Robust uncertainty quantification in popular estimators of the instantaneous reproduction number

Nicholas Steyn*[1] and Dr Kris V Parag[2]

[1]Department of Statistics, University of Oxford, Oxford, United Kingdom
[2]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

Abstract

The instantaneous reproduction number (R_t) is a widely used measure of the rate of spread of an infectious disease. Correct quantification of the uncertainty of R_t estimates is crucial for making well-informed decisions. Popular methods for estimating R_t leverage smoothing techniques to distinguish signal from noise. Examples include EpiEstim and EpiFilter, each are controlled by a single "smoothing parameter", which is traditionally chosen by the user. We demonstrate that the values of these smoothing parameters are unknown and vary markedly with epidemic dynamics. We argue that data-driven smoothing choices are crucial for accurately representing uncertainty about R_t estimates. We derive model likelihoods for the smoothing parameters in both EpiEstim and EpiFilter. Adopting a flexible Bayesian framework, we use these likelihoods to automatically marginalise out the relevant smoothing parameters from these models when fitting to incidence time-series. Applying our methods, we find that the default parameterisations of these models can negatively impact inferences of R_t, delaying detection of epidemic growth, and misrepresenting uncertainty (typically by producing overconfident estimates), with substantial implications for public health decision-making. Our extensions mitigate these issues, provide a principled approach to uncertainty quantification, and improve the robustness of inference of R_t in real-time.

The instantaneous reproduction number R_t, defined as the average number of secondary infections generated per effective primary case at time t, is a popular measure of epidemic spread [1]. A value of $R_t < 1$ indicates a declining epidemic while a value of $R_t > 1$ indicates a growing epidemic. This quantity is particularly useful for policymakers as it gives the proportional change in transmission rates required to control the epidemic, thus informing decisions to trigger and release public health interventions [2, 3, 4, 5, 6]. As a stark example, in June 2020, an R_t estimate of 1.01 was used to justify continued school closures in Greater Manchester, England [7, 8]. In addition to real-time decision-making, estimates of R_t are also used for forecasting, scenario analysis, and understanding the impact of interventions [9].

Many models exist to estimate R_t from such data, including mechanistic models such as compartmental SIR models, and semi-mechanistic models such as the renewal-model [10], the latter of which underlies most contemporary methods for the real-time estimation of R_t, including EpiEstim [11] and EpiFilter [12]. We focus on quantifying uncertainty in these two models, although our approach generalises to any real-time model.

Correct quantification of the uncertainty of R_t is crucial for making well-informed decisions. If one reports a 95% credible interval for R_t, then it is expected that the true value of R_t should fall within this interval 95% of the time. Undercoverage occurs when the true value falls within the credible interval less often than expected, leading to over-confident and biased decision-making, while overcoverage occurs when the true value falls within the credible interval more often than expected, leading to under-confident and highly uncertain decision-making. A model that produces correct coverage is said to be well-calibrated.

*Corresponding author: nicholas.steyn@univ.ox.ac.uk

Despite the importance of uncertainty quantification, calibration is often neglected in epidemiological models. For example, during the COVID-19 pandemic, SPI-M in England used estimates from multiple groups to produce consensus estimates of R_t. They were unable to combine the estimates using standard techniques due to models "*providing estimates with lower levels of uncertainty that are not fully accounting for inherent uncertainties*" [13]. Even when correct coverage is explicitly targeted, epidemic models frequently fail to achieve it. For instance, nearly all models submitted to the *open challenge to advance probabilistic forecasting for dengue epidemics* [14] produced over-confident predictions for the various forecasting targets. A simple baseline model (which was included for comparison) demonstrated superior calibration compared to all 16 submitted models when predicting the peak week, for example.

In particular, smoothing assumptions (such as penalized likelihoods [15], piecewise constant/trailing window models [16, 17, 11, 18], and latent-space models [12, 19, 20, 21]) are a key source of model miscalibration. Oversmoothed estimates result in delayed and overconfident estimates, while undersmoothed estimates are noisy and lack precision. Even with perfect case reporting (i.e. no observation noise), inherent stochasticity in the transmission of infectious diseases necessitates the use of smoothing to distinguish signal from noise. All popular estimators of R_t employ some type of smoothing.

Despite the importance and ubiquity of these assumptions, the philosophical and practical treatment of smoothing parameters varies by method. Some methods treat these parameters as unknown quantities to be estimated alongside R_t [19, 17, 22], while other methods treat the choice of smoothing parameter(s) as a model selection problem and seek to find some optimal point-value of the parameter [23]. Some methods allow the user to choose their own values or provide heuristic default values [11, 15, 22].

We argue that, because the true dynamics of R_t are always unknown and depend on both the pathogen and the population in which the pathogen is spreading, uncertainty about the nature of these dynamics should not be ignored. This uncertainty factors in both the choice of the dynamic model itself (structural uncertainty), and the parameters associated with the chosen dynamic model (parametric uncertainty). We focus on parametric uncertainty in this paper, which on its own can cause substantial model miscalibration, and demonstrate that correct marginalisation of smoothing parameters generally improves model robustness, even if the structure of the dynamic model does not accurately reflect reality.

In this paper, we derive novel likelihoods for the smoothing parameters in two popular models: EpiEstim [11] and EpiFilter [12]. We then use these likelihoods to marginalise out the smoothing parameters, presenting estimates of R_t that appropriately account for uncertainty in these parameters. We also derive predictive posterior distributions and demonstrate their use in model comparison via the continuous ranked probability score (CRPS) [24]. We validate our methods on both simulated data, where model estimates can be compared to a known truth, and real-world data, where we consider the practical implications of our methods on decision-making during the COVID-19 pandemic in New Zealand.

Methods

We provide a high-level overview of our methods here, with full mathematical derivations and technical details provided in supplementary section 1.

