Likelihood-based estimation and prediction for misspecified epidemic models: an application to measles in Samoa

SMB 2021

David Wu[1]
with Helen Petousis-Harris[2], Janine Paynter[2], Vinod Suresh[1,3], and Oliver J. Maclaren[1]

June 2021

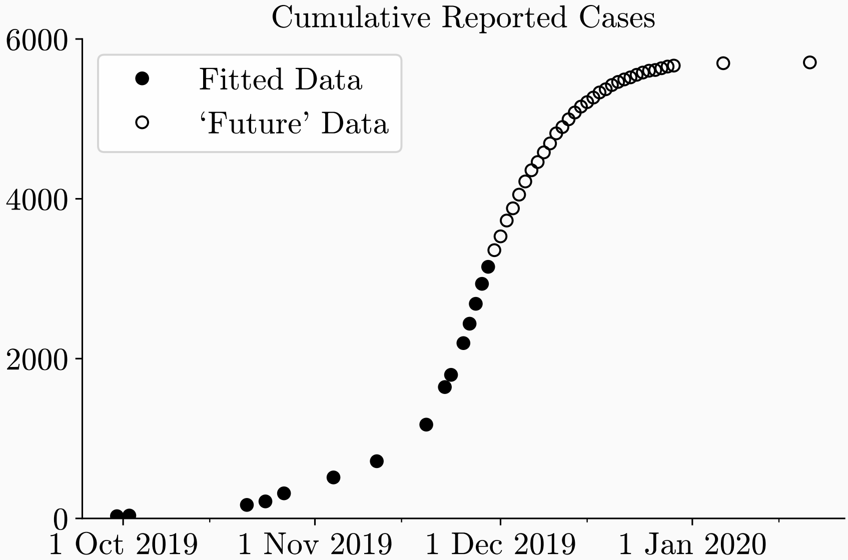
[1]Dept. Engineering Science, University of Auckland
[2]Dept. of General Practice and Primary Health Care, University of Auckland
[3]Auckland Bioengineering Institute, University of Auckland

2019 Samoan Measles Outbreak

Measles outbreak in the small South Pacific island nation of Samoa

- September 2019 – January 2020
- Over 5700 cases (pop. 200 000)
- 83 deaths
- Anomalously low vaccination coverage in infants
    - 40% MCV1
    - 28% MCV2

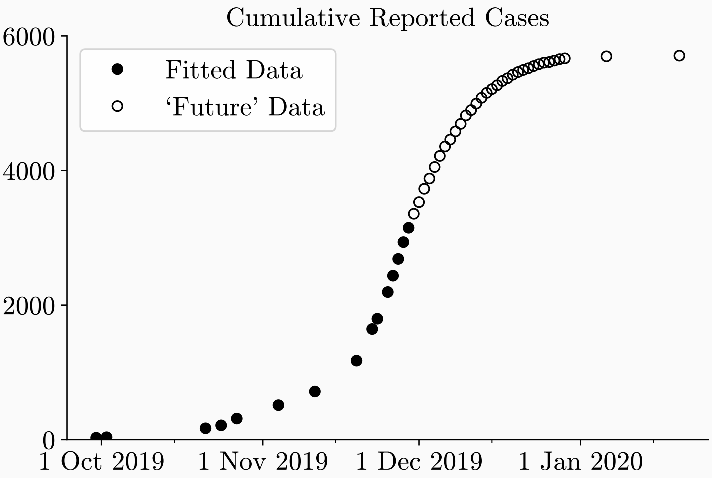
2019 Samoan Measles Outbreak



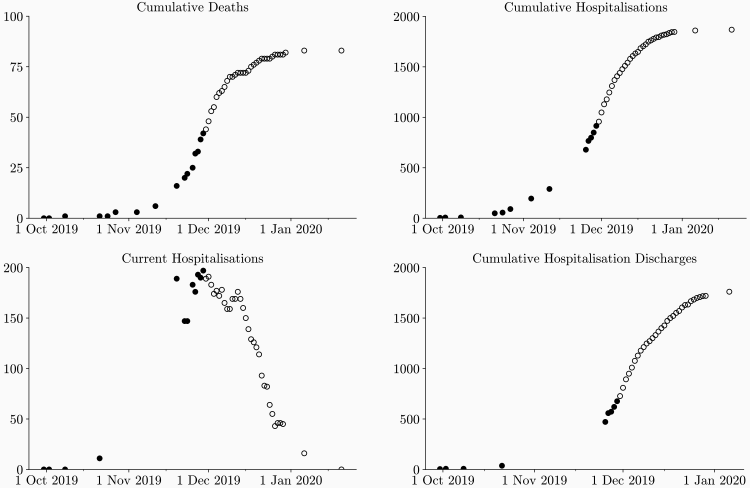
Core Questions

- How many cases in total?
- How many deaths in total?

Data collected is usually *noisy* and *incomplete*. Our models are only *partially observed*.



Data collected is usually *noisy* and *incomplete*. Our models are only *partially observed*.



Epidemic models are often modelled using *differential equations*.

$$\frac{dx}{dt} = f(x, \theta)$$

Some Problems: Models

Epidemic models are often modelled using *differential equations*.

$$\dot{S} = -\beta S I$$
$$\dot{I} = \beta S I - \alpha I$$

Some Problems: Models

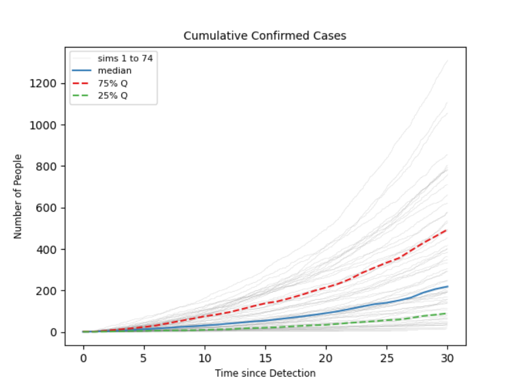
Epidemic models are often modelled using *differential equations*.

$$\frac{dx}{dt} = f(x, \theta)$$

These models are *idealised* and may not correctly reflect realistic processes.

