Infectious disease outbreaks: inference and prediction under model misspecification and partially observed data

ANZIAM 2020

David Wu

February 2020

Department of Engineering Science
University of Auckland

Overview

1. Samoan measles outbreak (late 2019)

Overview

1. Samoan measles outbreak (late 2019)

2. Inferring model parameters, and making forecasts

Overview

1. Samoan measles outbreak (late 2019)
2. Inferring model parameters, and making forecasts
    - Relaxing the Nonlinear Least Squares Method

Overview

1. Samoan measles outbreak (late 2019)
2. Inferring model parameters, and making forecasts
    - Relaxing the Nonlinear Least Squares Method
    - Generalised Profiling

Overview

1. Samoan measles outbreak (late 2019)

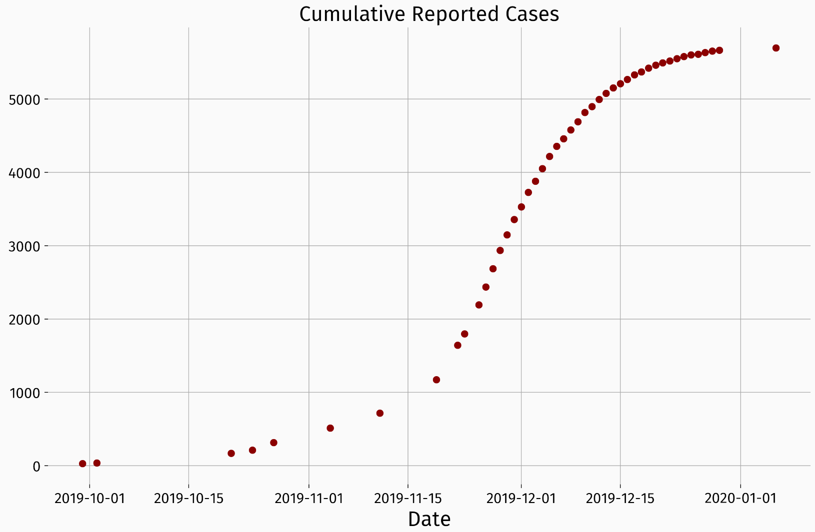
2. Inferring model parameters, and making forecasts
    - Relaxing the Nonlinear Least Squares Method
    - Generalised Profiling

3. Tuning the Method and Learnings about Model Misspecification

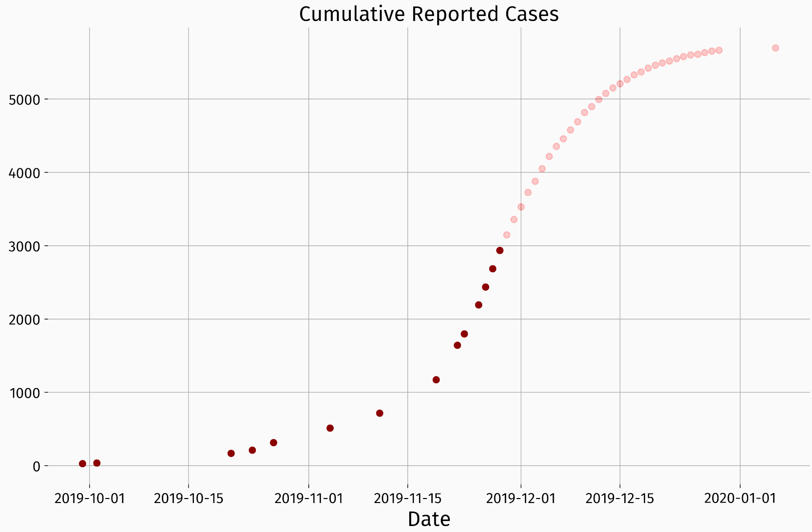
Context

Samoan Measles Outbreak 2019

Samoa Measles Outbreak 2019



Samoa Measles Outbreak 2019



Questions

1. How many people would be infected?

Questions

1. How many people would be infected?
2. For how long would this outbreak last?

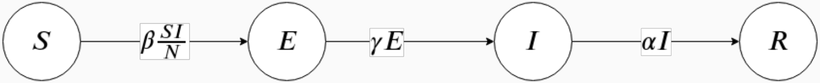
Questions

1. How many people would be infected?
2. For how long would this outbreak last?
3. How many people would die from measles?