Background

Both EpiEstim and EpiFilter leverage the *Poisson renewal model* for R_t estimation. Specifically, let C_s be the number of cases reported at time s and w_u be the probability that a secondary case is reported u days after the primary case (often approximated by the serial interval), the *total infectiousness* is defined as $\Lambda_t = \sum_{u=1}^{t} C_{t-u} w_u$. Given Λ_t and the current value of R_t, the number of cases at time t is then assumed to be distributed according to:

$$C_t \sim \text{Poisson}(R_t \Lambda_t) \tag{1}$$

EpiEstim assumes that, on each day t, R_t has been fixed for the preceding k days (a trailing window of length k). Larger values of k imply that R_t has been fixed for a longer period, and thus result in smoother estimates. The likelihood of observing cases between time-steps $t-k+1$ and t (denoted $C_{t-k+1:t}$) is the product of the daily Poisson likelihoods (equation 1) in the trailing window. A conjugate Gamma(α, β) prior distribution is assumed for R_t, resulting in a Gamma$(\alpha_{t,k}, \beta_{t,k})$ posterior distribution for R_t given $C_{1:t-1}$, where the shape-parameter $\alpha_{t,k}$ and rate-parameter $\beta_{t,k}$ are given by:

$$\alpha_{t,k} = \alpha + \sum_{s=t-k+1}^{t} C_s, \quad \beta_{t,k} = \beta + \sum_{s=t-k+1}^{t} \Lambda_s \quad (2)$$

Instead of assuming fixed trailing windows, EpiFilter assumes that R_t follows a Gaussian random walk with standard deviation equal to $\eta\sqrt{R_{t-1}}$. Larger values of η allow R_t to vary faster, resulting in less smooth estimates. A grid-approximation to the exact Bayesian filtering equations is used to derive the posterior distribution of R_t given $C_{1:t}$ (i.e. reported case data up-to time t). While EpiFilter also allows for the estimation of the smoothing distribution (R_t given past and future data), we focus on real-time estimation, so do not consider this here.

Model likelihoods

We use the same general framework to derive likelihoods for the smoothing parameters of both methods. Formal derivations for each model are included in supplementary section 1. Letting θ denote an arbitrary smoothing parameter, we begin with the predictive decomposition of the likelihood:

$$\log P(C_{1:T}|\theta) = \sum_{t=1}^{T} \log P(C_t|C_{1:t-1}, \theta) \quad (3)$$

and note that the one-step-ahead-likelihood can be written as:

$$P(C_t|C_{1:t-1}, \theta) = \int P(C_t|R_t, C_{1:t-1}) P(R_t|C_{1:t-1}, \theta)\, dR_t \quad (4)$$

where $P(C_t|R_t, C_{1:t-1})$ is the renewal model (equation 1). Thus, deriving model likelihoods relies on the derivation of the one-step-ahead predictive distribution for R_t: $P(R_t|C_{1:t-1}, \theta)$.

For EpiEstim, R_t depends on reported cases only on days $t-k+1$ to t, however the predictive distribution explicitly ignores data on day t, so EpiEstim's predictive distribution for R_t is Gamma-distributed with shape $\alpha_{t-1,k-1}$ and rate $\beta_{t-1,k-1}$. In this case, equation 4 is a Gamma-Poisson mixture, hence $C_t|C_{1:t-1}$ follows a negative binomial distribution with parameters $r = \alpha_{t-1,k-1}$ and $p = \frac{\beta_{t-1,k-1}}{\beta_{t-1,k-1}+\Lambda_t}$. The likelihood of EpiEstim's k is simply the sum of log-negative binomial probability mass functions for each day t.

For EpiFilter, the predictive distribution of R_t is a by-product of the Bayesian filtering equations, found by propagating the estimated distribution of R_{t-1} given $C_{1:t-1}$ forward according to the assumed Gaussian random walk. This allows us to find the one-step-ahead likelihood of C_t by taking a weighted average of the Poisson likelihood of C_t with respect to the predictive distribution of R_t given $C_{1:t-1}$.

We provide full implementations of these methods and the corresponding likelihoods in the github repository.

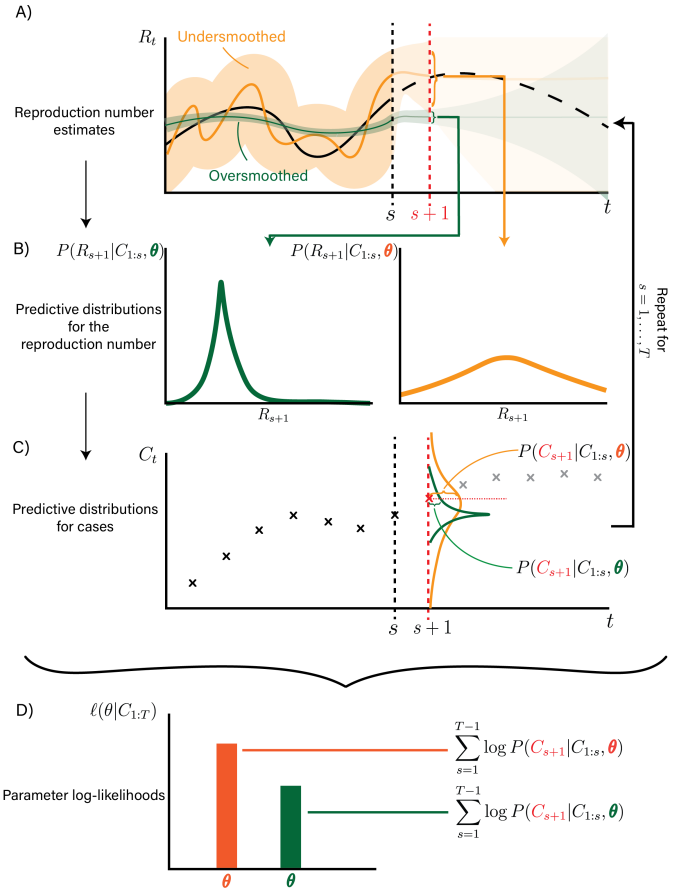

Figure 1: Schematic demonstrating the derivation of parameter likelihoods using the predictive decomposition (equation 3). R_t estimates are projected forward one time-step according to the relevant dynamic model (panel A), giving predictive distributions for R_{s+1} (panel B). These are combined with the renewal model to give the probability of observing C_{s+1} conditional on past data $C_{1:s}$ and the chosen smoothing parameter θ (panel C). In this time-step, the undersmoothed model is more likely than the oversmoothed model. Repeating this process and summing the log-predictive probabilities for all time-steps produces the model likelihood (Panel D). These log-likelihoods are later used to marginalise out uncertainty about θ, allowing for robust reporting of uncertainty about R_t. The form of the R_t estimates and predictive distributions for R_t depend on the specific model.