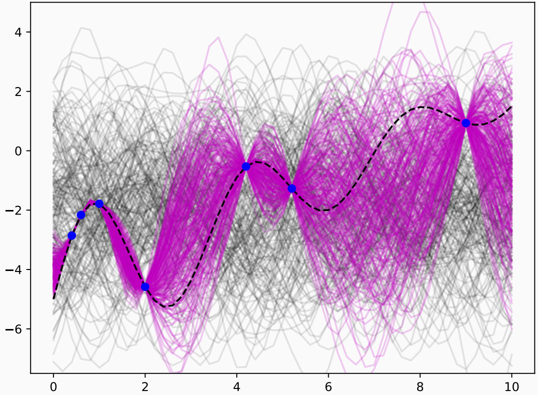
Dealing with Model Misfit

Our model may not correctly reflect reality – we can recover some of this by allowing for misfit as a random effect e.g. with a stochastic differential equation.



But this means we have to marginalise over the realisations!

A less expensive way is to use Gaussian processes to model the discrepancy*.



*Brynjarsdóttir and O'Hagan 2014, "Learning about physical parameters: the importance of model discrepancy".

Typical approach is to use nonlinear least-squares:

$$l(\theta) = \left\| y(t) - g\left(\int_{t_0}^{t} f(\tau, \theta) d\tau\right) \right\|^2$$

Can then pass to other algorithms to get estimators and uncertainties.

Standard Approaches

Typical approach is to use nonlinear least-squares:

$$l(\theta) = \left\| y(t) - g\left(\int_{t_0}^{t} f(\tau, \theta) d\tau \right) \right\|^2$$

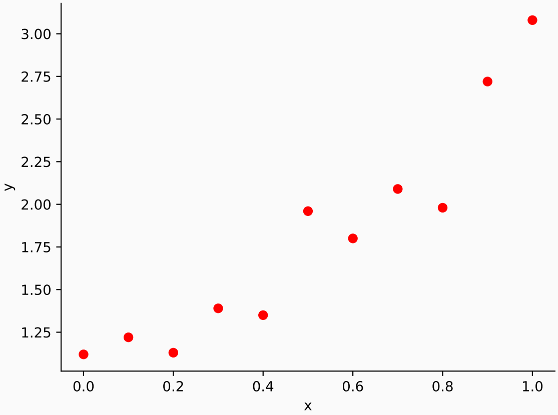
Can then pass to other algorithms to get estimators and uncertainties.

Lots of numerical integration!

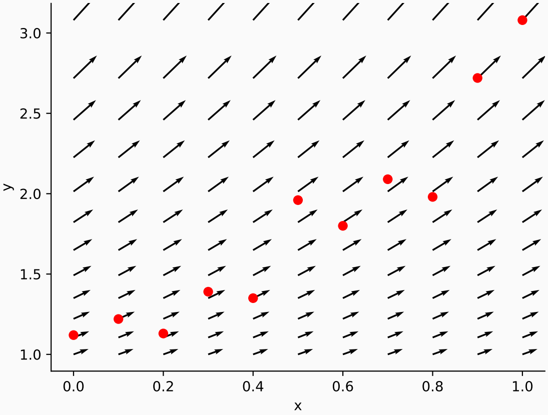
Enforces the model exactly!

Avoiding Integration: Two-Stage LS

Fit the derivatives to the ODE model directly!

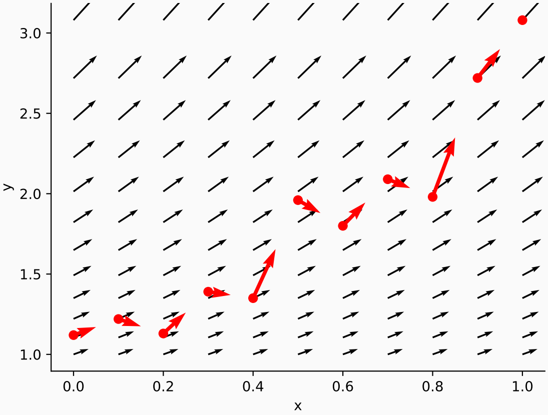


Fit the derivatives to the ODE model directly!



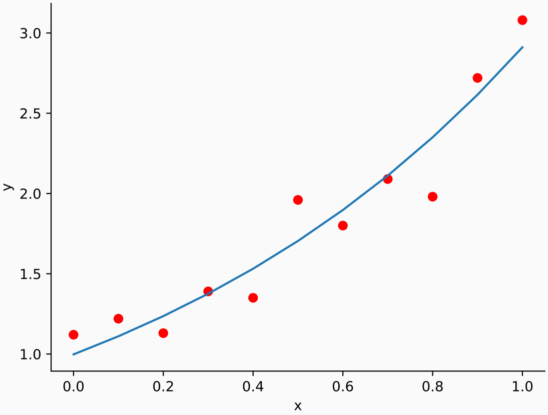
Avoiding Integration: Two-Stage LS

Fit the derivatives to the ODE model directly!



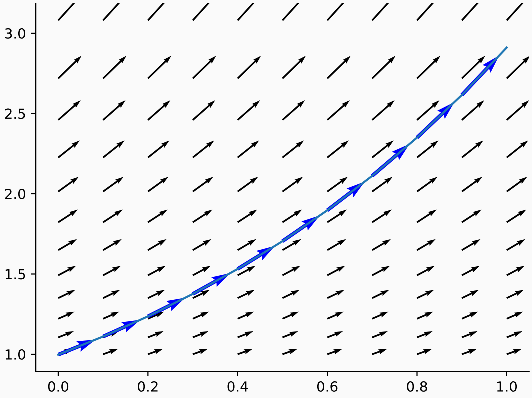
Avoiding Integration: Two-Stage LS

Fit the derivatives to the ODE model directly!



Avoiding Integration: Two-Stage LS

Fit the derivatives to the ODE model directly!



Avoiding Integration: Two-Stage LS

Fit the derivatives to the ODE model directly!

$$s(t) = \sum_{i}^{K} \varphi_i(t) c_i$$

$$\hat{s}(t) = \arg\min_{s} \left\{ \|y(t) - g(s(t))\|^2 + \lambda \left\| \frac{d^2 s}{dt^2} \right\|^2 \right\}$$

$$l(\theta) = \left\| \frac{d\hat{s}}{dt} - f(\hat{s}, \theta) \right\|^2$$

- Need to tune hyperparameter
- Breaks down if we only have partially observed models (can't compute f)

Generalised profiling: Penalise the smooth using the model!

$$s(t) = \sum_{i}^{K} \varphi_i(t) c_i$$

$$\hat{s}_\theta(t) = \arg\min_{s} \left\{ \|y(t) - g(s(t))\|^2 + \lambda \left\| \frac{ds}{dt} - f(s(t), \theta) \right\|^2 \right\}$$

$$l(\theta) = \|y - g(\hat{s}_\theta)\|^2$$

first by Ramsay et al.[†], in the field of functional data analysis.