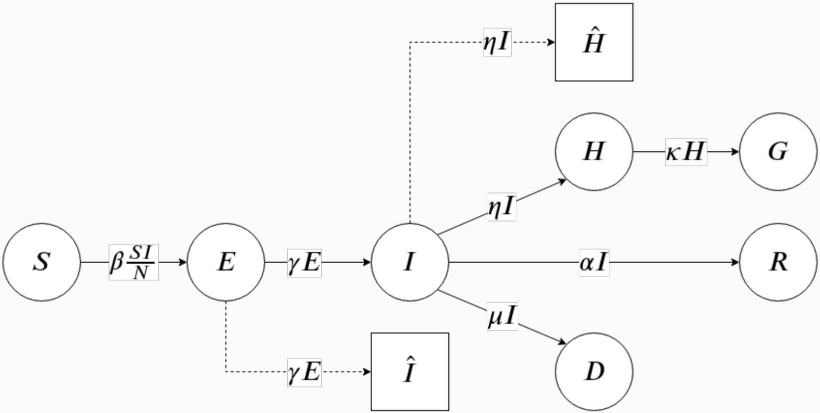
Modelling

Model

Simple SEIR model

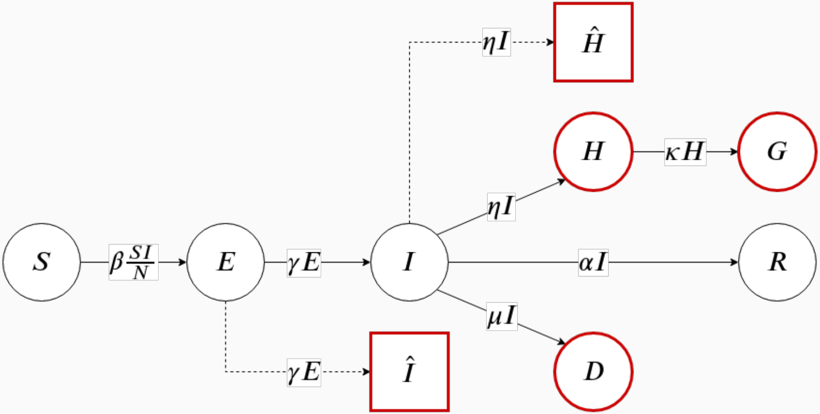


SEIR model with additional compartments



Model

Partially observed model



Inference

Parameter Inference – Nonlinear Least Squares

Given an ODE model

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

Parameter Inference – Nonlinear Least Squares

Given an ODE model

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

which observed by g, parameters θ can be fitted to data y by computing

$$\min_{\theta, x_0} \quad \left\| y - g\left(\int_0^t f(x(\tau); \theta) d\tau\right) \right\|^2,$$

$$x(0) = x_0$$

Parameter Inference – Nonlinear Least Squares

$$\min_{\theta, x_0} \quad \left\| y - g\left(\int_0^t f(x(\tau);\theta)d\tau\right)\right\|^2,$$

$$x(0) = x_0$$

Problems:

- Enforces the model exactly
- Can be expensive to compute the integral

Parameter Inference – Nonlinear Least Squares

$$\min_{\theta, x_0} \quad \left\| y - g\left(\int_0^t f(x(\tau); \theta) d\tau \right) \right\|^2,$$

$$x(0) = x_0$$

Problems:

- Enforces the model exactly
- Can be expensive to compute the integral

Standard variations and extensions of this method typically do not solve these problems!

Parameter Inference – Generalised Profiling

From the functional data analysis literature[*][†]

$$\min_{x,\theta} \quad \|y - g(x)\|^2 + \rho \left\| \frac{dx}{dt} - f(x;\theta) \right\|^2$$

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

[†] Hooker et al. 2011, "Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario."

Parameter Inference – Generalised Profiling

From the functional data analysis literature[*†]

$$\min_{x,\theta} \quad \underbrace{\|y - g(x)\|^2}_{\substack{\text{state estimation} \\ \text{data misfit}}} + \rho \underbrace{\left\| \frac{dx}{dt} - f(x;\theta) \right\|^2}_{\substack{\text{parameter estimation} \\ \text{model misfit}}}$$

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

[†] Hooker et al. 2011, "Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario."

Parameter Inference – Generalised Profiling

From the functional data analysis literature[*][†]

$$\min_{x,\theta} \quad \underbrace{\|y - g(x)\|^2}_{\substack{\text{state estimation}\\ \text{data misfit}}} + \rho \underbrace{\left\| \frac{dx}{dt} - f(x;\theta) \right\|^2}_{\substack{\text{parameter estimation}\\ \text{model misfit}}}$$

Approximate x by projecting onto a basis:

$$x \approx \Phi c$$

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

[†] Hooker et al. 2011, "Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario."

