Progress 08 - 10/01/2021

Pete Dockrill - 11/01/2021

Progress

- Found several useful features of the Covasim framework which help quantify fit quality, such as Mean Absolute Error
- Have managed to achieve a total loss of 2.517 by fitting to the first wave using β and *pop_infected*, but general trend of the data still not reflected in the simulations
- The median simulation does not depend on the number of simulations when run with the same parameters (misunderstanding?)
- Attempted to perform a more detailed parameter search around the optimal parameters, but did not manage to improve the fit
- Completely forgot to include data up until the end of August so data currently runs until beginning of July (will recalibrate for further studies using data up to end of August)
- Materials from "*Model fitting and inference for infectious disease dynamics*" short course at the LSHTM would be particularly useful to improve my understanding and fill in any gaps

My current understanding

Calibration (*Calibration_wave1.py*):

- Rough grid search for the parameter values that minimise the loss

Simulation (*Simulation_wave1.py*):

- Search through region around the optimal parameter values in parameter space to find any further decrease in loss

- Calculate the median of the top 10/50/100/etc. simulations

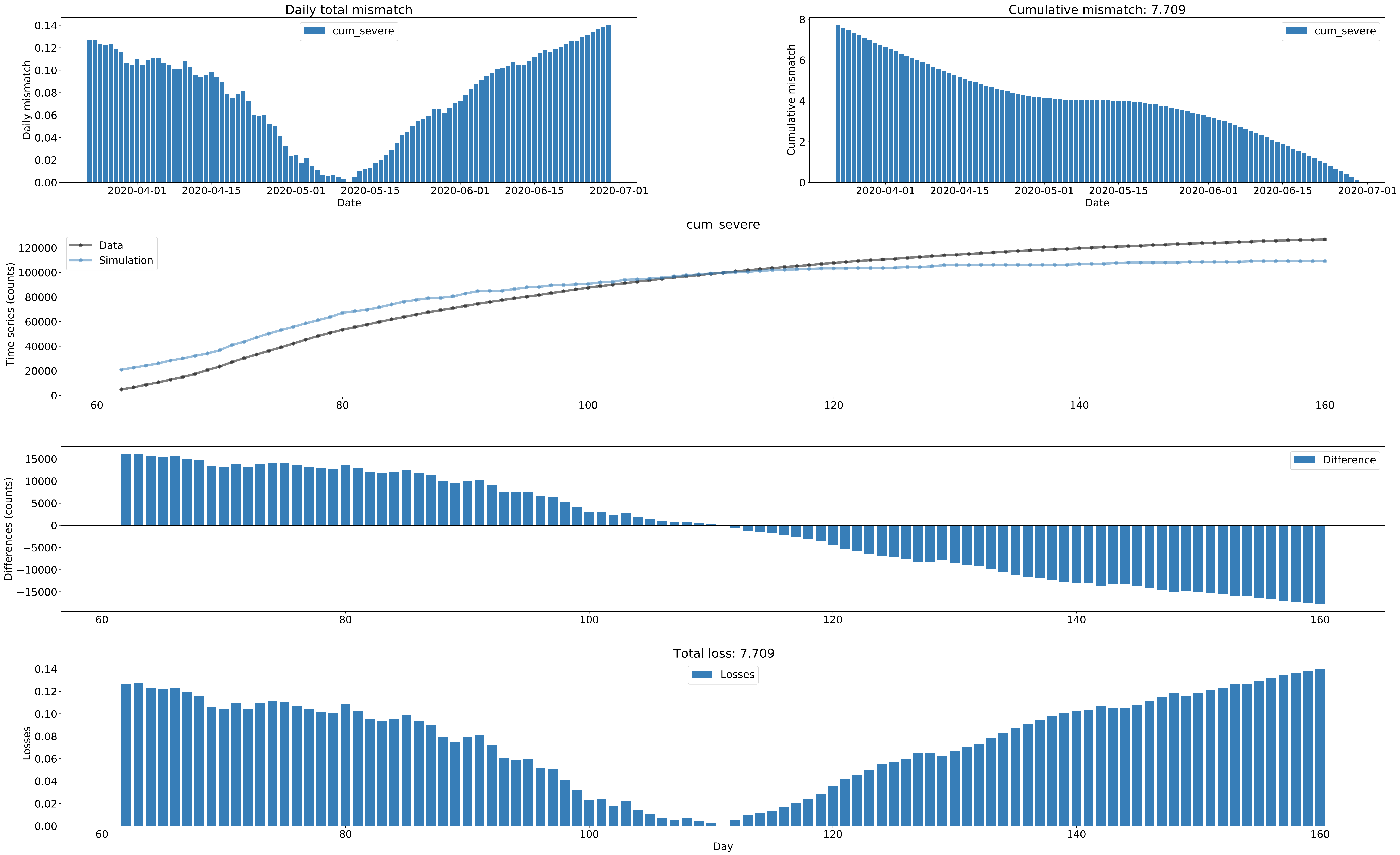
Note: all of these studies use the ‘cumAdmissions’ dataset from the UK Coronavirus website