† Ramsay et al. 2007, "Parameter estimation for differential equations: a generalized smoothing approach".

GP Likelihood: Formulation

Reframe as likelihoods by assuming additive Gaussian noise model, and modelling the process as an SDE.

$$s(t) = \sum_{i}^{K} \varphi_i(t) c_i$$

$$l(\theta) = \|L(y - g(s))\|^2 + \left\| W \left(\frac{ds}{dt} - f(s, \theta) \right) \right\|^2$$

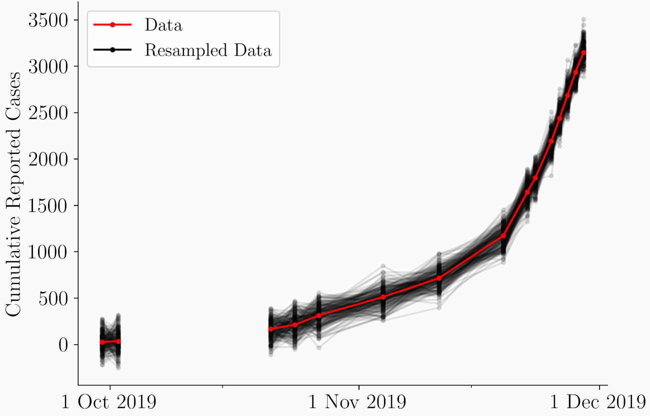
L and W come out as the whitening matrices of the covariance matrices of the associated noise.

GP Likelihood: Benefits

- Can add regularisation naturally by stacking terms
- Can use generalised least squares to automatically find L, W
- Allows for natural state estimation into the future
- Incorporates a model misfit
- No numerical integration

Can now just use frequentist methods to quantify uncertainty.

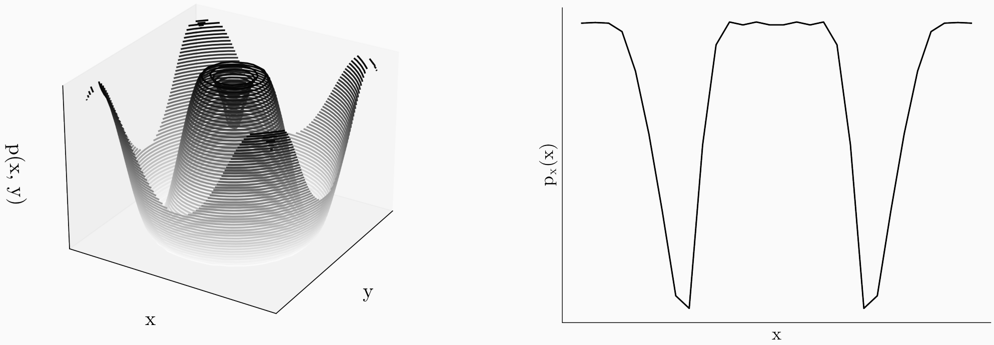
→ Bootstrap by resampling "data error" and "model error".



Uncertainty

Can now just use frequentist methods to quantify uncertainty.

$\rightarrow$ Compute profile likelihoods using constrained optimisation.



Back to the Case Study...

On 29th November 2019:

- over 3000 cases
- nearly 200 cases in hospital
- 42 deaths

Back to the Case Study...

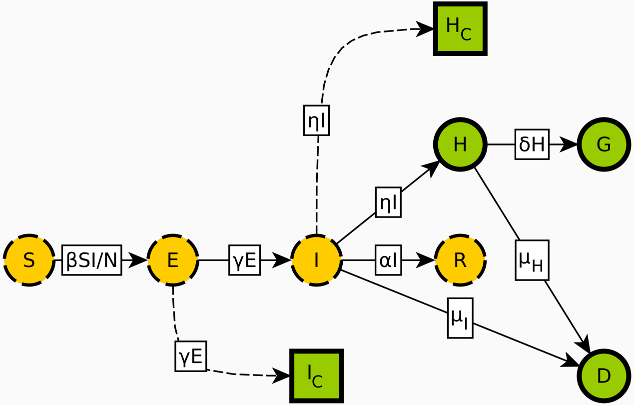
On 29th November 2019:

- over 3000 cases
- nearly 200 cases in hospital
- 42 deaths

At the time, we made a prediction of between 4500 and 6500 cases, and around 70 deaths, most of the cases within a month.

Model

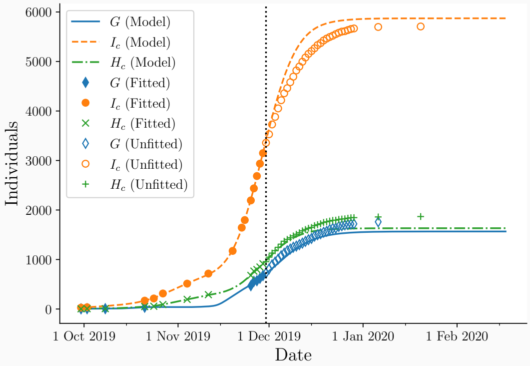
Use an SEIR model with hospitalisation.

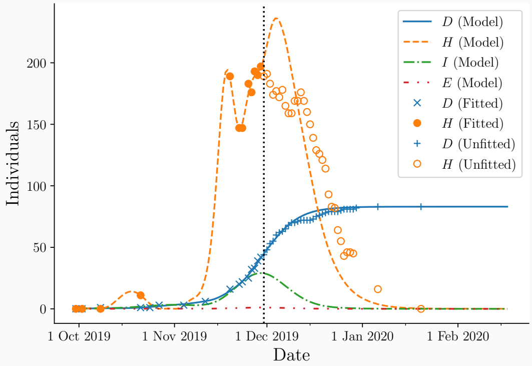


Only green, solid-bordered states are available in publicly available data.