Parameter Inference – Generalised Profiling

From the functional data analysis literature[*][†]

$$\min_{x,\theta} \quad \underbrace{\|y - g(x)\|^2}_{\substack{\text{state estimation} \\ \text{data misfit}}} + \rho \underbrace{\left\| \frac{dx}{dt} - f(x;\theta) \right\|^2}_{\substack{\text{parameter estimation} \\ \text{model misfit}}}$$

Approximate x by projecting onto a basis:

$$x \approx \Phi c$$

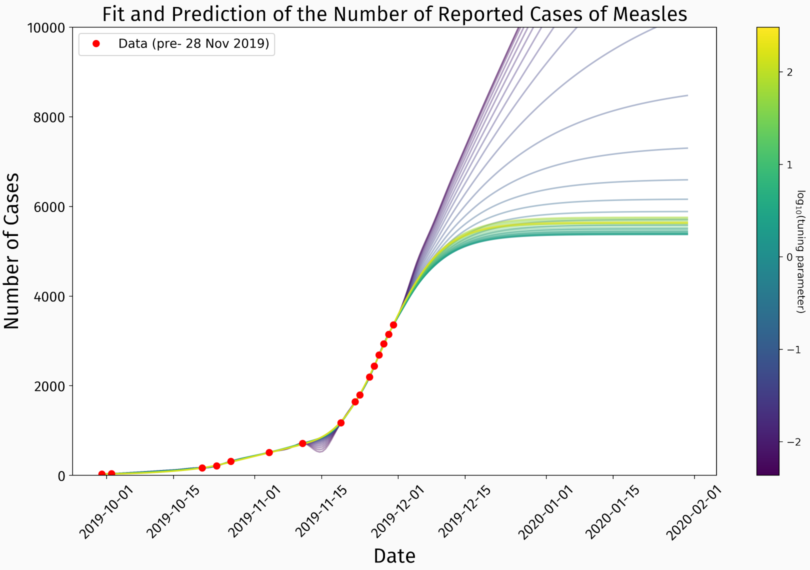
- Allows for model misspecification
- Cheap to evaluate the objective function (no integration)

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

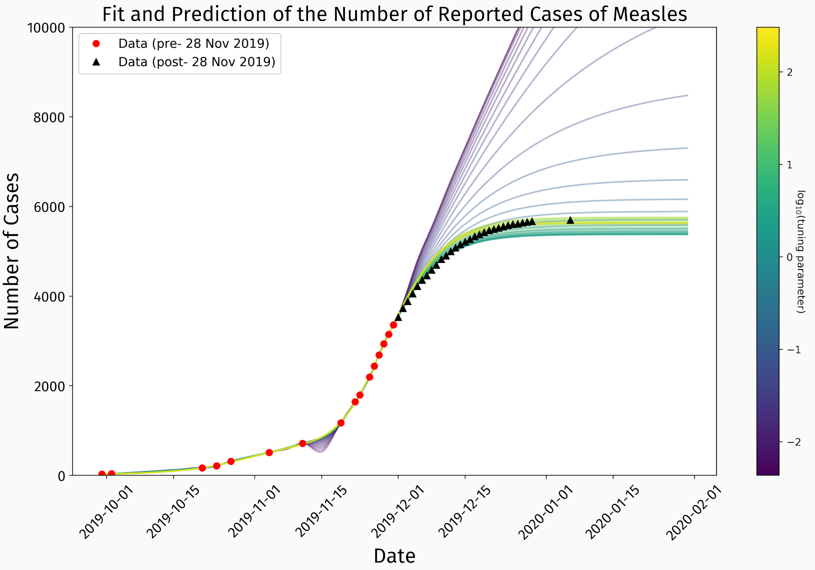
[†] Hooker et al. 2011, "Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario."