Posterior distributions

These methods admit log-likelihoods for k and η given $C_{1:t}$, denoted $\ell(k|C_{1:t})$ and $\ell(\eta|C_{1:t})$ respectively. We use these to derive posterior distributions (denoted $P(k|C_{1:t})$ and $P(\eta|C_{1:t})$), typically using a uniform prior distribution over $k \in \{1, 2, \ldots, 30\}$ (assuming a maximum monthly dependence) and $\eta \in [0, 1]$ (assuming a maximum Poisson diffusion).

We are ultimately interested in estimates of R_t that account for uncertainty about k or η. To achieve this, we marginalise out these parameters from the posterior distribution of R_t, a procedure that is rare in the literature. For EpiEstim, we leverage the discrete nature of k to write exactly:

$$P(R_t|C_{1:t}) = \sum_{k=1}^{30} P(R_t|C_{1:t}, k)P(k|C_{1:t}) \tag{5}$$

whereas for EpiFilter, we use a grid-approximation ($\eta \in \mathcal{E}$, supplementary section 1.1.2) to the desired integral:

$$\begin{aligned} P(R_t|C_{1:t}) &= \int P(R_t|C_{1:t}, \eta)P(\eta|C_{1:t})\ d\eta \\ &\approx \sum_{\eta\in\mathcal{E}} P(R_t|C_{1:t}, \eta)P(\eta|C_{1:t}) \end{aligned} \tag{6}$$

We can also marginalise out the smoothing parameter from the predictive distributions for C_t (equation 4). This allows us to present marginal one-step-ahead predictive distributions for C_t under both models, useful for model comparison and probabilistic forecasting.

Model evaluation

We argue that a "good" model is one that maximises precision (i.e. minimises the width of uncertainty intervals), subject to being well-calibrated [24]. Choosing the "best" model thus involves a trade-off: how much (if any) miscoverage are we willing to accept in exchange for more precise estimates? A commonly used metric in these cases is the Continuous Ranked Probability Score (CRPS), which measures the distance between the estimated predictive distribution and the empirical distribution of the data. The smaller this distance is, the more closely the model's predictive uncertainty aligns with the observed data variability. We employ the CRPS to compare the performance of our models. Full details of how we calculate CRPS for each model is provided in supplementary section 1.4.

Data

We test our models on three different simulated datasets, each assuming a different dynamic model for R_t: a Gaussian random walk (matching the dynamic model assumed by EpiFilter), a sinusoidal curve, and a step-change model. These models cover a range of smooth to sharp changes in R_t. We also compare model outputs on real-world data from the COVID-19 pandemic in New Zealand [25], chosen as an example of high-quality data with limited reporting biases. We explicitly relate real-world policy decisions to the inferences made by our models. A common serial interval from the COVID-19 literature, a Gamma distribution with a mean of 6.5 days and standard deviation of 4.2 days, is used for both the simulation study and the application to real-world data [26, 27]. Further information on simulated data is provided in supplementary section 2.

Results

Simulation study

Fitting EpiEstim and EpiFilter to a single realisation of each simulated epidemic (figure 2) demonstrates that default parameterisations of both EpiEstim and EpiFilter result in oversmoothed estimates of R_t, relative to the level of smoothing estimated from the data. The exception is EpiFilter when fit to the random walk simulation, where the true value of η is deliberately chosen to match EpiFilter's default $\eta = 0.1$.

Using these default parameters results in credible intervals for R_t that typically undercover the true value. This is more noticeable for EpiEstim, where coverage of the 95% credible intervals for R_t in the default model ranges from just 8.9% in the sinusoidal simulation, to 74.4% in the step-change simulation. Marginalising out k considerably improves coverage in these models to 81.1% and 94.4% respectively. Default EpiFilter is generally more robust, partially as a result of the default $\eta = 0.1$ being less extreme with respect to the posterior distribution of η, although marginalising out this parameter still improves coverage of R_t from 81.1% to 92.2% in the sinusoidal model.

The one-step-ahead predictive coverage of reported cases is also improved by marginalising out the smoothing parameter. While this is true for all models and simulations considered, the effect is more pronounced in EpiEstim.