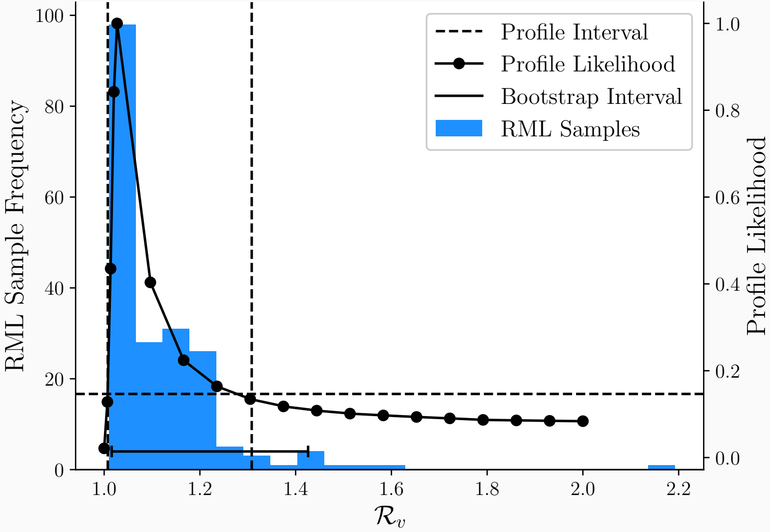
Maximum Likelihood Estimate

We fit to data from 30 Sept. 2020 to 29 Nov. 2020

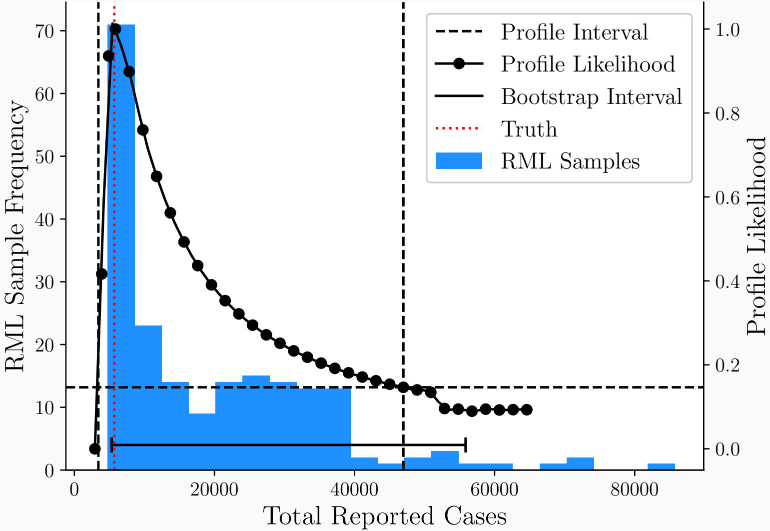




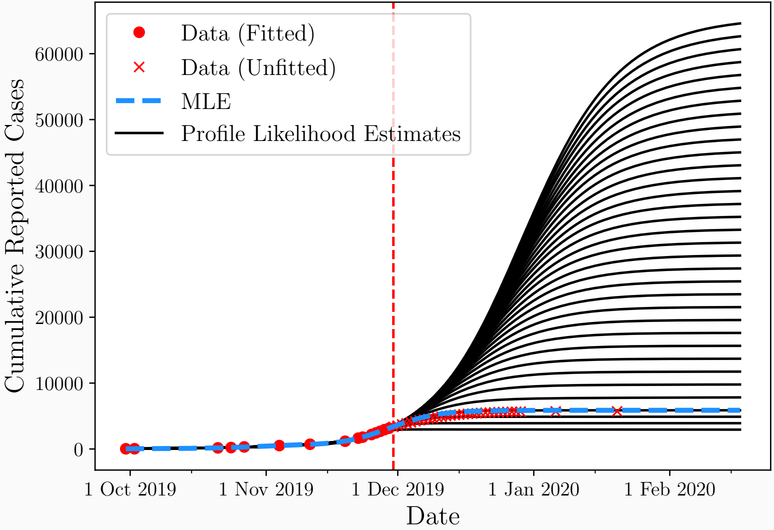
Quantifying Uncertainty: $\mathcal{R}_v$



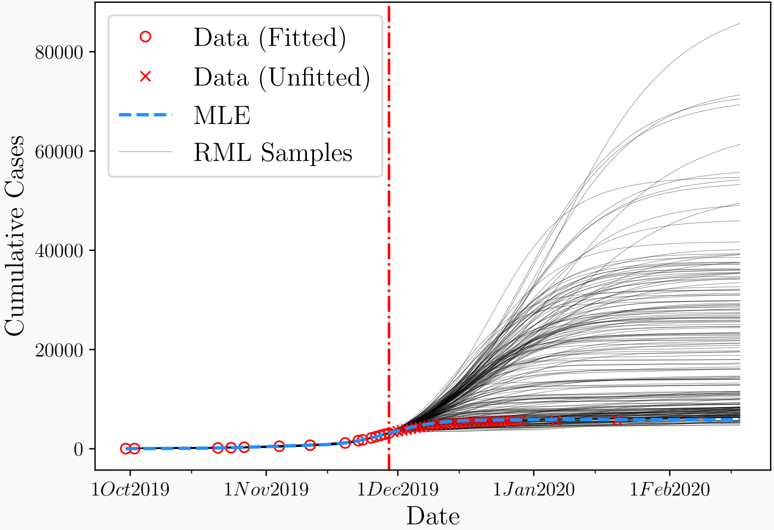
Quantifying Uncertainty: Total Cases



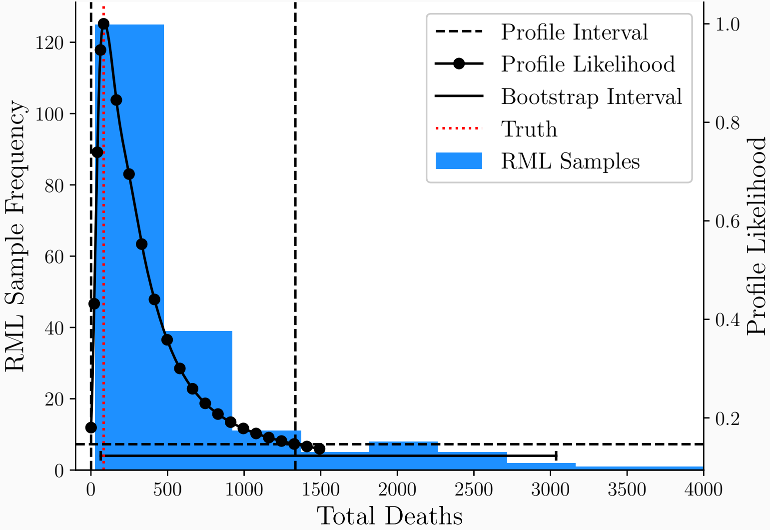
Quantifying Uncertainty: Total Cases



Quantifying Uncertainty: Total Cases



Quantifying Uncertainty: Total Deaths



Inference Takeaways

- Outbreak was likely to be around 6000 cases, and under 100 deaths.
- *but* the tail of the CI extends to horrifically high numbers.