Results

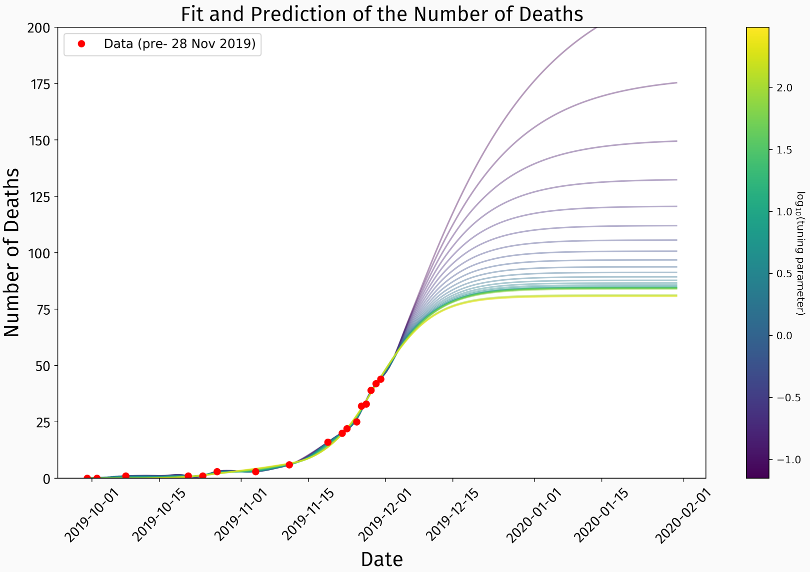
Results: Fit and Prediction



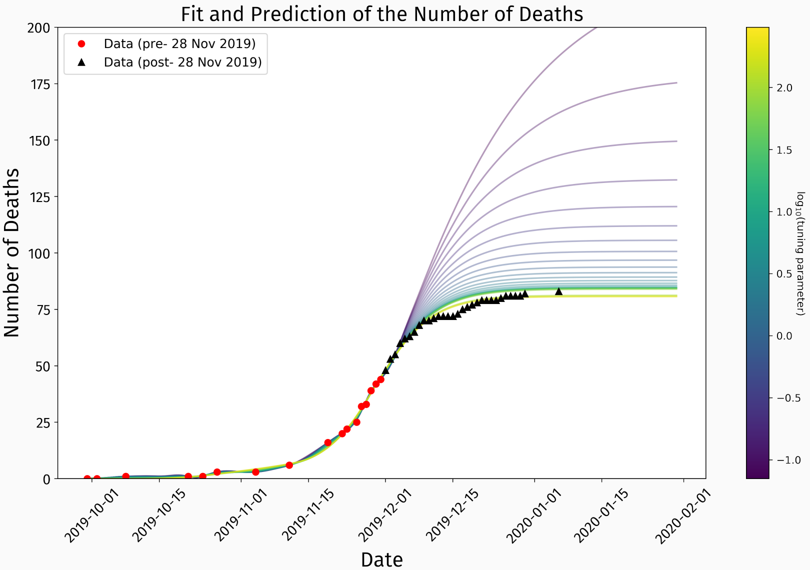
Results: Fit and Prediction



Results: Fit and Prediction



Results: Fit and Prediction



Choosing the Trade-off Parameter

Choosing ρ

$$\min_{x,\theta} \quad \|y - g(x)\|^2 + \rho \left\| \frac{dx}{dt} - f(x;\theta) \right\|^2$$

ρ acts as a regularisation.

As ρ increases, the model is imposed more strongly.

Choosing ρ

$$\min_{x,\theta} \quad \|y - g(x)\|^2 + \rho \left\| \frac{dx}{dt} - f(x;\theta) \right\|^2$$

ρ acts as a regularisation.

As ρ increases, the model is imposed more strongly.

- Model might be misspecified: don't necessarily want to impose exactly
- Without a model can't extrapolate or predict out of sample

Choosing ρ

$\rightarrow$ Can emphasize prediction of future data

$\rightarrow$ Can emphasize parameter estimation

Choosing ρ

$\rightarrow$ Can emphasize prediction of future data

$\rightarrow$ Can emphasize parameter estimation

$\Rightarrow$ Initially look at stability of parameter estimates and predictions as they vary with *rho*

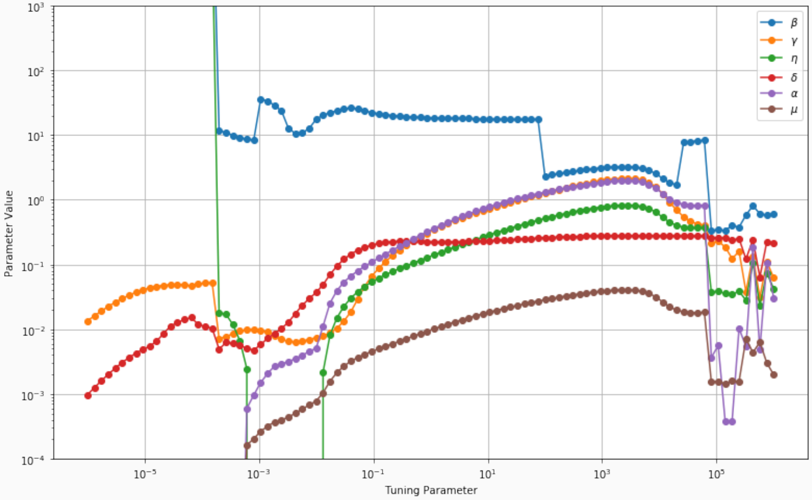
Choosing ρ

- $\rightarrow$ Can emphasize prediction of future data
- $\rightarrow$ Can emphasize parameter estimation

