# Using Approximate Bayesian Computation to Infer Disease Parameter Uncertainty in a COVID-19 Microsimulation*

**Abstract.** A particular challenge with individual-level models, that often arises as a consequence of their necessarily stochastic nature, is in how to quantify uncertainty in model outcomes. This is particularly problematic when models are used for policy development as it is vital that forecast uncertainties are properly understood and communicated to policy makers. In this paper, the technique of Approximate Bayesian Computation (ABC) is used to dynamically infer the parameters of an individual-level COVID-19 simulation and to use the parameter posteriors as a means of quantifying the uncertainty in model forecasts. This not only reveals potentially useful insight into the evolution of the disease by also points to the (in)ability of the model to make predictions under high uncertainty.

**Keywords:** COVID-19 · calibration · Approximate Bayesian Computation · individual-based modelling.

## 1 Introduction

The COVID-19 pandemic forced policymakers and the public to place ever greater trust in modelling (7) and individual-based models, in particular, were instrumental in the formation of many government policies (e.g 4). However, properly understanding, quantifying and communicating uncertainty in the forecasts of individual-level models can be a challenging. It is possible to reduce some forecast uncertainty by including newly-emerging observational data (i.e. data that were not available when the model was initially calibrated) as they arise.

This paper adapts an existing, large-scale, individual-based COVID-19 model called the Dynamic Model for Epidemics (DyME: 12) and applies the technique of Approximate Bayesian Computation (2; 6; 13) to: (i) recalibrate the model dynamically as new COVID-19 positive test data become available; (ii) explore the parameter uncertainty that varies as the disease evolves; and (iii) produce forecasts that reflect the uncertainty in the model parameter estimates. The key findings are: (i) we potentially learn something about the evolution of the COVID-19 pandemic by observing the evolving parameter distributions produced by ABC; and (ii) forecast uncertainty reduces as new case data are included in the model calibration. These are important outcomes as uncertainty in

* This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 757455).

model predictions is often overlooked when models are used in policy (11) and it demonstrates that the is value in combining ABC with dynamic individual-level models.

## 2 Methods

The Dynamic Model for Epidemics (DyME) is a spatial microsimulation developed in 2020-21 in response to COVID-19 (12). It creates a synthetic representation of all people and households in a study area – in this case ∼800,000 individuals in Devon, UK – and combines methods from epidemiological modelling, spatial modelling, synthetic population generation and dynamic microsimulation. A model iteration corresponds to a simulated day, with individuals able to visit shops, schools and workplaces, as well as spending time with the rest of their household at home. The risk of an individual contracting the disease is proportional to the amount of time they spend at a particular location (school, workplace, shop, home) and the *hazard* associated with that location. The hazard is proportional to the number of infectious people who visit the place and the amount of time they spend there.

Data used in the study derive from the number of positive PCR tests recorded by the UK Government from April 2020 onwards. Multipliers created by (8) were used to scale-up the raw test data to account the large number of infections that go un-reported due to asymptomatic infections and a lack of testing capacity. In addition, the positive cases were lagged by six days so that they are more likely to represent the point of *infection*, rather than the PCR test, which is, after all, what the model simulates. The source data have been permanently archived[1]; for full details see (1).

There are a large number of model parameters that must be optimised for the model to accurately recreate the observed case data. Here we use Approximate Bayesian Computation (ABC) for that purpose. ABC has advantages over more commonly-used calibration approaches because it provides an assessment of the uncertainty in each parameter estimate (5). We re-calibrate the model a number of times as new case data emerge which potentially reveals new information about how the model is responding to the new data (and, hence, how the pandemic may be evolving in the real world). The general problem that ABC attempts to solve is the calculation of a posterior distribution, $\pi(\theta \mid Y)$, of a set of model parameters, $\theta$, given some observed data, $Y$, using the likelihood, $L(Y \mid \theta)$, and a prior estimate of the parameters, $\pi(\theta)$:

$$\pi(\theta \mid Y) \propto L(Y \mid \theta)\pi(\theta). \tag{1}$$

In practice the likelihood function is often intractable, so ABC uses a 'forward model' (here the DyME simulation) to explore combinations of parameters that create data that are sufficient similar to the real data. Naive sampling methods that search the parameter space evenly can be extremely inefficient for