Inference Takeaways

- Outbreak was likely to be around 6000 cases, and under 100 deaths.
- *but* the tail of the CI extends to horrifically high numbers.

The Government of Samoa implemented a curfew system on the 2nd Dec. 2019, and began a mass-vaccination campaign, and mitigated the likelihood of severe outcomes.

Recap

- We can avoid integration by using collocation methods like generalised profiling, which naturally gives us model misfit.
- Formulating it in a likelihood framework allows for natural uncertainty quantification and hyperparameter tuning.
- It works well (for a certain definition of "well") for the Samoan measles outbreak case study.

Thanks

Oliver Maclaren

Vinod Suresh

Helen Petousis-Harris

Janine Paynter

THE UNIVERSITY OF AUCKLAND NEW ZEALAND

samoaobserver

Thanks for coming!

Questions?

arXiv: 2103.16058

References

Brynjarsdóttir, Jenný and Anthony O'Hagan (Nov. 2014). "Learning about physical parameters: the importance of model discrepancy". In: *Inverse Problems* 30.11, p. 114007. ISSN: 0266-5611. DOI: *10.1088/0266-5611/30/11/114007*.

Ramsay, J. O. et al. (Nov. 2007). "Parameter estimation for differential equations: a generalized smoothing approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.5, pp. 741–796. ISSN: 13697412. DOI: *10.1111/j.1467-9868.2007.00610.x*.

Mathematical Formulation

Formulation

Model:

$$y = g(x) + \epsilon$$
$$dx = f(x, \theta)dt + \Sigma dW_t$$
$$\epsilon \sim \mathcal{N}(0, \Gamma)$$

Log-Likelihood function:

$$-2 \log \mathcal{L}(x, \theta) = \frac{1}{n} \|L(y - g(x))\|^2 + \frac{1}{m} \left\| W \left(\frac{dx}{dt} - f(x, \theta) \right) \right\|^2 + 2 \log |L| + 2 \log |W|$$
$$x = \Phi c, \quad L^T L = \Gamma^{-1}, \quad W^T W = \Sigma^{-1}$$
$$y \in \mathbb{R}^n, \quad \Phi \in \mathbb{R}^{m \times k}, \quad c \in \mathbb{R}^k$$

Regularisation as stacking

Have a regularisation term or additional data:

$$r(x,\theta) \sim \mathcal{N}(0,(R^T R)^{-1})$$

Can stack more terms to add regularisation:

$$\begin{aligned}-2\log\mathcal{L}(x,\theta) = &\frac{1}{n}\left\|L(y-g(x))\right\|^2 + \frac{1}{m}\left\|W\left(\frac{dx}{dt}-f(x,\theta)\right)\right\|^2 + \left\|Rr(\theta,x)\right\|^2 \\ &+ 2\log|L| + 2\log|W| + 2\log|R|\end{aligned}$$

Similar to Bayesian framing of priors and posteriors, except we are *maximising* over the distributions, instead of *marginalising*.

GLS/IRLS

$$-2\log\mathcal{L}(x,\theta)=\frac{1}{n}\left\|L(y-g(x))\right\|^2+\frac{1}{m}\left\|W\left(\frac{dx}{dt}-f(x,\theta)\right)\right\|^2+2\log|L|+2\log|W|$$

IRLS is repeated iterations of:

$$\hat{x}^{(i)},\hat{\theta}^{(i)}=\underset{x,\theta}{\arg\min}\left\{-2\log\mathcal{L}(x,\theta|\hat{L}^{(i)},\hat{W}^{(i)})\right\}$$

$$\hat{L}^{(i+1)},\hat{W}^{(i+1)}=\underset{L,W}{\arg\min}\left\{-2\log\mathcal{L}(L,W|\hat{x}^{(i)},\hat{\theta}^{(i)})\right\}$$

RML bootstrapping

$$
\begin{aligned}
-2\log\mathcal{L}(x,\theta) &= \frac{1}{n}\|L(y-g(x))\|^2 + \frac{1}{m}\left\|W\left(\frac{dx}{dt}-f(x,\theta)\right)\right\|^2 + 2\log|L| + 2\log|W| \\
&= \frac{1}{n}\|L(y-g(x))\|^2 + \frac{1}{m}\left\|W\left(r-\left(\frac{dx}{dt}-f(x,\theta)\right)\right)\right\|^2 + 2\log|L| + 2\log|W| \\
&\text{where } r = 0
\end{aligned}
$$

Resampled data is:

$$
\begin{aligned}
y^* &\sim \mathcal{N}(y,(L^TL)^{-1}) \\
r^* &\sim \mathcal{N}(r,(W^TW)^{-1})
\end{aligned}
$$

Profile Likelihood

The profile likelihood of a likelihood $\mathcal{L}(x, \theta)$ over function $\Omega(x, \theta)$ is:

$$\mathcal{L}_{\Omega(x,\theta)}(\omega) = \max_{x,\theta} \mathcal{L}(x, \theta | \Omega(x, \theta) = \omega)$$

Typically we see this used when profiling over a particular parameter:

$$\theta = (\theta^*, \tilde{\theta})$$

$$\mathcal{L}_{\theta^*}(\omega) = \max_{x,\tilde{\theta}} \mathcal{L}(x, \theta | \theta^* = \omega)$$

We also typically normalise the profile likelihood:

$$\mathcal{L}^*_{\Omega(x,\theta)}(\omega) = \frac{\mathcal{L}_{\Omega(x,\theta)}(\omega)}{\max_\omega \mathcal{L}_{\Omega(x,\theta)}(\omega)}$$

Nonlinear least squares

Model:

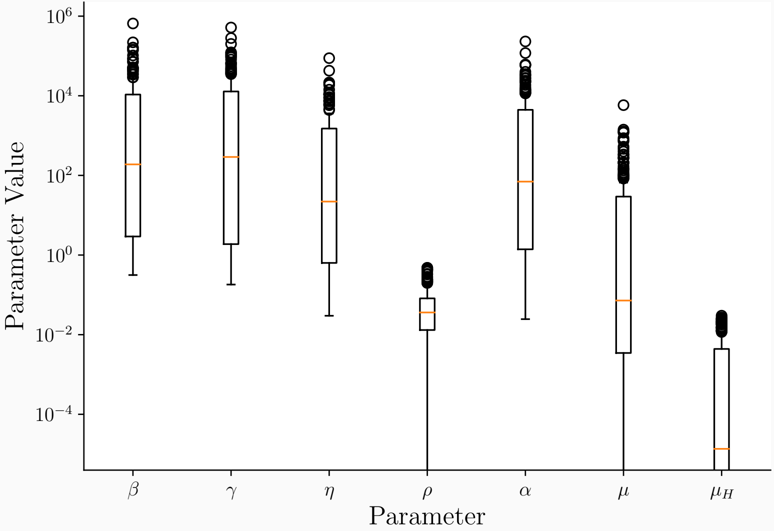
$$y = g(x) + \epsilon$$
$$\frac{dx}{dt} = f(x, \theta)$$
$$\epsilon \sim \mathcal{N}(0, \Gamma)$$

NLS is an optimisation problem with an embedded integration task.

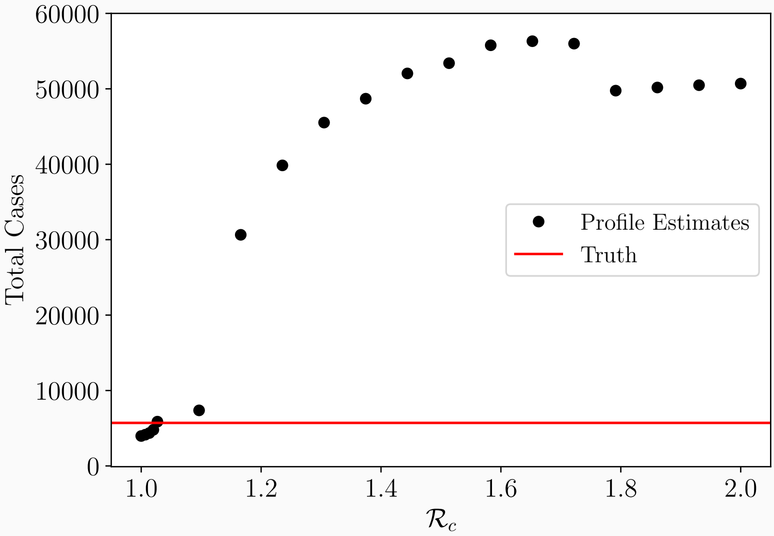
$$\min_{\theta, x_0} -2 \log \mathcal{L}(\theta, x_0) = \min_{\theta, x_0} \left\{ r^T \Gamma^{-1} r \right\}$$
$$r = y(t) - g\left(x(t)\right)$$
$$x(t) = \int_{t_0}^{t} f(\tau, \theta) d\tau$$
$$x(t_0) = x_0$$

Additional Results

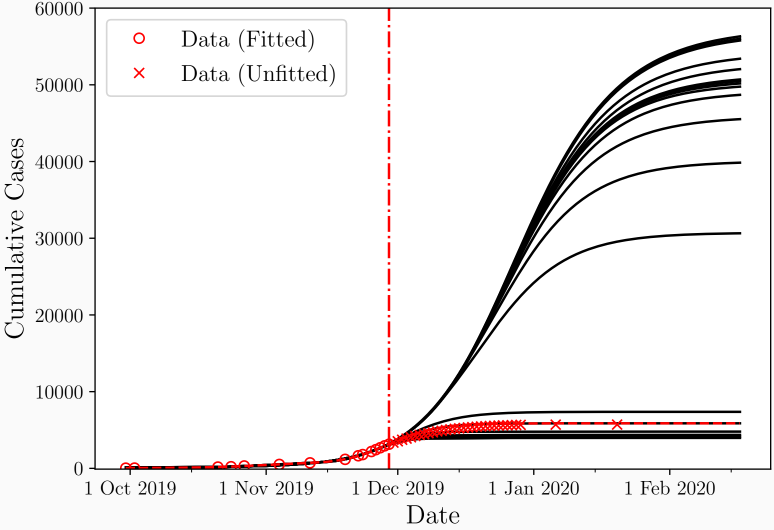
RML bootstrap parameter estimates



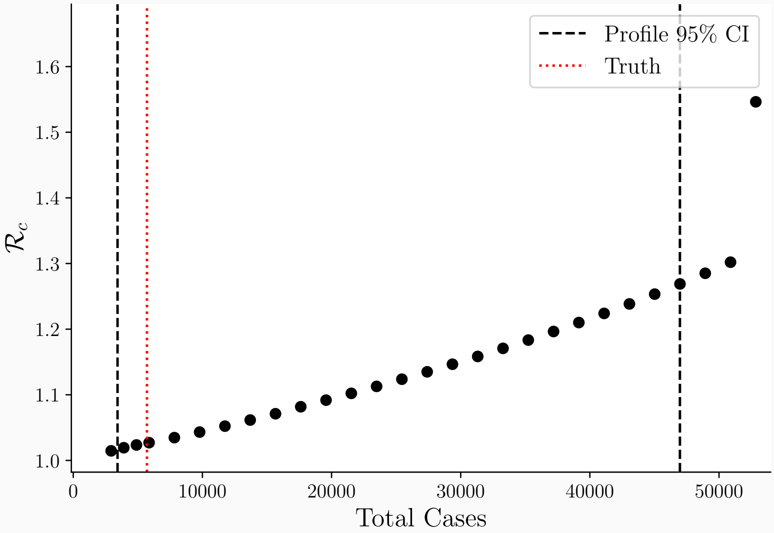
Total Cases along $\mathcal{R}_c$ profile