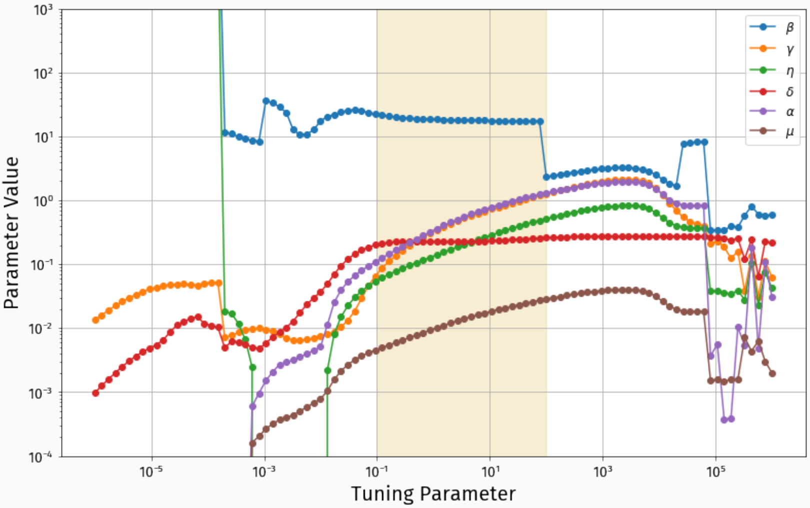
$\Rightarrow$ Initially look at stability of parameter estimates and predictions as they vary with *rho*

- Simple, intuitive
- But not formal statistical uncertainty!

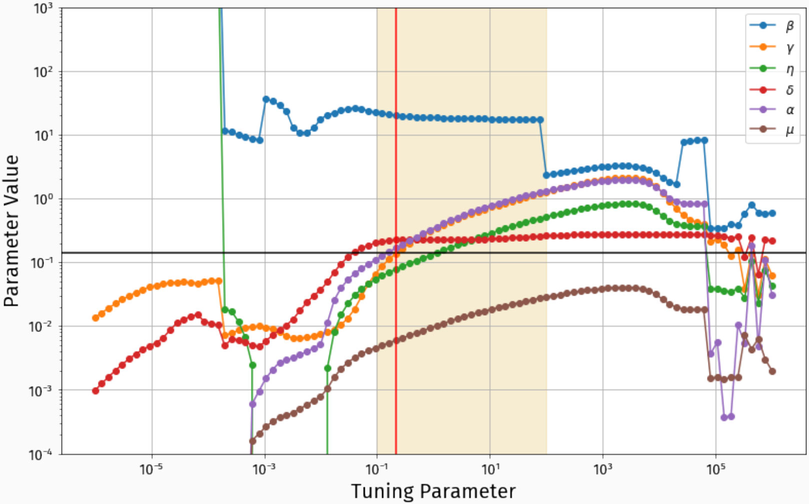
Estimated Parameter Values by Tuning Parameter



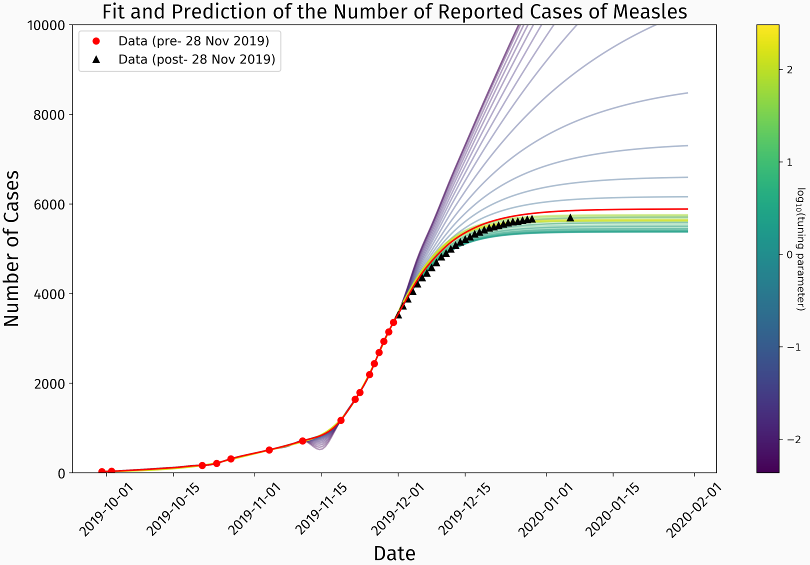
Estimated Parameter Values by Tuning Parameter