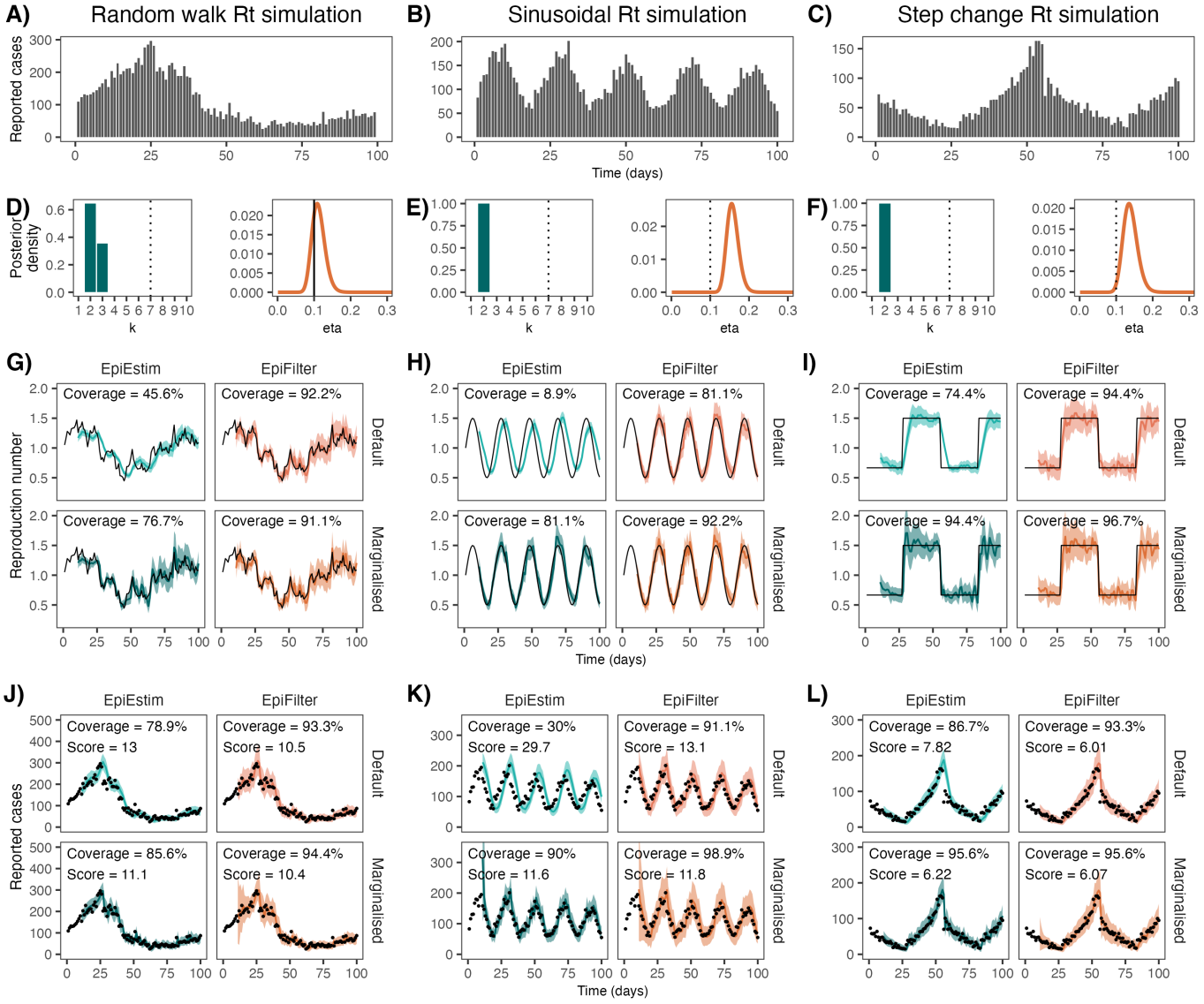

Figure 2: Simulated case data (A, B, C), posterior distributions for smoothing parameters at $t = 100$ (D, E, F), estimates of R_t (G, H, I), and estimates of predictive cases (J, K, L) for one realisation of each simulated epidemic. The first column (A, D, G, J) shows results for the Gaussian random walk simulation with $\eta = 0.1$, a dynamic model that precisely matches default EpiFilter. The second column (B, E, H, K) shows results for the sinusoidal simulation, and the third column (C, F, I, L) shows results for the step-change simulation. Vertical dotted lines in the parameter posterior distributions denote the default parameter values, while the vertical solid line denotes both the default parameter value and the true value of the parameter (in EpiFilter when fit to the random walk simulation). Black lines (in R_t estimates) and black dots (in predictive C_t estimates) denote the true values of R_t and C_t respectively. Predictive coverage of the 95% credible intervals (closer to 95% is better) and the CRPS (lower is better) are shown within each figure.

Finally, marginalising out the smoothing parameter almost always improves (decreases) the CRPS, suggesting that marginalised models produce more accurate predictive distributions of cases than default models. The sole exception is EpiFilter in the step-change simulation, where the CRPS worsens (increases) slightly, although we show in supplementary section 3 that, on average, marginalisation also improves EpiFilter when fit to step-change simulations. The worsening CRPS in this specific simulation suggests that, according to the trade-off implied by the CRPS, the narrower credible intervals produced by default EpiFilter may be worth sacrificing some coverage of predictive cases. We observe a similar effect in the sinusoidal simulation when comparing EpiEstim and EpiFilter, where the CRPS is slightly lower in EpiEstim (indicating it as the better model), despite EpiFilter having better coverage of predictive cases.

These results depend on the simulated data, and we consider differing simulated epidemic dynamics in supplementary section 3, showing how the appropriate level of smoothing depends on the underlying epidemic dynamics. We find that marginalisation becomes more important as the standard deviation of the simulated random walk increases, the period of the sinusoidal curve decreases, or the step-change becomes more frequent, as (in all three examples) the level of oversmoothing in default models increases.

Contrary to common intuition, fitting to a greater number of daily cases does not necessarily improve the quality of inference. Supplementary figure S1 demonstrates that, while marginalised EpiFilter is largely robust to sample size, EpiEstim's coverage worsens as sample size increases in both the default and marginalised models. This occurs as the model, which has a guaranteed level of misspecification (R_t cannot be constant at one value on $[t-k+1, t]$ and then constant at a different value on $[t-k+2, t+2]$) results in the model becoming more confident in the incorrect estimate. This increased confidence in a misspecified model means that, in the presence of observation noise, EpiFilter is also expected to degrade as sample size increases.

The COVID-19 pandemic in New Zealand, August-December 2021

After largely containing the spread of COVID-19, in August 2021 an outbreak of the delta-variant was detected in Auckland, New Zealand, leading to the imposition of strict non-pharmaceutical interventions. The outbreak was characterised by an initial peak of cases in late August, followed by a period of decline, before a second peak in mid-November (figure 3). The reproduction number was repeatedly cited in decision-making by multiple officials, including the Prime Minister and the Director-General of Health [3, 28, 29, 30, 4, 31, 32].

We fit all four models to reported cases (smoothed using a 5-day moving average to decrease reporting noise) between 11 August 2021 and 1 December 2021 (figure 3). Both EpiEstim and EpiFilter exhibit an improved CRPS after marginalisation, and EpiEstim exhibits improved predictive coverage after marginalisation. As R_t is unknown, we are unable to evaluate the calibration of this variable.

As in the simulated epidemics, we find that the default models oversmooth relative to the marginalised models. This is most noticeable for EpiEstim, where almost all posterior mass is on $k = 2$. For EpiFilter, the default value of $\eta = 0.1$ is within the 95% credible interval for the final value of η, so, in this case, default EpiFilter is better calibrated than default EpiEstim, however table 1 still highlights some delays in detection of epidemic growth or decline.

On 31 August, marginalised EpiFilter first detected that R_t may be less than 1, while 4 days later, this was detected by the default EpiEstim, the last model to do so. In fact, both marginalised models were confident that $R_t < 1$ by 3 September, before the possibility of decline had been detected by default EpiEstim. This was a period characterised by daily press conferences and strict interventions (including stay-at-home orders), which were estimated to have cost the city NZ$56m per day [33].

On 13 September, the Government announced that because restrictions had *"...reduced that value [R_t] down to consistently below one"*, restrictions in Auckland would be relaxed the following week on 21 September [30]. There was considerable interest in R_t over the following week, as a resurgence could have necessitated prolonged restrictions. The timing of this announcement coincides exactly with the first detection of a potential resurgence by the marginalised models (on 13 September), while the default models are still confident that $R_t < 1$.