[1] `https://dx.doi.org/10.17605/OSF.IO/QZW6F`

ABC (13; 3), so Monte-Carlo methods have been proposed as a means of reducing the number of rejected model runs (i.e. those that do not produce data that are sufficiently similar to the observations). Here, Sequential Monte Carlo (SMC: 10), also known as 'particle filtering' (13), is used as implemented using the Python package `pyabc` (9). The outcome of the ABC algorithm is a joint posterior distribution over the parameters. By drawing parameter values from that joint distribution it is possible to create probabilistic forecasts of the future disease evolution. In addition, examining the parameter distributions can reveal information about how the disease has evolved. For further detail about ABC, the interested reader can refer to Turner and Zandt's excellent tutorial (13).

The code used to run the experiments is available on GitHub[2]. The DyME model was implemented using OpenCL which reduced the run-time to a few seconds on a typical computer. Nevertheless, tens of thousands of individual model runs can be required to generate a reliable posterior so the time taken to evaluate a complete experiment ranges from 24 to 48 hours.

## 3 Results

Due to space constraints, here we present just the analysis of the evolving parameter posteriors, not the probabilistic forecasts that can be created by drawing from the joint posterior distribution.

Every 14 days the model is re-calibrated using ABC drawing on the most recent 14 days of positive COVID-19 test data. Figure 1 illustrates how the posterior distributions (Equation 1) vary over each 14 day 'window'. The first four parameters (`retail`, `primary_school`, `secondary_school` and `work`) are multipliers that influence how hazardous those particular locations are. The remaining three (`presymptomatic`, `symptomatic` and `asymptomatic`) are multipliers that are applied to the three different disease states. Interestingly the posteriors suggest that the exposure risks associated with visiting shops or primary schools, as well as the hazards associated with symptomatic transmission, appeared to increase over time. This may tell us something about how society is responding to the disease (e.g. people may become less transmission-conscious in shops as the disease evolves) or point to the influence of government policies (e.g. shutting schools entirely or mandating mask wearing for some age groups). For other parameters, such as `work` or `asymptomatic`, the distribution changes little over the course of the experiment. Finally it is also possible to observe that some parameters are much more certain than others; e.g. there is a very small range of asymptomatic transmission that causes the model to perform accurately, whereas values associated with schools and retail are much less certain.

[2] `https://github.com/Urban-Analytics/RAMP-UA/tree/master/experiments/calibration`

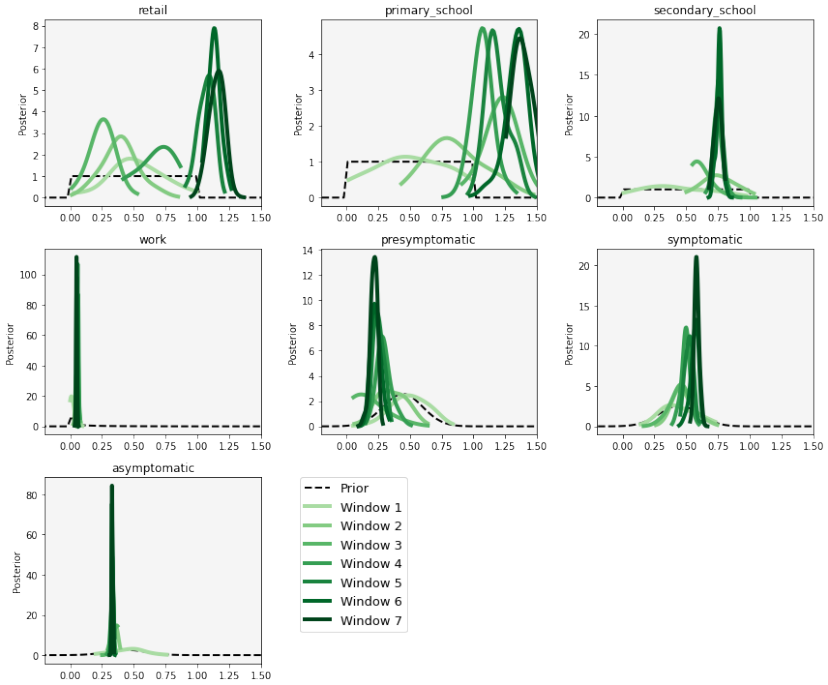


**Fig. 1.** Posterior distribution of each parameter value at the end of each data calibration window, with the prior parameter distributions shown by the black dotted line