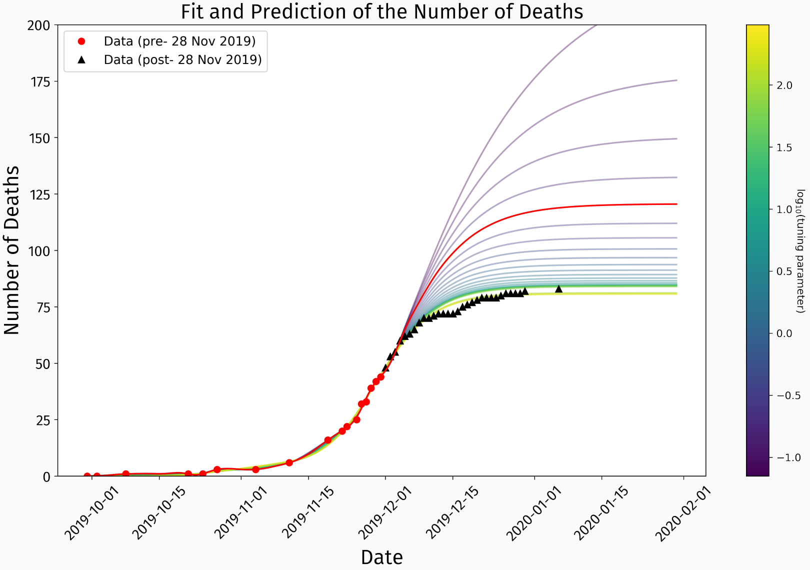
Tuning by Enforcing Parameter Priors



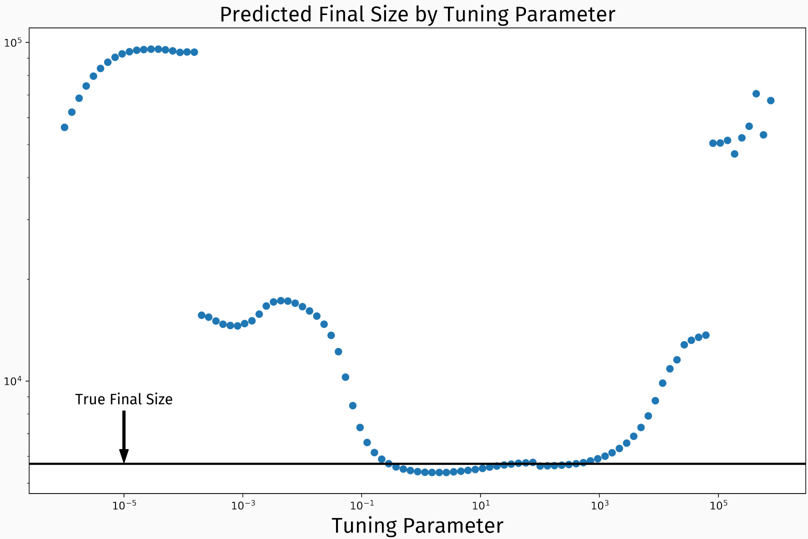
Tuning by Enforcing Parameter Priors



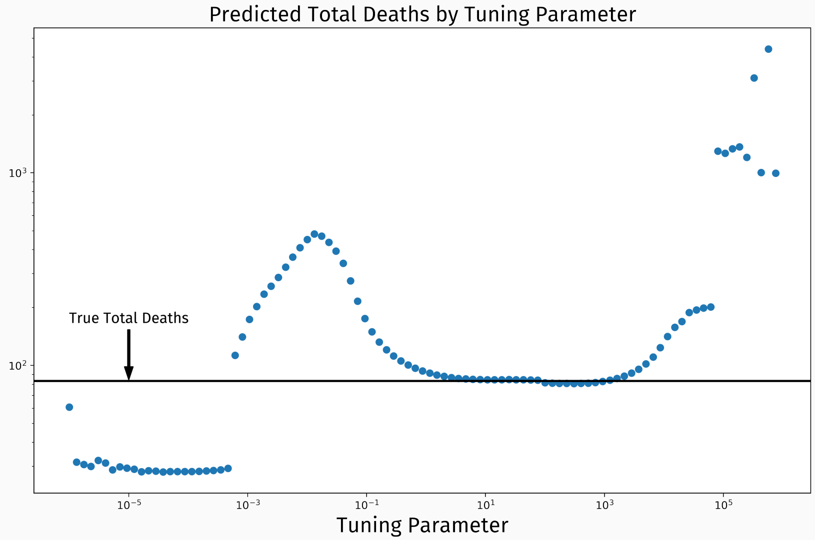
Tuning by Enforcing Parameter Priors



Final Size Cross-Validation



Total Deaths Cross-Validation



Observations

- There is a stable, *intermediate* zone for the regularisation strength
- This zone corresponds to
    - good prediction when compared to future data
    - reasonable parameter estimates when compared to prior information
- Akin to bias-variance tradeoffs

Observations

- There is a stable, *intermediate* zone for the regularisation strength
- This zone corresponds to
    - good prediction when compared to future data
    - reasonable parameter estimates when compared to prior information
- Akin to bias-variance tradeoffs

$\Rightarrow$ Neither *exact* model or *no* model appears to be best for *either* prediction or estimation.