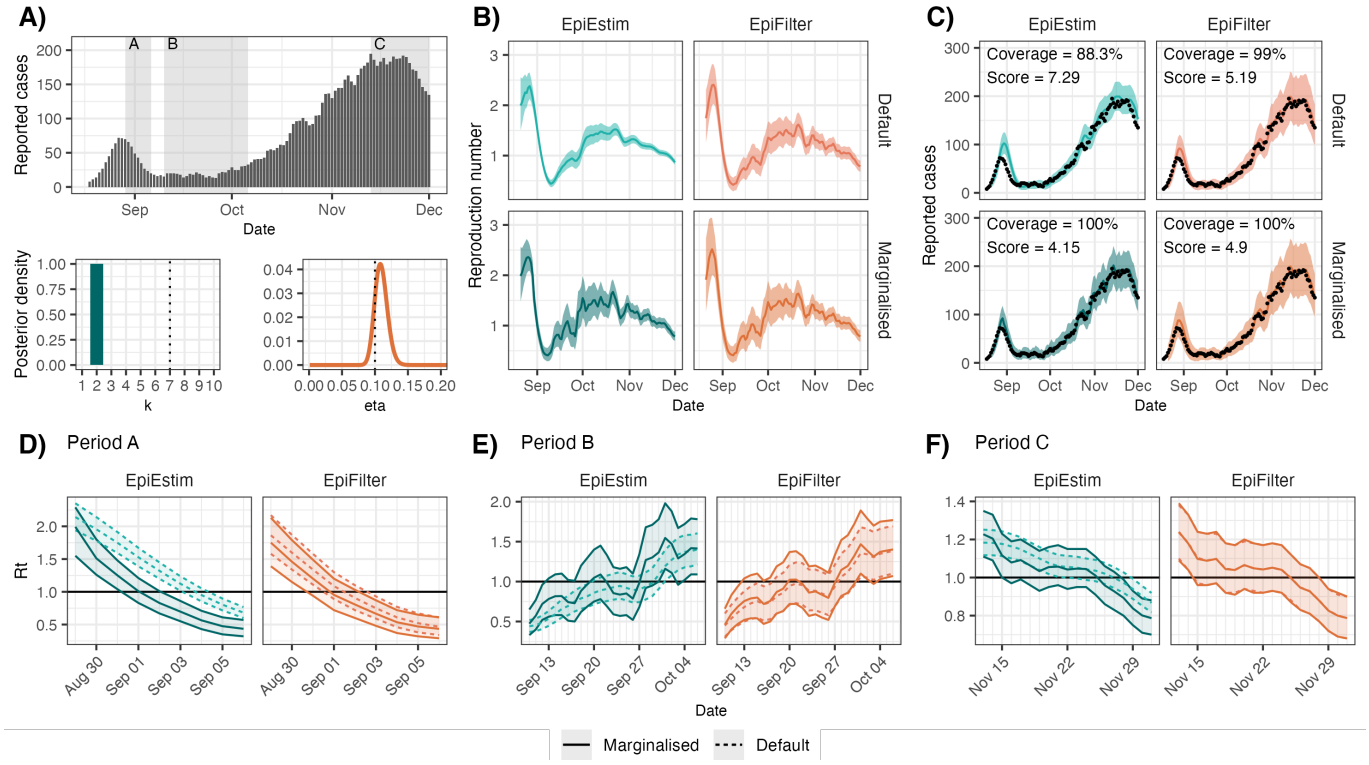

Figure 3: Reported case data and parameter posterior estimates (A), reproduction number estimates (B), predictive cases (C) and reproduction number estimates for selected periods (D, E, F) from fitting the four models to reported case data from the outbreak of the delta-variant of SARS-CoV-2 in Auckland, New Zealand between 11 August 2021 and 1 December 2021. Our methods improve predictive coverage of EpiEstim and the CRPS of both models.

	EpiEstim		EpiFilter	
	Default	Marginalised	Default	Marginalised
Period A				
First detection	4 Sep (+4)	1 Sep (+1)	1 Sep (+1)	31 Aug (+0)
Expected $R_t < 1$	4 Sep (+3)	2 Sep (+1)	2 Sep (+1)	1 Sep (+0)
Confidence in $R_t < 1$	5 Sep (+2)	3 Sep (+0)	3 Sep (+0)	3 Sep (+0)
Period B				
First detection	20 Sep (+7)	13 Sep (+0)	18 Sep (+5)	13 Sep (+0)
Expected $R_t > 1$	30 Sep (+10)	20 Sep (+0)	28 Sep (+8)	20 Sep (+0)
Confidence in $R_t > 1$	1 Oct (+0)	1 Oct (+0)	1 Oct (+0)	1 Oct (+0)
Period C				
First detection	22 Nov (+7)	16 Nov (+1)	15 Nov (+0)	15 Nov (+0)
Expected $R_t < 1$	27 Nov (+2)	26 Nov (+1)	25 Nov (+0)	25 Nov (+0)
Confidence in $R_t < 1$	29 Nov (+1)	28 Nov (+0)	29 Nov (+1)	29 Nov (+1)

Table 1: Dates of the first detection of growth or decline (defined as the upper bound on the 95% credible interval of R_t crossing 1 for growth, or the lower bound crossing 1 for decline), the date that the central estimates of R_t first crossed 1, and the date that the models are first confident in growth or decline (defined as the lower bound on the 95% credible interval of R_t crossing 1 for growth, or the upper bound crossing 1 for decline) for the COVID-19 outbreak in Auckland, New Zealand that started in August 2021. The number of days after the first detection is shown in parentheses, highlighting the substantial delay often observed in default models.

Default EpiFilter first detected a potential resurgence on 18 September (5 days after the marginalised models) and default EpiEstim on 20 September (7 days after the marginalised models, and the day before restrictions were due to be relaxed). While the marginalised models first produced central estimates of $R_t > 1$ on the same day (20 September), it wasn't until 10 days later, on 30 September, that default EpiEstim produced a central estimate of $R_t > 1$, an entire 17 days after the announcement, and 9 days after the restrictions were relaxed.

Daily reported cases continued to increase until mid-November, before appearing to plateau. While the marginalised models (and default EpiFilter) became uncertain about continued epidemic growth around 15 November, default EpiEstim was still confident that $R_t > 1$ until 22 November (7 days later), a clear example of oversmoothed models being overconfident.