Cumulative Case Trajectory along $\mathcal{R}_c$ profile



$\mathcal{R}_c$ along Total Cases profile



Synthetic SEIR experiment

Observation Window Experiment

Have SEIR model

$$\dot{S} = -\beta SI/N \qquad \dot{E} = \beta SI/N - \gamma E$$

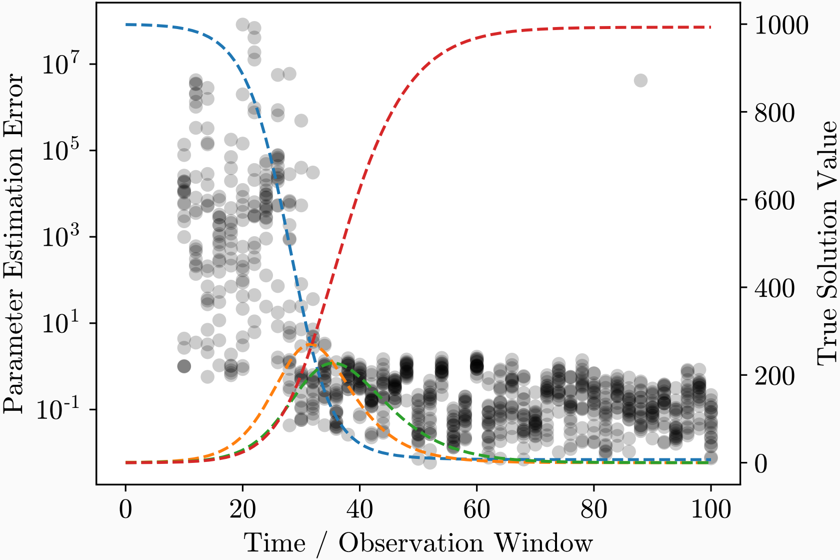
$$\dot{I} = \gamma E - \alpha I \qquad \dot{R} = \alpha I$$

Generate data y:

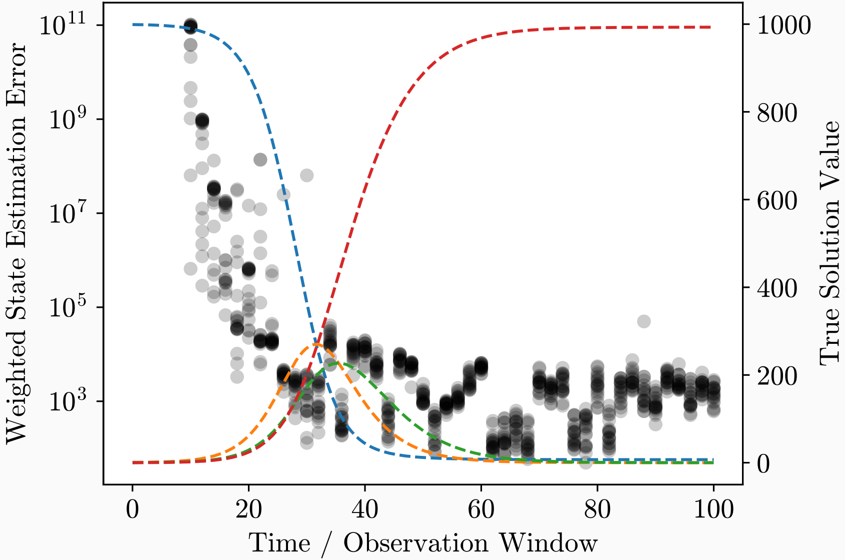
$$y_i \sim \text{Poisson}([S(t_i), R(t_i)]) \qquad t_i \in [0, 1, \ldots, T]$$

for $T \in [10, 11, \ldots, 100]$.

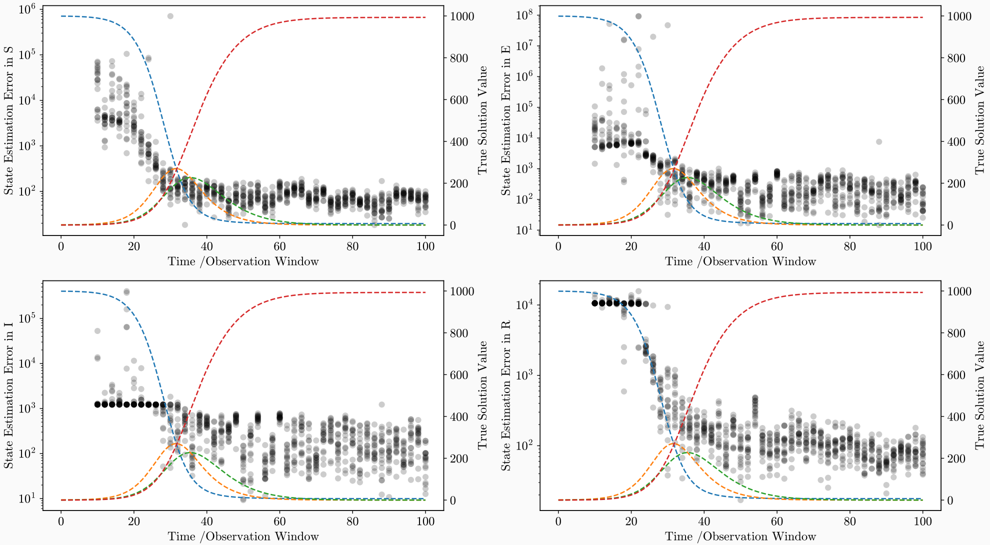
Observation Window Effects (SEIR model, synthetic data)



Observation Window Effects (SEIR model, synthetic data)



Observation Window Effects (SEIR model, synthetic data)