Calibration

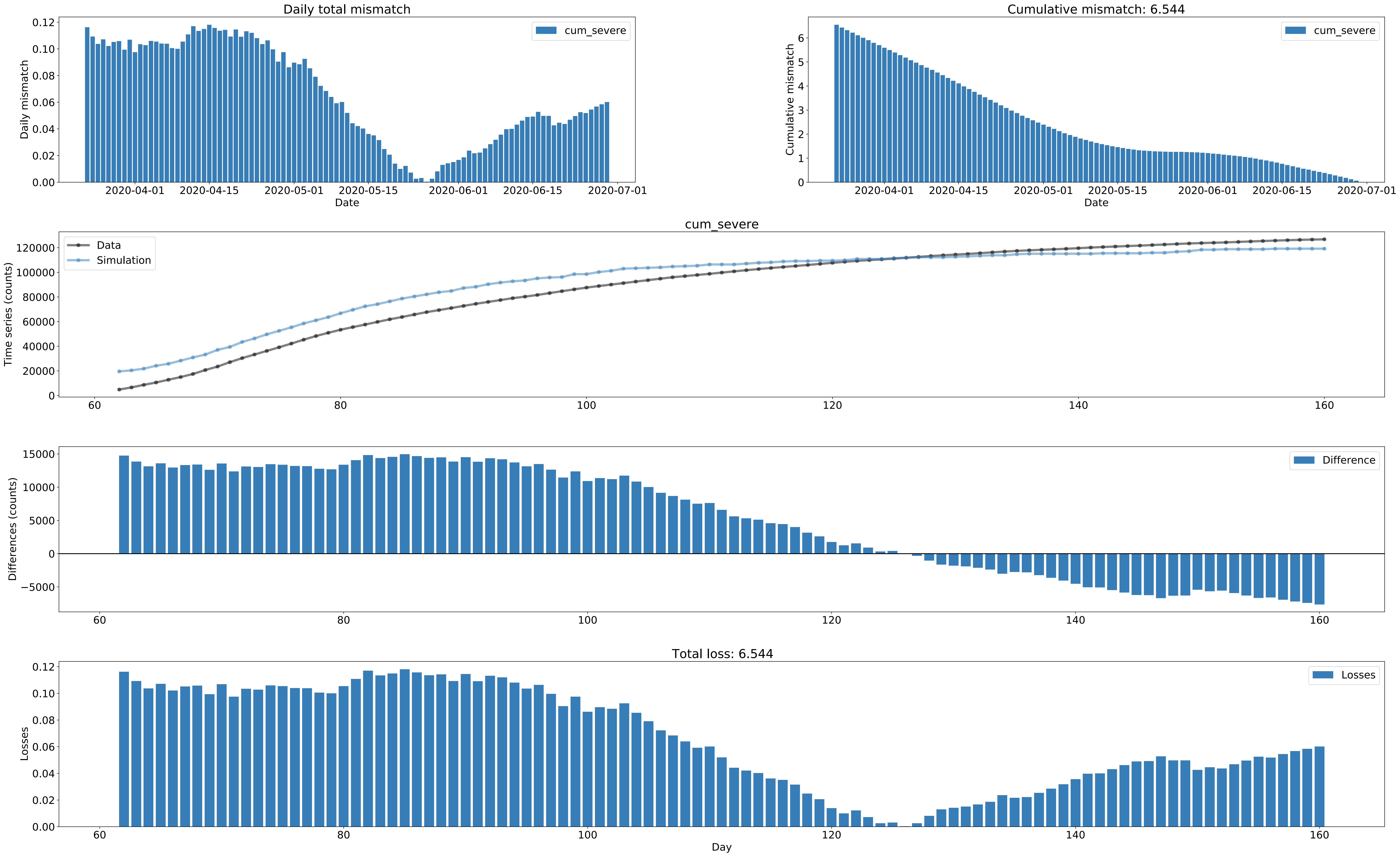
Method

- A data source is chosen (e.g cases, hospitalisations/admissions or deaths)
- Calibration
  - An Optuna trial is initialised with initial value(s) of the parameter(s)
  - A simulation is run using parameter value(s) selected by Optuna, assuming the interventions included in UK_Masks_TTI.py
  - Mismatch between the data and the simulation is calculated (Objective function)
  - The process is repeated using different values of the parameters for a predetermined number of Optuna trials
- A simulation is run using the optimal parameters found in the calibration
- Goodness of fit statistics are plotted to quantify the quality of the fit
- The Optuna study is saved as an SQL database

