BAYESIAN CALIBRATION OF DIFFERENTIABLE AGENT-BASED MODELS

Arnau Quera-Bofarull
Department of Computer Science
University of Oxford
`arnau.quera-bofarull@cs.ox.ac.uk`

Ayush Chopra
Media Lab
Massachusetts Institute of Technology
`ayushc@mit.edu`

Anisoara Calinescu
Department of Computer Science
University of Oxford
`ani.calinescu@cs.ox.ac.uk`

Michael Wooldridge
Department of Computer Science
University of Oxford
`mjw@cs.ox.ac.uk`

Joel Dyer
Department of Computer Science & Institute for New Economic Thinking
University of Oxford
`joel.dyer@cs.ox.ac.uk`

ABSTRACT

Agent-based modelling (ABMing) is a powerful and intuitive approach to modelling complex systems; however, the intractability of ABMs' likelihood functions and the non-differentiability of the mathematical operations comprising these models present a challenge to their use in the real world. These difficulties have in turn generated research on approximate Bayesian inference methods for ABMs and on constructing differentiable approximations to arbitrary ABMs, but little work has been directed towards designing approximate Bayesian inference techniques for the specific case of differentiable ABMs. In this work, we aim to address this gap and discuss how generalised variational inference procedures may be employed to provide misspecification-robust Bayesian parameter inferences for differentiable ABMs. We demonstrate with experiments on a differentiable ABM of the COVID-19 pandemic that our approach can result in accurate inferences, and discuss avenues for future work.

1 INTRODUCTION

Agent-based models (ABMs) are growing in popularity as a modelling paradigm for complex systems in various fields, such as economics (Baptista et al., 2016; Paulin et al., 2019) and epidemiology (Aylett-Bullock et al., 2021). Such models simulate the interactions and decisions of a set of autonomous entities, where the rules governing those decisions and interactions are often nonlinear and stochastic.

While this modelling approach provides considerable flexibility to the modeller, the complex structure and stochastic nature of many ABMs raise two key difficulties in deploying them in practice:

- ABMs typically lack a tractable likelihood function – denoted with $p(\mathbf{x} \mid \boldsymbol{\theta})$, where $\mathbf{x}$ is the model output and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ are the d-dimensional free parameters of the model – which complicates the problem of calibrating the model's free parameters;
- the mathematical expressions comprising the ABM's specification are typically non-differentiable, which presents a barrier to the use of gradient-based methods in problems such as model calibration.

Over recent years, two active lines of research have emerged that seek to address each of these problems: a growing literature on approximate parameter inference techniques for ABMs, which has

become increasingly focused on Bayesian methods (see e.g. Grazzini et al., 2017; Platt, 2021; Dyer et al., 2022a); and the development of techniques for building differentiable approximations to initially non-differentiable ABMs (Chopra et al., 2022b; Monti et al., 2022). However, the combination of Bayesian parameter calibration methods and differentiable ABMs has not yet been considered.

In this paper, we examine the problem of performing approximate Bayesian parameter inference for differentiable ABMs, and consider an approach that exploits the differentiability of the differentiable agent-based simulator. We discuss how the approach we consider – which is derived from the literature on generalised Bayesian inference (see e.g. Bissiri et al., 2016; Knoblauch et al., 2022) – may enjoy favourable robustness properties in comparison to alternative techniques, enabling the differentiable ABM to be applied more successfully in misspecified settings.

2   BACKGROUND

2.1   SIMULATION-BASED BAYESIAN INFERENCE FOR AGENT-BASED MODELS

Simulation-based inference (SBI) algorithms are a set of procedures for performing parameter inference for simulation models, such as ABMs. Bayesian approaches to SBI for ABMs have gained popularity over recent years with the use of approximate Bayesian computation (ABC) (Tavaré et al., 1997; Pritchard et al., 1999; Beaumont et al., 2002), see e.g. Grazzini et al. (2017); van der Vaart et al. (2015). Broadly speaking, ABC targets an approximate posterior q_{ABC} of the form

$$q_{\text{ABC}}(\boldsymbol{\theta} \mid \mathbf{s_y}) \propto \hat{p}_{\text{ABC}}(\mathbf{s_y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta}), \tag{1}$$

where $\mathbf{s_x}$ is some summary statistic of $\mathbf{x}$, $\hat{p}_{\text{ABC}}(\mathbf{s_y} \mid \boldsymbol{\theta})$ is some approximation to the model's likelihood function, and π is a prior density over $\boldsymbol{\Theta}$. A variety of choices for $\hat{p}_{\text{ABC}}(\mathbf{s_y} \mid \boldsymbol{\theta})$ have been explored in the literature; see Grazzini et al. (2017) and Platt (2021) for examples and Dyer et al. (2022a) for a broader review.

