 1 grant), the National University of Singapore (through a Reimagine Research grant), and the Singapore Ministry of Health's National Medical Research Council

under its National Epidemic Preparedness and Response R&D Funding Initiative (MOH-001041) Programme for Research in Epidemic Preparedness And REsponse (PREPARE).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.idm.2023.06.002.

References

Allen, T., Murray, K. A., Zambrana-Torrelio, C., Morse, S. S., Rondinini, C., Di Marco, M., Breit, N., Olival, K. J., & Daszak, P. (2017). Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications, 8*(1), Article 1. https://doi.org/10.1038/s41467-017-00923-8

Beguin, J., Martino, S., Rue, H., & Cumming, S. G. (2012). Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. *Methods in Ecology and Evolution, 3*(5), 921–929. https://doi.org/10.1111/j.2041-210X.2012.00211.x

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology, 4*, 33–49. https://doi.org/10.1016/j.sste.2012.12.001

Bureau of Meteorology Australia. (2021). Climate data online. http://www.bom.gov.au/climate/data/.

Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the Total Environment, 728*(138882). https://doi.org/10.1016/j.scitotenv.2020.138882

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

De Smedt, T., Simons, K., Van Nieuwenhuyse, A., & Molenberghs, G. (2015). Comparing MCMC and INLA for disease mapping with Bayesian hierarchical models. *Archives of Public Health, 73*(Suppl 1), O2. https://doi.org/10.1186/2049-3258-73-S1-O2

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research, 2*(1), 23–41. https://doi.org/10.1177/096228029300200103

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature, 584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2021). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Google. (2020). COVID-19 community mobility report. COVID-19 community mobility report. https://www.google.com/covid19/mobility?hl=en.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour, 5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Jiang, F., Deng, L., Zhang, L., Cai, Y., Cheung, C. W., & Xia, Z. (2020). Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *Journal of General Internal Medicine, 35*(5), 1545–1549. https://doi.org/10.1007/s11606-020-05762-w

Jin, S., Dickens, B. L., Lim, J. T., & Cook, A. R. (2022). EpiRegress: A method to estimate and predict the time-varying effective reproduction number. *Viruses, 14*(7), Article 7. https://doi.org/10.3390/v14071576

Kano, T., Yasui, K., Mikami, T., Asally, M., & Ishiguro, A. (2021). An agent-based model of the interrelation between the COVID-19 outbreak and economic activities. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 477*(2245), Article 20200604. https://doi.org/10.1098/rspa.2020.0604

Kupferschmidt, K., & Wadman, M. (2021). Delta variant triggers new phase in the pandemic. *Science, 372*(6549), 1375–1376. https://doi.org/10.1126/science.372.6549.1375

Liu, Y., & Rocklöv, J. (2021). The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *Journal of Travel Medicine, taab124*. https://doi.org/10.1093/jtm/taab124

Macali, A. (2023). Coronavirus cases in NSW - COVID live. https://covidlive.com.au/nsw.

Ministry of Health New Zealand. (2021). COVID-19: News and media updates. https://www.health.govt.nz/covid-19-novel-coronavirus/covid-19-news-and-media-updates.

Ministry of Health Singapore. (2021). Past updates on COVID-19 local situation. https://www.moh.gov.sg/covid-19/past-updates.

Ministry of Manpower Singapore. (2023). Newsroom. https://www.mom.gov.sg/newsroom.

Parag, K. V. (2021). Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology, 17*(9), Article e1009347. https://doi.org/10.1371/journal.pcbi.1009347

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., & Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine, 29*(12), 1325–1339. https://doi.org/10.1002/sim.3858

Pfefferbaum, B., & North, C. S. (2020). Mental health and the covid-19 pandemic. *New England Journal of Medicine, 383*(6), 510–512. https://doi.org/10.1056/NEJMp2008017

Re3data.Org. (2012). re3data.org - registry of research data repositories. *GISAID*. https://www.re3data.org/repository/r3d100010126.

Redfern, A. (2021). Live COVID-19 vaccination tracker. Covidvax.Live. http://covidvax.live/.

Rue, H., Lindgren, F., Niekerk, J.van, Krainski, E., & Fattah, E. A. (2023). R-INLA project. https://www.r-inla.org/.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B, 71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application, 4*(1), 395–421. https://doi.org/10.1146/annurev-statistics-060116-054045

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji, U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., & de Oliveira, T. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*, 1–10. https://doi.org/10.1038/s41586-022-04411-y

Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology, 160*(6), 509–516. https://doi.org/10.1093/aje/kwh255

WHO. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19%5f-11-march-2020.