We test our methods on additional real-world datasets in supplementary section 4.

Discussion

We have presented derivations for the model likelihood of two popular R_t estimators. We used these likelihoods to develop posterior distributions for the corresponding smoothing parameters, and then marginalised out uncertainty about these parameters from estimates of R_t. We have shown that doing this results in improved uncertainty quantification for R_t, thus allowing increased confidence in the calibration of reported credible intervals.

Robust uncertainty quantification is crucial for real-world decision-making. Our methods provide one of the first principled and computationally efficient ways of ensuring the robustness of reported uncertainty in two popular models: EpiEstim [11] and EpiFilter [12].

In addition to real-time decision-making, estimates of R_t are also used as the ground-truth in other models, such as models investigating the impact of non-pharmaceutical interventions [34, 35, 36, 37, 38, 39, 40, 41, 42], the effect of climate [43], or the relationship between mobility and transmission [44]. These citations all use EpiEstim to estimate R_t, most with the default value of $k = 7$. Some of these citations correct for delays induced by the smoothing parameter by shifting estimates by $k/2$ days, but this is a deterministic correction that ignores uncertainty about k. The Bayesian nature of our methods allows for the propagation of uncertainty in k through to estimates of R_t and thus to downstream models estimating the effect of interventions, climate, or mobility.

While we focus on parametric uncertainty, by fitting our models to simulated data from a range of dynamic models, one of our key findings is that marginalising out smoothing parameters results in increased model robustness, even when ignoring structural uncertainty. This is observed by improved coverage of predictive cases and decreased CRPS in simulations and on real data, and by improved coverage of R_t in simulations. As R_t is always unknown, it is not possible to compare coverage of this quantity on real data, and we rely on the CRPS on observables as a proxy for model performance (which is justified by our results on simulated data).

We used two metrics to evaluate model performance: coverage of the 95% credible intervals and CRPS. The former is a measure of calibration only, while the latter also factors in precision (and simultaneously considers all credible intervals, not just the 95% level). Calibration and precision are often a trade-off: correct calibration can usually be obtained by decreasing precision (i.e., widening credible intervals). The CRPS is a principled way of balancing these two goals. Alternative scoring rules may be appropriate depending on context. The documentation of [45] summarises scoring rules in an epidemiological context.

Our results suggest that improved CRPS implies better estimates of R_t, even when the model is misspecified (i.e., when the structure neither matches the ground or simulated truth). However, there is no guarantee of performance under model misspecification. Since some misspecification is inevitable when modelling infectious diseases, any scoring rule should be interpreted with caution. One example is observation noise: EpiEstim and EpiFilter assume that the epidemic follows the Poisson renewal model and that the appropriate level of smoothing is fixed over time (supplementary section 5 investigates this further). We demonstrate that our methods help with robustness to observation noise in supplementary section 3.3, but the lack of explicitly allowing for observation noise remains a limitation. We chose the New Zealand dataset as an example of high-quality epidemiological data with limited reporting biases, but in many real-world applications, observation noise is a significant issue that must be factored into analysis.

Alternative parameter fitting procedures for EpiEstim have previously been proposed. In the supple-

mentary material of [11], the authors suggest selecting k such that the window contains sufficient cases as to reduce the posterior coefficient-of-variation to a desired level. Depending on philosophy, this is either a subjective decision about the bias-variance trade-off, or a way of choosing parameters to obtain a desired level of confidence in R_t estimates. In either case, choosing a value of k with relatively low likelihood is likely to lead to poor model calibration. Alternatively, APEestim [23] proposed an information-theoretic approach to selecting k. While their approach is principled, it results in the selection of k shifted by one unit compared to our method, and does not allow for the marginalisation of uncertainty when there are multiple plausible values of k. We compare APEestim with our methods in supplementary section 6. Finally, as both EpiEstim and EpiFilter are Bayesian estimators, we argue that marginalising out the smoothing parameter is a more justified approach than selecting a single value.

While our methods can be applied to any sequential model (i.e., any model that estimates R_t using $C_{1:t}$, and then R_{t+1} using $C_{1:t+1}$, and so on), other methods exist that approach the smoothing problem differently. For example, EpiNow2 [19] also uses the renewal model, but models R_t using a Gaussian process. The smoothness of this model is primarily determined by the length-scale, which can either be assumed (in much the same way as default EpiEstim or EpiFilter) or estimated. Rather than using the sequential structure of the data, EpiNow2 fits the length-scale alongside R_t (and other parameters) using a Markov-chain Monte Carlo algorithm. Understanding the nuances of these various approaches to smoothing and hence inference-based decision-making will form a topic of future study.

The August 2021 outbreak of SARS-CoV-2 in Auckland, New Zealand provides a pertinent example of the importance of smoothing assumptions. R_t was first reported by officials as being less than 1 on 29 August, two days before any of our models (marginalised or default). While the official models were also based on the renewal model, they approached the smoothing problem differently: they assumed R_t was fixed at one value prior to the lockdown on 18 August, and then fixed at a different value after the lockdown [46], a piecewise constant model [16, 17]. These assumptions, provided they were correct, allowed researchers to be more confident in their estimates of R_t.

It is often argued that public policy decision-making should be "data-driven", with R_t frequently featuring as an example of "data" that could inform these decisions [47]. However, without careful model construction and accurate representation of uncertainty, estimates of R_t risk being influenced more by model assumptions than the underlying data. As we have demonstrated on both simulated and real-world data, these decisions will also be delayed relative to decisions made using models with more robust uncertainty quantification. Fortunately, our methods provide a simple and computationally efficient way to improve the robustness of these estimates, and to benchmark the uncertainty introduced or removed by smoothing assumptions.

Additional material

In addition to the supplementary material, all code is available on the GitHub repository. This repository contains all code necessary to reproduce the results in this paper, as well as Julia implementations of both EpiEstim and EpiFilter. Tutorials for the use of these methods are provided on the corresponding website.

Acknowledgements

N.S. acknowledges support from the Oxford-Radcliffe Scholarship from University College, Oxford, the EPSRC CDT in Modern Statistics and Statistical Machine Learning (Imperial College London and University of Oxford), and A. Maslov for studentship support. K.V.P. acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (Reference No. MR/X020258/1) funded by the UK Medical Research Council. This UK-funded grant is carried out in the frame of the Global Health EDCTP3 Joint Undertaking. The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Declaration of interests

The authors declare no competing interests.