Observations

- There is a stable, *intermediate* zone for the regularisation strength
- This zone corresponds to
  - good prediction when compared to future data
  - reasonable parameter estimates when compared to prior information
- Akin to bias-variance tradeoffs

$\Rightarrow$ Neither *exact* model or *no* model appears to be best for *either* prediction or estimation.

Instead, an *approximately* enforced model seems 'best' for both goals.

Future Work

- Formalising tuning methods for ρ
- More formal uncertainty quantification

Summary and Closing

Summary

1. Generalised profiling is an efficient method for estimation under model misspecification
2. Allows the prediction of a typical measles outbreak (Samoa 2019)
3. Misspecification can cause problems for tuning fitting methods

Thanks

Oliver Maclaren

Vinod Suresh

Helen Petousis-Harris

Thank You!

Questions?

References

Hooker, Giles et al. (July 2011). “Parameterizing state-space models for infectious disease dynamics by generalized profiling: measles in Ontario.”. In: *Journal of the Royal Society, Interface* 8.60, pp. 961–74. ISSN: 1742-5662. DOI: **10.1098/rsif.2010.0412**.

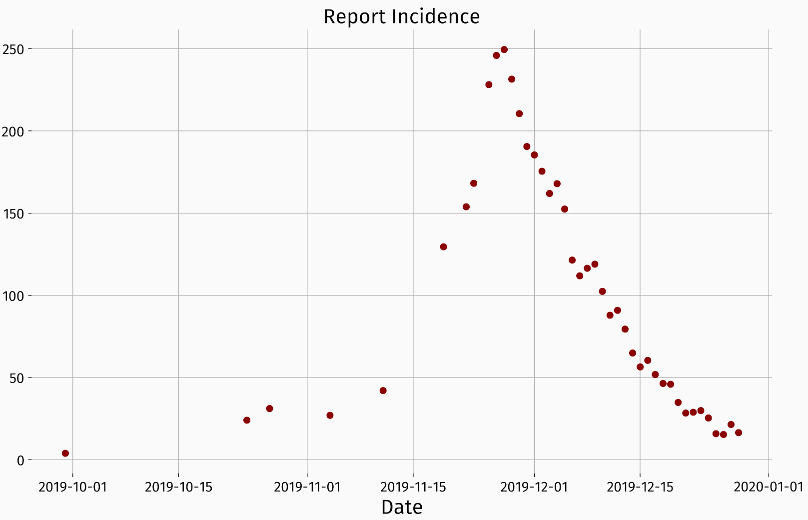
Ramsay, J. O. et al. (Nov. 2007). “Parameter estimation for differential equations: a generalized smoothing approach”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.5, pp. 741–796. ISSN: 13697412. DOI: **10.1111/j.1467-9868.2007.00610.x**.

Model (Equation Form)

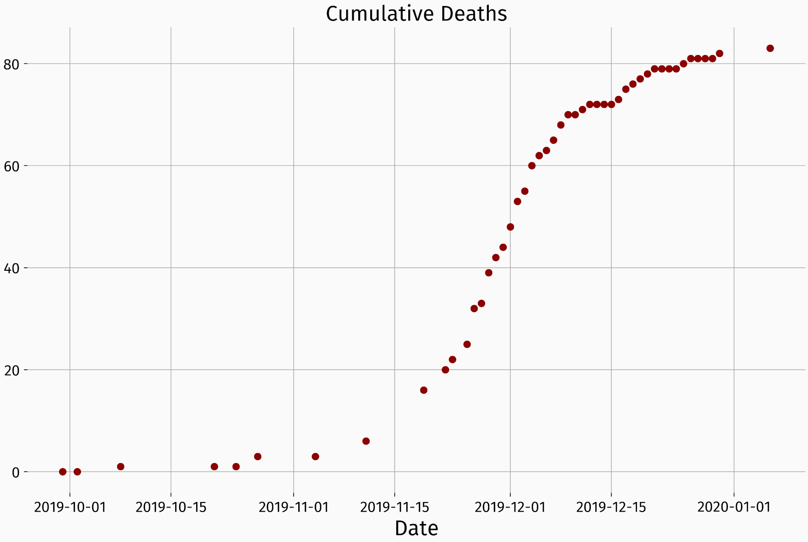
$$
\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{SI}{N} \\
\frac{dE}{dt} &= \beta\frac{SI}{N} - \gamma E \\
\frac{dI}{dt} &= \gamma E - \alpha I - \mu I \\
\frac{dR}{dt} &= (1-\eta)\alpha I
\end{aligned}
$$