Beyond ABC, Dyer et al. (2022a) introduces and motivates the use of a class of more recently developed SBI procedures. These approaches offer considerable benefits, such as a vastly reduced simulation burden compared to alternative techniques, and the ability to flexibly accommodate data of different kinds (e.g. dynamic graph data, see Dyer et al. (2022b)) – but their performance can deteriorate in misspecified settings (Cannon et al., 2022), motivating research into methods for improving their robustness (Ward et al., 2022; Kelly et al., 2023).

2.2   DIFFERENTIABLE AGENT-BASED MODELS

Differentiable agent-based models (GRADABMs) are a recent class of tensorized ABMs that offer a number of benefits, such as compatibility with gradient-based learning via automatic differentiation, and the ability to simulate million-size populations in a few seconds on commodity hardware, integrate with deep neural networks, and ingest heterogeneous data sources. Recent prior work has demonstrated the utility of GRADABMs for scalable and fast forward simulations (Chopra et al., 2021), end-to-end calibration by coupling with neural networks (Chopra et al., 2022a), as well as efficient validation using gradient-based sensitivity analyses (Quera-Bofarull et al., 2023). Direct access to simulator gradients in GRADABMs allows the modeller to leverage a statistical model for calibration without the need to build approximate surrogates. While this has shown to help integrate diverse data sources to enable more accurate calibration, prior work has only leveraged such gradients to generate point estimates of parameters. However, maximising the real-world utility of GRADABMs will also require uncertainty quantification during model calibration tasks; this motivates the use of Bayesian approaches for model calibration.

3   METHOD

To perform approximate Bayesian inference for a differentiable agent-based simulator, we consider a simulation-based Bayesian inference procedure with appealing robustness properties. In particular, we find a posterior density q as the solution to a generalised variational inference (GVI) problem, which has been shown to possess favourable robustness properties in comparison to methods that target the classical Bayesian posterior (Knoblauch et al., 2022). This enables us to obtain practically
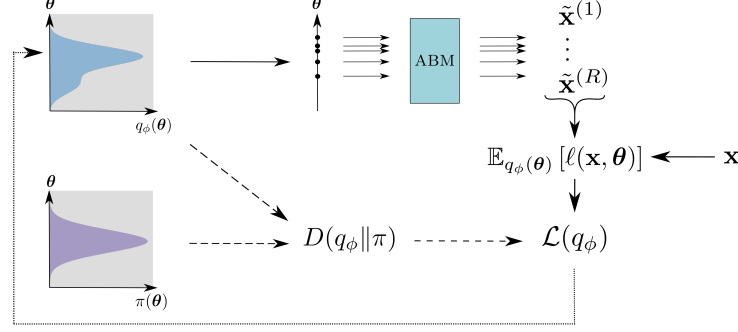
Figure 1: A schematic of the method we employ. The posterior density estimator q_ϕ is encouraged to remain similar to the prior π through some divergence D (dashed arrows), while also being encouraged to generate simulations from the ABM that closely match the data $\mathbf{x}$ to which the model is being calibrated (solid arrows). The overall loss is a linear combination of these contributing terms and is used to inform the shape of q_ϕ (dotted arrow).

useful updates to the initial belief distribution, captured by the prior density π, when the agent-based model is not a perfect representation of the true data-generating mechanism (as is the case in reality).

GVI proceeds by first establishing a triple consisting of: a lower-bounded loss function/scoring rule $\ell : \mathcal{X} \times \mathbf{\Theta} \to \mathbb{R}_{\geq 0}$ capturing some notion of discrepancy between the observed data $\mathbf{x}$ and the behaviour of the simulator at parameter value $\boldsymbol{\theta}$; a divergence D between the posterior q and prior π; and a search space $\mathcal{Q}$ of permitted solutions for q. Bayesian inference is then performed by solving the following optimisation problem:

$$q^* = \arg\min_{q \in \mathcal{Q}} \mathcal{L}(q), \qquad \mathcal{L}(q) := \mathbb{E}_{q(\boldsymbol{\theta})}\left[\ell(\mathbf{x}, \boldsymbol{\theta})\right] + D\left(q \| \pi\right). \tag{2}$$

The solution q^* is the generalised Bayesian posterior, while the *classical* Bayesian posterior is obtained with $\ell(\mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{x} \mid \boldsymbol{\theta})$ and D as the Kullback-Liebler divergence D_{KL}.