Data availability statement

All data used in this study is publicly available from [25]. All data and code used in our analysis is available at https://github.com/nicsteyn2/RobustRtEstimators.

References

[1] Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly. "Are Epidemic Growth Rates More Informative than Reproduction Numbers?" In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement_1 (Nov. 2022), S5–S15. ISSN: 0964-1998. DOI: 10.1111/rssa.12867. (Visited on 06/27/2024).

[2] Ruth McCabe and Christl A. Donnelly. "Disease Transmission and Control Modelling at the Science–Policy Interface". In: *Interface Focus* 11.6 (Oct. 2021), p. 20210013. DOI: 10.1098/rsfs.2021.0013. (Visited on 09/12/2023).

[3] Ministry of Health NZ. *COVID-19 Update 30 August 4pm*. https://www.health.govt.nz/news/covid-19-update-30-august-4pm. Aug. 2021. (Visited on 09/04/2024).

[4] Ministry of Health NZ. *COVID-19 Update 11 October 2021 4pm*. https://www.health.govt.nz/news/covid-19-update-11-october-2021-4pm. Oct. 2021. (Visited on 09/04/2024).

[5] Center for Infectious Disease Research and Policy. *No 'reset' with Ebola Outbreak, WHO Official Says*. https://www.cidrap.umn.edu/ebola/no-reset-ebola-outbreak-who-official-says. June 2019. (Visited on 09/05/2024).

[6] 10 Downing Street Prime Minister's Office. *Prime Minister's Statement on Coronavirus (COVID-19): 10 May 2020*. https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-10-may-2020. May 2020. (Visited on 09/05/2024).

[7] Lorenzo Pellis et al. "Estimation of Reproduction Numbers in Real Time: Conceptual and Statistical Challenges". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 185.Supplement_1 (Nov. 2022), S112–S130. ISSN: 0964-1998. DOI: 10.1111/rssa.12955. (Visited on 09/05/2024).

[8] P. A. Media. "Schools in North-West of England Postpone Reopening Plans after New Coronavirus Data". In: *The Guardian* (June 2020). ISSN: 0261-3077. (Visited on 10/09/2024).

[9] The Royal Society. *Reproduction Number (R) and Growth Rate (r) of the COVID-19 Epidemic in the UK*. Tech. rep. 2020.

[10] Anne Cori and Adam Kucharski. "Inference of Epidemic Dynamics in the COVID-19 Era and Beyond". In: *Epidemics* 48 (Sept. 2024), p. 100784. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2024.100784. (Visited on 09/02/2024).

[11] Anne Cori et al. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics". In: *American Journal of Epidemiology* 178.9 (Nov. 2013), pp. 1505–1512. ISSN: 0002-9262. DOI: 10.1093/aje/kwt133. (Visited on 09/09/2023).

[12] Kris V. Parag. "Improved Estimation of Time-Varying Reproduction Numbers at Low Case Incidence and between Epidemic Waves". In: *PLOS Computational Biology* 17.9 (Sept. 2021), e1009347. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009347. (Visited on 09/10/2023).

[13] Thomas Maishman et al. "Statistical Methods Used to Combine the Effective Reproduction Number, R(t), and Other Related Measures of COVID-19 in the UK". In: *Statistical Methods in Medical Research* 31.9 (Sept. 2022), pp. 1757–1777. ISSN: 0962-2802. DOI: 10.1177/09622802221109506. (Visited on 09/05/2024).

[14] Michael A. Johansson et al. "An Open Challenge to Advance Probabilistic Forecasting for Dengue Epidemics". In: *Proceedings of the National Academy of Sciences* 116.48 (Nov. 2019), pp. 24268–24274. DOI: 10.1073/pnas.1909865116. (Visited on 09/09/2023).

[15] Niel Hens et al. "Estimating the Effective Reproduction Number for Pandemic Influenza from Notification Data Made Publicly Available in Real Time: A Multi-Country Analysis for Influenza A/H1N1v 2009". In: *Vaccine* 29.5 (Jan. 2011), pp. 896–904. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2010.05.010. (Visited on 04/17/2024).

[16] Christophe Fraser et al. "Influenza Transmission in Households During the 1918 Pandemic". In: *American Journal of Epidemiology* 174.5 (Sept. 2011), pp. 505–514. ISSN: 0002-9262. DOI: 10.1093/aje/kwr122. (Visited on 04/17/2024).

[17] Seth Flaxman et al. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe". In: *Nature* 584.7820 (Aug. 2020), pp. 257–261. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2405-7. (Visited on 04/17/2024).

[18] R. N. Thompson et al. "Improved Inference of Time-Varying Reproduction Numbers during Infectious Disease Outbreaks". In: *Epidemics* 29 (Dec. 2019), p. 100356. ISSN: 1755-4365. DOI:

10.1016/j.epidem.2019.100356. (Visited on 09/09/2023).

[19] Sam Abbott et al. "Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts". In: *Wellcome Open Research* 5 (Dec. 2020), p. 112. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.16006.2. (Visited on 09/09/2023).

[20] Shinsuke Koyama, Taiki Horie, and Shigeru Shinomoto. "Estimating the Time-Varying Reproduction Number of COVID-19 with a State-Space Method". In: *PLOS Computational Biology* 17.1 (Jan. 2021), e1008679. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008679. (Visited on 09/05/2024).

[21] Amin Azmon, Christel Faes, and Niel Hens. "On the Estimation of the Reproduction Number Based on Misreported Epidemic Data". In: *Statistics in Medicine* 33.7 (2014), pp. 1176–1192. ISSN: 1097-0258. DOI: 10.1002/sim.6015. (Visited on 09/05/2024).

[22] Oliver Eales et al. "Appropriately Smoothing Prevalence Data to Inform Estimates of Growth Rate and Reproduction Number". In: *Epidemics* 40 (Sept. 2022), p. 100604. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100604. (Visited on 09/06/2023).

[23] Kris V. Parag and Christl A. Donnelly. "Using Information Theory to Optimise Epidemic Models for Real-Time Prediction and Estimation". In: *PLOS Computational Biology* 16.7 (July 2020), e1007990. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007990. (Visited on 04/18/2024).