Synthetic SIR experiment

SIR Synthetic Data Validation

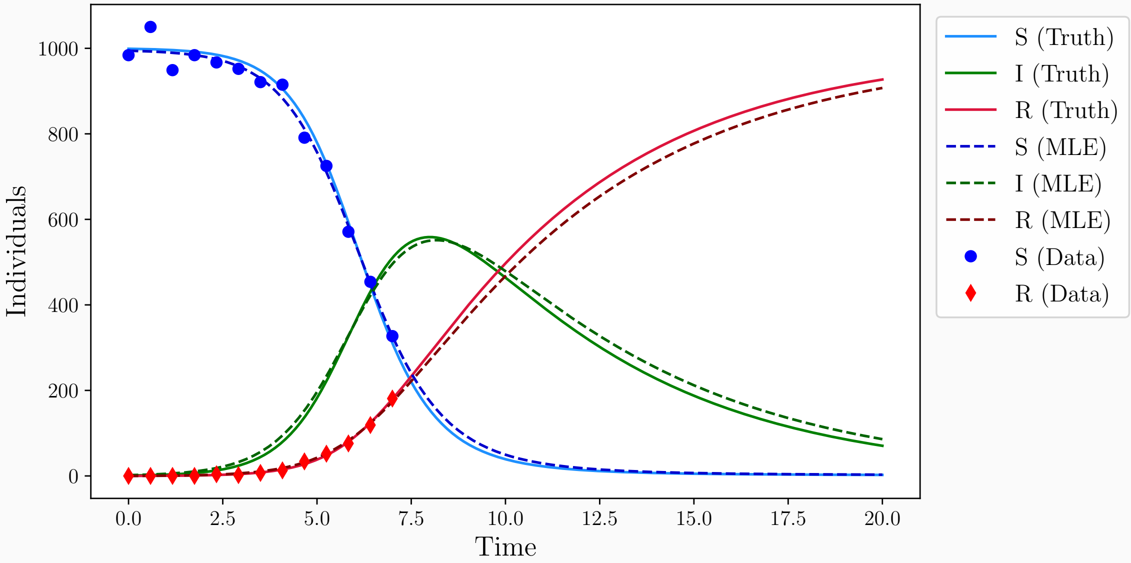
Have SIR model

$$\dot{S} = -\beta SI/N$$
$$\dot{I} = \beta SI/N - \alpha I$$
$$\dot{R} = \alpha I$$

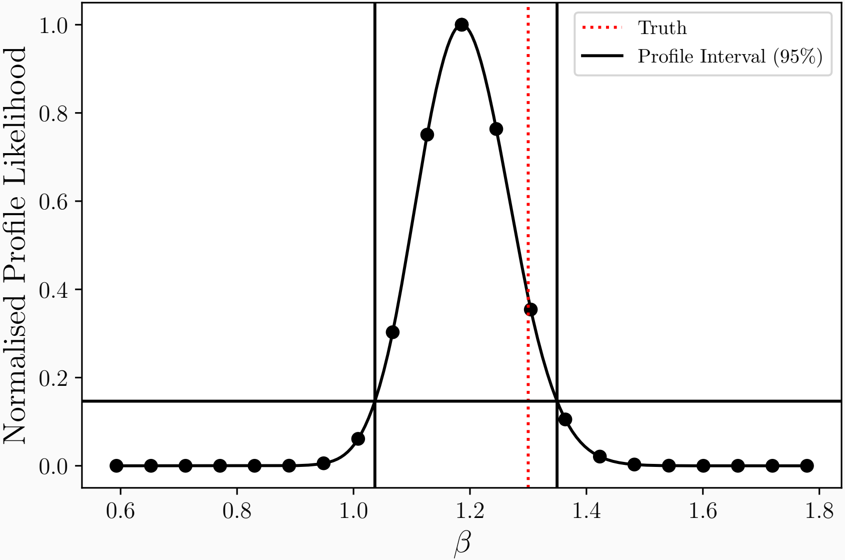
Generate data y

$$y_i \sim \text{Poisson}([S(t_i), R(t_i)])$$
$$t_i \in [0, \ldots, 7]$$

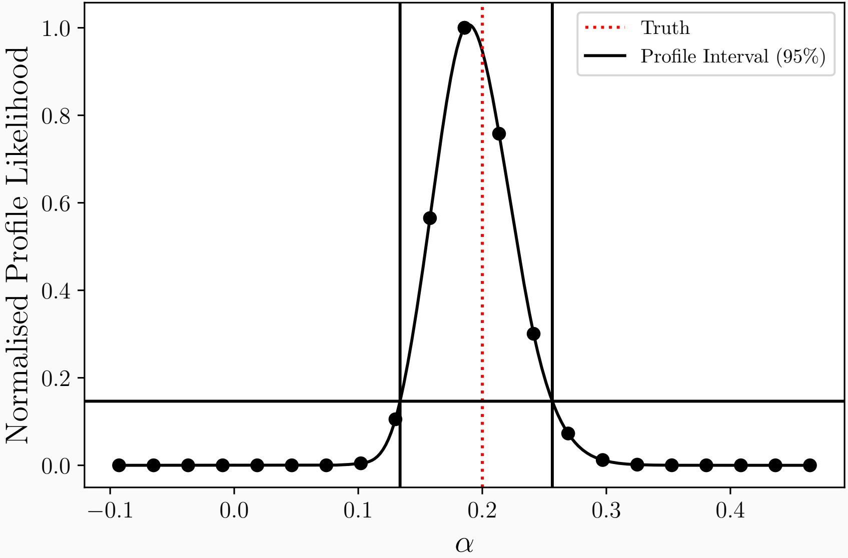
Fit with SIR model



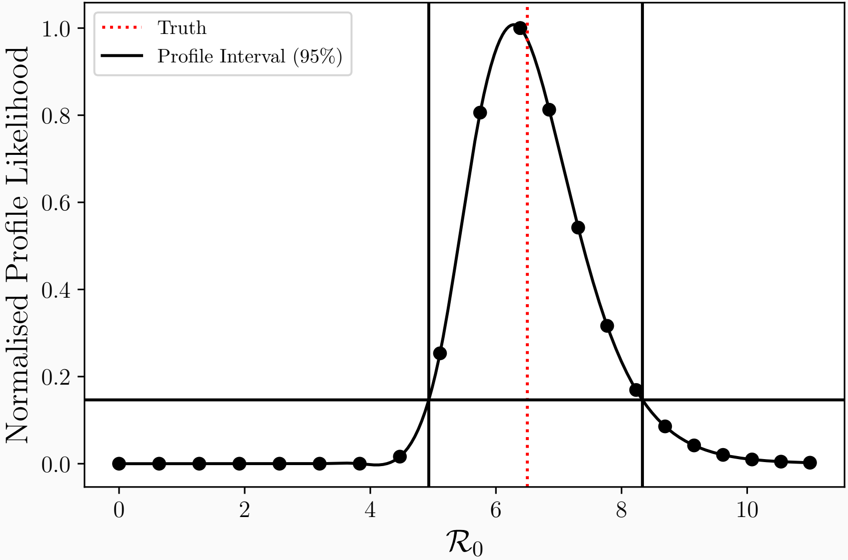
Profiles (SIR)



Profiles (SIR)



Profiles (SIR)



RML Bootstrap