$$
\begin{aligned}
\frac{dH}{dt} &= \eta\alpha I - \kappa H \\
\frac{dG}{dt} &= \kappa H \\
\frac{dD}{dt} &= \mu I \\
\frac{dI_c}{dt} &= \gamma E \\
\frac{dH_c}{dt} &= \eta\alpha I
\end{aligned}
$$

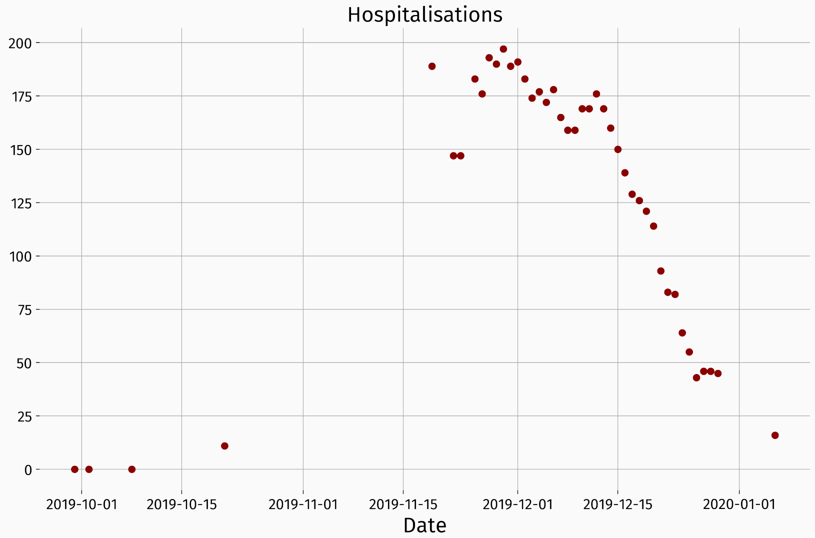
Report Incidence



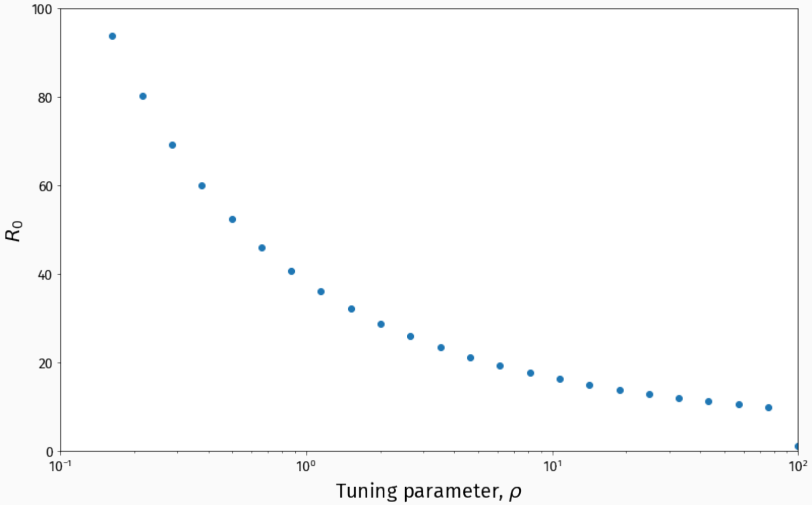
Deaths


Cumulative Deaths
80
60
40
20
0
2019-10-01
2019-10-15
2019-11-01
2019-11-15
2019-12-01
2019-12-15
2020-01-01
Date

Hospitalisations



R_0 by Tuning Parameter



Uncertainty?

We can capture the uncertainty in prediction with prediction intervals. The form of generalised profiling can be roughly interpreted as

$$y - g(x;\theta) \sim \mathcal{N}(0, \sigma_o^2)$$

$$\frac{dx}{dt} - f(x;\theta) \sim \mathcal{N}(0, \sigma_m^2)$$

$$\sigma_m^2 = \rho\sigma_o^2$$

We can approximate σ_o^2, and use that to estimate σ_m^2, which can be used to inform the prediction interval in time.

An Inherent Tradeoff