[24] Tilmann Gneiting and Adrian E Raftery. "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association* 102.477 (Mar. 2007), pp. 359–378. ISSN: 0162-1459. DOI: 10.1198/016214506000001437. (Visited on 09/13/2024).

[25] Ministry of Health NZ. *New Zealand COVID-19 Data*. (Visited on 09/13/2024).

[26] Kris V. Parag, Benjamin J. Cowling, and Christl A. Donnelly. "Deciphering Early-Warning Signals of SARS-CoV-2 Elimination and Resurgence from Limited Data at Multiple Scales". In: *Journal of The Royal Society Interface* 18.185 (Dec. 2021), p. 20210569. ISSN: 1742-5662. DOI: 10.1098/rsif.2021.0569. (Visited on 10/03/2024).

[27] N Ferguson et al. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand*. Tech. rep. Imperial College London, Mar. 2020. DOI: 10.25561/77482. (Visited on 09/09/2023).

[28] Ministry of Health NZ. *COVID-19 Update 1 September 2021*. https://www.health.govt.nz/news/covid-19-update-1-september-2021. Sept. 2021. (Visited on 09/15/2024).

[29] Ministry of Health NZ. *COVID-19 Update 2 September 2021*. https://www.health.govt.nz/news/covid-19-update-2-september-2021. Sept. 2021. (Visited on 09/15/2024).

[30] Ministry of Health NZ. *COVID-19 Update 13 September 2021 4pm*. https://www.health.govt.nz/news/covid-19-update-13-september-2021-4pm. Sept. 2021. (Visited on 09/13/2024).

[31] Ministry of Health NZ. *COVID-19 Update 18 October 2021 4pm*. https://www.health.govt.nz/news/covid-19-update-18-october-2021-4pm. Oct. 2021. (Visited on 09/15/2024).

[32] Ministry of Health NZ. *COVID-19 Update 19 October 2021*. https://www.health.govt.nz/news/covid-19-update-19-october-2021. Oct. 2021. (Visited on 09/15/2024).

[33] *Covid Crunch: Auckland Lockdown Cost $8 Billion*. https://www.nzherald.co.nz/nz/covid-19-delta-outbreak-auckland-lockdown-cost-8-billion/AGURSE2ZSR475B4FZ2LBFRKPZI/. Jan. 2022. (Visited on 09/13/2024).

[34] Louis Yat Hin Chan, Baoyin Yuan, and Matteo Convertino. "COVID-19 Non-Pharmaceutical Intervention Portfolio Effectiveness and Risk Communication Predominance". In: *Scientific Reports* 11.1 (May 2021), p. 10605. ISSN: 2045-2322. DOI: 10.1038/s41598-021-88309-1. (Visited on 09/04/2024).

[35] Chigozie A. Ogwara et al. "Impact of Public Health Policy and Mobility Change on Transmission Potential of Severe Acute Respiratory Syndrome Coronavirus 2 in Rhode Island, March 2020 – November 2021". In: *Pathogens and Global Health* 118.1 (Jan. 2024), pp. 65–79. ISSN: 2047-7724. DOI: 10.1080/20477724.2023.2201984. (Visited on 09/04/2024).

[36] Xiaoshuang Liu et al. "Differential Impact of Non-Pharmaceutical Public Health Interventions on COVID-19 Epidemics in the United States". In: *BMC Public Health* 21.1 (May 2021), p. 965. ISSN: 1471-2458. DOI: 10.

1186/s12889-021-10950-2. (Visited on 09/04/2024).

[37] Vesna Barros et al. "A Causal Inference Approach for Estimating Effects of Non-Pharmaceutical Interventions during Covid-19 Pandemic". In: *PLOS ONE* 17.9 (Sept. 2022), e0265289. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0265289. (Visited on 09/04/2024).

[38] Sylvia K. Ofori et al. "SARS-CoV-2 Transmission Potential and Rural-Urban Disease Burden Disparities across Alabama, Louisiana, and Mississippi, March 2020 — May 2021". In: *Annals of Epidemiology* 71 (July 2022), pp. 1–8. ISSN: 1047-2797. DOI: 10.1016/j.annepidem.2022.04.006. (Visited on 09/04/2024).

[39] Nina Haug et al. "Ranking the Effectiveness of Worldwide COVID-19 Government Interventions". In: *Nature Human Behaviour* 4.12 (Dec. 2020), pp. 1303–1312. ISSN: 2397-3374. DOI: 10.1038/s41562-020-01009-0. (Visited on 09/04/2024).

[40] Yacong Bo et al. "Effectiveness of Non-Pharmaceutical Interventions on COVID-19 Transmission in 190 Countries from 23 January to 13 April 2020". In: *International Journal of Infectious Diseases* 102 (Jan. 2021), pp. 247–253. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.10.066. (Visited on 09/04/2024).

[41] David Rubin et al. "Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States". In: *JAMA Network Open* 3.7 (July 2020), e2016099. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2020.16099. (Visited on 09/04/2024).

[42] Michelle Kendall et al. "Epidemiological Changes on the Isle of Wight after the Launch of the NHS Test and Trace Programme: A Preliminary Analysis". In: *The Lancet Digital Health* 2.12 (Dec. 2020), e658–e666. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30241-7. (Visited on 09/04/2024).

[43] Rachel E. Baker et al. "Assessing the Influence of Climate on Wintertime SARS-CoV-2 Outbreaks". In: *Nature Communications* 12.1 (Feb. 2021), p. 846. ISSN: 2041-1723. DOI: 10.1038/s41467-021-20991-1. (Visited on 09/04/2024).

[44] Pierre Nouvellet et al. "Reduction in Mobility and COVID-19 Transmission". In: *Nature Communications* 12.1 (Feb. 2021), p. 1090. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21358-2. (Visited on 09/06/2023).

[45] Nikos Bosse et al. *Scoringutils: Utilities for Scoring and Assessing Predictions*. Nov. 2023. (Visited on 10/21/2024).

[46] Nicholas Steyn et al. "Technical Report: Update to Modelling 7 September 2021". In: (Sept. 2021).

[47] Gavin Freeguard. "The Story of the R Number: How an Obscure Epidemiological Figure Took over Our Lives. Part 4: The Politics of R". In: *Significance* 21.4 (Aug. 2024), pp. 19–22. ISSN: 1740-9705, 1740-9713. DOI: 10.1093/jrssig/qmae058. (Visited on 09/11/2024).