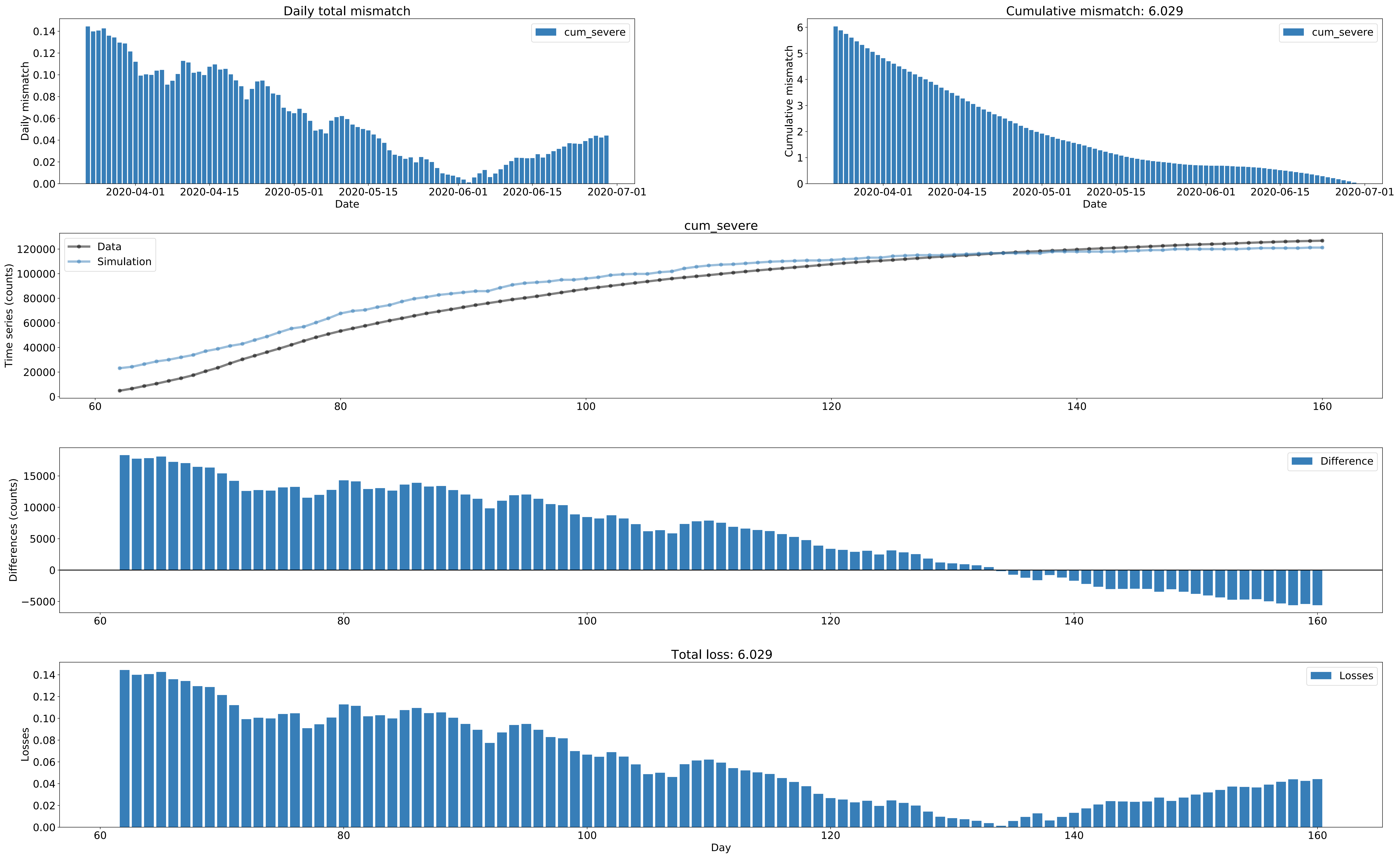
Fitting using β with 100 Optuna trials



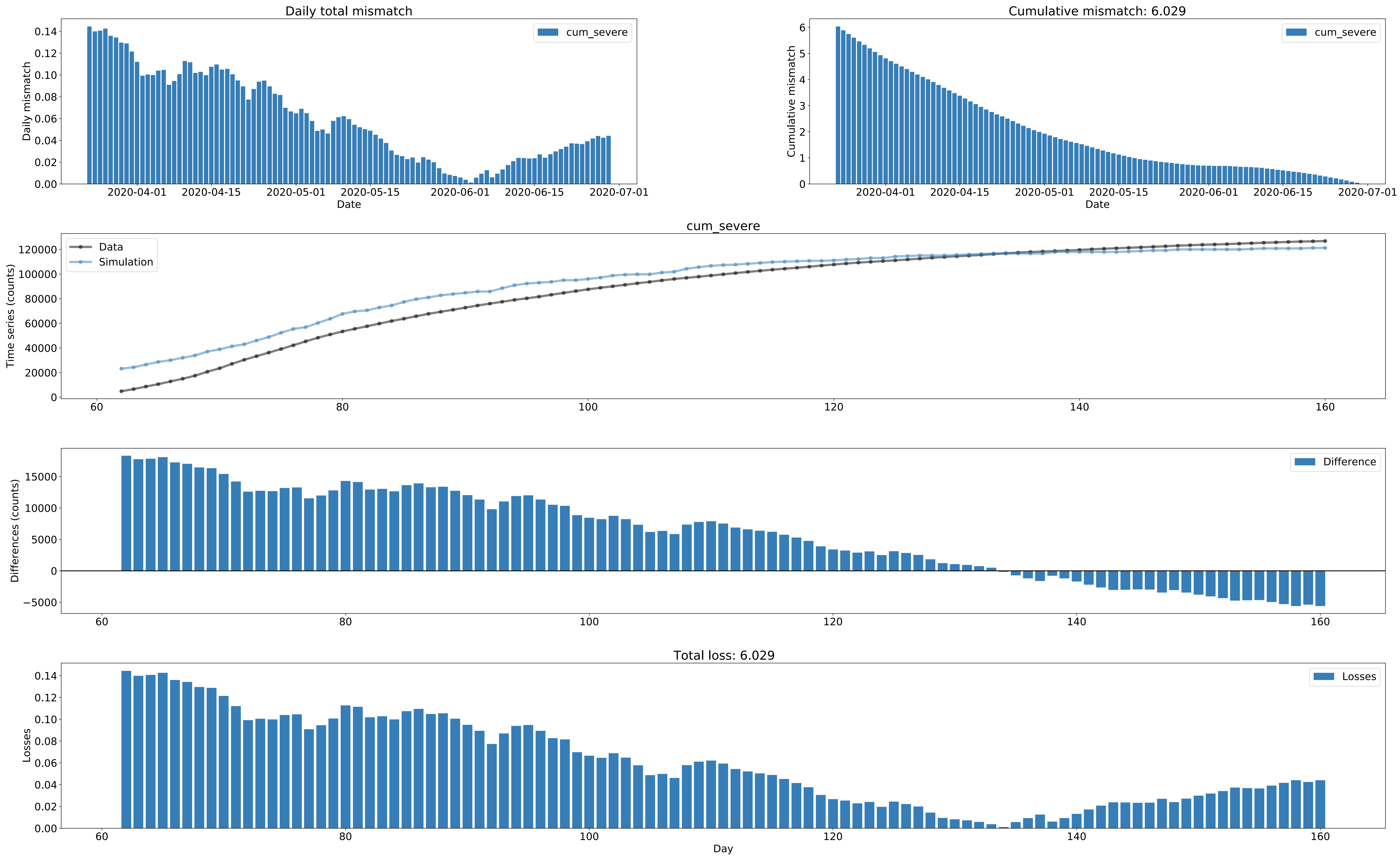
Fitting using β with 300 Optuna trials



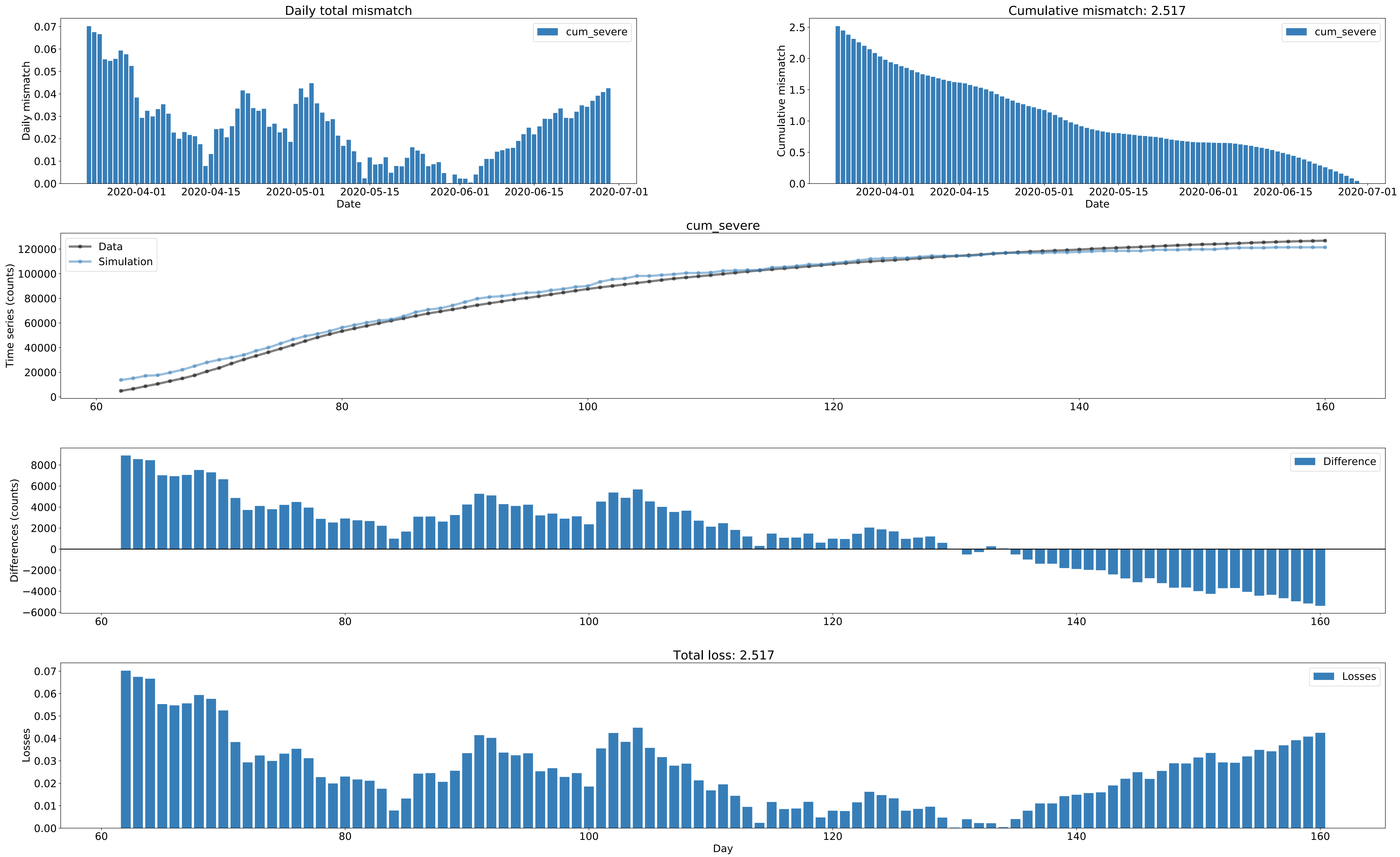
Fitting using β with 400 Optuna trials



Fitting using β with 700 Optuna trials



Fitting using β and pop_infected with 700 Optuna trials



Summary of results

№ Optuna trials	Time taken (s)	Optimal value of β	Optimal value of pop_infected	Total Loss
100	548	0.006704	N.A	7.709
300	1640	0.006876	N.A	6.544
400	2020	0.006921	N.A	6.029
700	4400	0.006921	N.A	6.029
700	4060	0.007515	1371.56	2.517