## 4 Conclusion

This paper demonstrates how a microsimulation (the DyME COVID-19 model) can be dynamically re-calibrated using Approximate Bayesian Computation (ABC) as new infection data become available. During each 14-day window, the model is re-calibrated to include the new case data that have emerged over that period. Unlike traditional one-shot calibration methods that use a single set of optimal parameters, this method retains the inherent parameter uncertainty that is expressed in the posterior distributions. It is important to note, however, that we only use a single optimisation algorithm (ABC-SMC) and, as (3) demonstrate, alternative algorithms may perform better.

Figure 1 illustrated that some parameters change distinctly as the disease evolves, but others are relatively stable. In addition, some parameters exhibit much greater uncertainty than others. Although these preliminary results are not yet sufficiently robust to draw firm conclusions about true disease dynamics, they do demonstrate that producing dynamic parameter posteriors might reveal useful information about the propagation of the disease, and society's evolving reaction to it.

# Bibliography


[1] Asher, M., Lomax, N., Morrissey, K., Spooner, F., Malleson, N.: Dynamic calibration with approximate bayesian computation for a microsimulation of disease spread. Nature Scientific Reports (2013), under review

[2] Beaumont, M.A., Zhang, W., Balding, D.J.: Approximate bayesian computation in population genetics. Genetics **162**(4), 2025–2035 (2002)

[3] Carrella, E.: No Free Lunch when Estimating Simulation Parameters. Journal of Artificial Societies and Social Simulation **24**(2), 7 (2021). https://doi.org/10.18564/jasss.4572

[4] Ferguson, N., Laydon, D., Nedjati Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunuba Perez, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L., Van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P., Winskill, P., Whittaker, C., Donnelly, C., Riley, S., Ghani, A.: Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Tech. rep., Imperial College London (2020). https://doi.org/10.25561/77482

[5] Grazzini, J., Richiardi, M.G., Tsionas, M.: Bayesian estimation of agent-based models. Journal of Economic Dynamics and Control **77**, 26–47 (2017). https://doi.org/10.1016/j.jedc.2017.01.014

[6] Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S.: Markov chain monte carlo without likelihoods. Proceedings of the National Academy of Sciences **100**(26), 15324–15328 (2003)

[7] McBryde, E.S., Meehan, M.T., Adegboye, O.A., Adekunle, A.I., Caldwell, J.M., Pak, A., Rojas, D.P., Williams, B.M., Trauer, J.M.: Role of modelling in COVID-19 policy development. Paediatric Respiratory Reviews **35**, 57–60 (2020). https://doi.org/10.1016/j.prrv.2020.06.013

[8] Phipps, S.J., Grafton, R.Q., Kompas, T.: Robust estimates of the true (population) infection rate for covid-19: a backcasting approach. Royal Society Open Science **7**(11), 200909 (2020)

[9] Schälte, Y., Klinger, E., Alamoudi, E., Hasenauer, J.: pyabc: Efficient and robust easy-to-use approximate bayesian computation. arXiv (2022). https://doi.org/10.48550/arxiv.2203.13043, `https://arxiv.org/abs/2203.13043`

[10] Sisson, S.A., Fan, Y., Tanaka, M.M.: Sequential monte carlo without likelihoods. Proceedings of the National Academy of Sciences **104**(6), 1760–1765 (2007)

[11] Spiegelhalter, D.: Risk and Uncertainty Communication. Annual Review of Statistics and Its Application **4**(1), 31–60 (2017). https://doi.org/10.1146/annurev-statistics-010814-020148

[12] Spooner, F., Abrams, J.F., Morrissey, K., Shaddick, G., Batty, M., Milton, R., Dennett, A., Lomax, N., Malleson, N., Nelissen, N., Coleman, A., Nur,

J., Jin, Y., Greig, R., Shenton, C., Birkin, M.: A dynamic microsimulation model for epidemics. Social Science & Medicine **291**, 114461 (2021). https://doi.org/10.1016/j.socscimed.2021.114461

[13] Turner, B.M., Van Zandt, T.: A tutorial on approximate Bayesian computation. Journal of Mathematical Psychology **56**(2), 69–85 (2012). https://doi.org/10.1016/j.jmp.2012.02.005