To construct a flexible class of feasible solutions, we take $\mathcal{Q} = \{q_\phi : \phi \in \Phi\}$, where q_ϕ is a normalising flow with trainable parameters ϕ taking values in some set Φ. The flow parameters are then found as

$$\phi^* = \arg\min_{\phi \in \Phi} \left\{ \mathbb{E}_{q_\phi(\boldsymbol{\theta})}\left[\ell(\mathbf{x}, \boldsymbol{\theta})\right] + D\left(q_\phi \| \pi\right) \right\}. \tag{3}$$

In this way – owing to the ability to backpropagate pathwise gradients through the GRADABM – we can use the misspecification-robust inference framework of GVI to calibrate the model parameters without recourse to potentially high-variance score-based gradient estimators (Mohamed et al., 2020). A schematic of our method is shown in Figure 1.

4 EXPERIMENTS

We evaluate the proposed inference procedure by calibrating GRADABM-JUNE (Quera-Bofarull, 2023), the differentiable version of the JUNE model (Aylett-Bullock et al., 2021). We use GRADABM-JUNE to model the spread of SARS-CoV-2 in the London Borough of Camden, comprising approximately 250,000 people. For this experiment, we restrict ourselves to modelling the spread of infection in households, companies, and schools, each of which is regulated by the β-parameters that parameterize the spread of infection at different locations (see Appendix A.1 for details).

We create a synthetic ground-truth time-series of SARS-CoV-2 infections for 30 days using the values $\beta_{\mathrm{household}} = 0.9$, $\beta_{\mathrm{school}} = 0.6$, and $\beta_{\mathrm{company}} = 0.3$,

and obtain a posterior density by training a Neural Spline Flow (NSF) (Durkan et al., 2019). We take

$$\ell\left(\mathbf{x}, \boldsymbol{\theta}\right) = \mathbb{E}_{\tilde{\mathbf{x}} \sim p(\cdot|\boldsymbol{\theta})}\left[\sum_{t=1}^{T} \frac{\|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^2}{w}\right] \quad (4)$$

in the loss function (Equation 2), where $\mathbf{x}_t$ is the logarithm of the number of infections per time-step and the hyperparameter $w > 0$ balances the relative influence of the scoring rule to the divergence-measuring term. We further choose $D = D_{\mathrm{KL}}$ for simplicity. Details of the neural network architecture and additional training hyperparameters are provided in Appendix A.2.

In this way, we obtain a posterior density over $\boldsymbol{\theta}$, which we show in Figure 2. We see that the flow assigns high posterior density to the generating parameters, suggesting that the flow has assigned posterior mass to appropriate regions of $\boldsymbol{\Theta}$. Of particular interest is the 2-dimensional projection of the density in the household-company and school-company plane, where we observe a trade-off between the contact intensity at households and companies. Indeed, it may be hard to distinguish where exactly infections are taking place when only the *overall* number of cases is observed.

Figure 2: The inferred posterior distribution over the three calibrated β parameters ($\beta_{\mathrm{household}}$, β_{school} and β_{company}). The marginal densities are shown on the diagonal, and the off-diagonals show the bivariate joint densities for all pairs of parameters. The parameters that generated the pseudo-true synthetic dataset are shown with blue lines and points.

We also show in Figure 3 a comparison between the pseudo-true dataset used to calibrate the model and simulations from the ABM generated by parameters drawn from the prior, the untrained flow, and the trained flow. From this, we see that the trained flow generates simulations that much better match the pseudo-true data than simulations generated by the prior or untrained flows, suggesting that our inference scheme has been successful in this model calibration task. Overall, the number of simulations required to train the flow was 2,500, which is small in comparison to e.g. ABC.