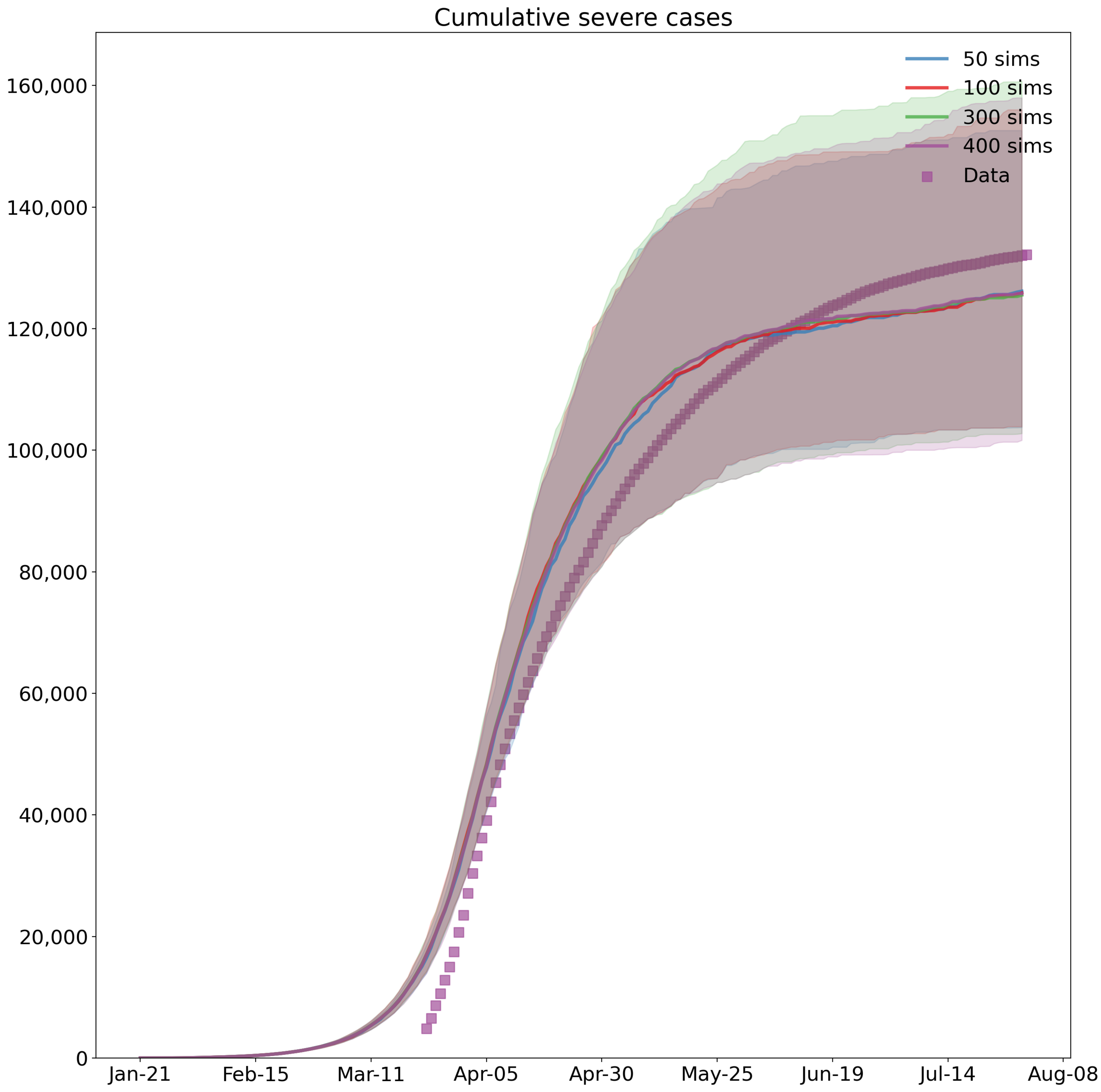
Comments

- Fit was significantly improved when optimising with β ***and*** *pop_infected* instead of only optimising with β
- When optimising with β only, the number of Optuna trials did not increase the quality of the fit beyond 400 Optuna trials (will re-run this)
- Simulation still doesn't match the trend of the data; the simulation begins above and ends below the curve
- I'm aware that all the *cum_severe* cases do not correspond to all the hospitalised cases

Simulation

Effect on the median simulation

- Might be a misunderstanding…
- Although the trend is clearly wrong, the number of simulations has little effect on the median simulation
- Will try to find a way to quantitatively compare the median simulations



Method - Local parameter search

- Ran a predetermined number of simulations using parameter values near to the optimal parameters i.e $\beta_{test} = \beta_{opt} \pm \varepsilon = 0.0075152767 \pm 0.0000001$
- Magnitude of ε was assigned randomly (ε = 0.0000005 x [random value between 1 and -1]) to form a disc of radius ε around the optimal value
- Value of *pop_infected* set to 1371.0 after seeing no significant change in the integer part
- No improvement in the loss seen after 138 trials

Going forward

- Find a way of fitting the first wave such that the trend of the data is reflected in the simulations
- Repeat first wave calibration using hospitalisations up to the end of August
- Create simulations based on the local parameter searches and calculate the median
- Make second wave projections
- Repeat process for the second wave
- Compare medians of the simulations run with the optimal parameters

Questions

- The trend of the simulations are not matching that of the data - what can I do to improve this?
- You mentioned that I should be able to fit the first wave by only using β - given that I could only improve the fit by fitting both beta and pop_infected, what am I doing wrong? How did you manage to do this?
- You mentioned that you've found a way of calibrating the model "by rejecting the priors that don't match what we already know with certainty" - what do you mean by this? Which priors? How are these priors related to the model?
- What is the maximum accuracy of β? Surely *pop_infected* must be an integer value in order to be physical?

Further questions

- I’ve calibrated the data assuming the interventions included in the UK_Masks_TTI.py script - could this be the reason why I can’t fit?

Applications

Applications

- A draft of CV and Imperial statement will be with you shortly if not already; I’d be grateful for your feedback ASAP