5 DISCUSSION & CONCLUSION

In this paper, we consider how the task of Bayesian parameter calibration may be performed for differentiable ABMs. We discuss how the ability to backpropagate gradients through the agent-based simulator in a pathwise manner provides us with immediate access to a class of Bayesian
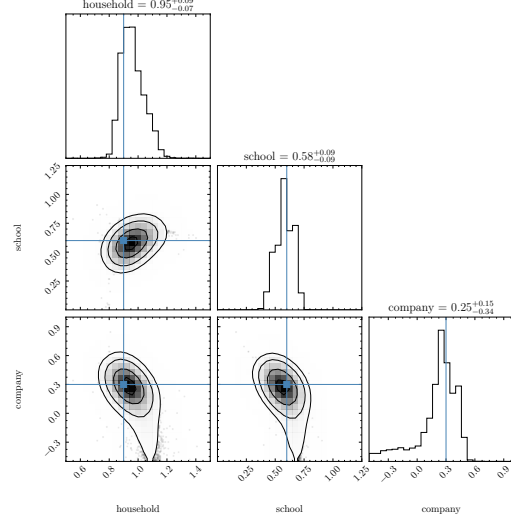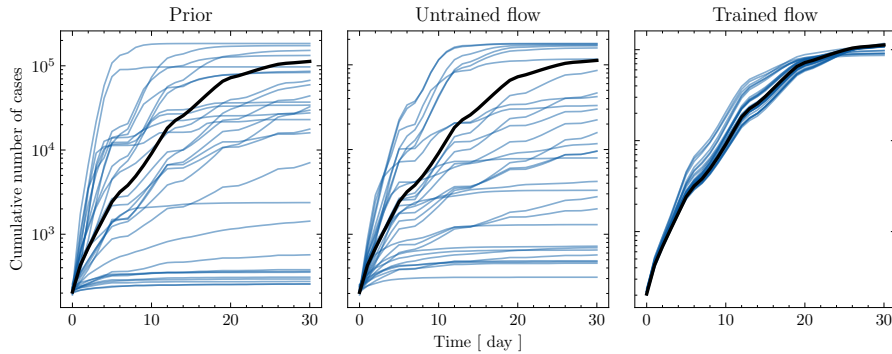

Figure 3: A comparison of the pseudo-true dataset (black curve) to simulations (blue curves) generated by parameters drawn from the prior density (**left**), the untrained normalising flow (**middle**), and the trained normalising flow (**right**).

4

inference methods known as generalised variational inference, and propose an approach drawn from this class of methods due to the fact that they may remedy misspecification-related problems more readily than existing approximate Bayesian inference methods for ABMs. Through experiments with GRADABM-JUNE, a differentiable ABM of the COVID-19 pandemic in England, we demonstrate that our approach can provide accurate Bayesian inferences. We aim to develop this work into a full paper, at which point we will release the code for reproducing these results.

In future, we will test this method on real-world data and compare its performance against alternative SBI techniques. We will also extend this work to the case of multiple observed *iid* datasets $\mathbf{x}$ from some real-world density $p(\mathbf{x})$ by considering the case of a conditional density estimator (e.g. a conditional normalising flow) and by training instead on the following loss function:

$$q^* = \arg\min_{q \in \mathcal{Q}} \left\{ \mathbb{E}_{q(\boldsymbol{\theta}|\mathbf{x})p(\mathbf{x})}\left[\ell(\mathbf{x}, \boldsymbol{\theta})\right] + \mathbb{E}_{p(\mathbf{x})}\left[D\left(q(\cdot \mid \mathbf{x}) \| \pi(\cdot)\right)\right] \right\}. \tag{5}$$

Once again, the choice $\ell(\mathbf{x}, \boldsymbol{\theta}) = -\log p(\mathbf{x} \mid \boldsymbol{\theta})$ and $D = D_{\mathrm{KL}}$ will yield classical Bayesian posteriors, while other choices generate generalised posteriors. This may enable us to deploy the same conditional density estimator and ABM over a variety of scenarios covered by the density $p(\mathbf{x})$, without retraining.

REFERENCES

Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Arnau Quera-Bofarull, Miguel Icaza-Lizaola, Aidan Sedgewick, Henry Truong, Aoife Curran, Edward Elliott, Tristan Caulfield, Kevin Fong, et al. June: open-source individual-based epidemiology simulation. *Royal Society open science*, 8(7):210506, 2021.

Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Arnau Quera-Bofarull, Anjali Katta, Katherine Hoffmann Pham, Benjamin Hoover, Hendrik Strobelt, Rebeca Moreno Jimenez, Aidan Sedgewick, Egmond Samir Evers, David Kennedy, Sandra Harlass, Allen Gidraf Kahindo Maina, Ahmad Hussien, and Miguel Luengo-Oroz. Operational response simulation tool for epidemics within refugee and IDP settlements: A scenario-based case study of the Cox's Bazar settlement. *PLOS Computational Biology*, 17(10):e1009360, October 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009360.

Rafa Baptista, J Doyne Farmer, Marc Hinterschweiger, Katie Low, Daniel Tang, and Arzu Uluc. Macroprudential policy in an agent-based model of the UK housing market. *Bank of England Working Paper*, 2016.

Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

Pier Giovanni Bissiri, Chris Holmes, and Stephen G Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(5):1103, 2016.

Patrick Cannon, Daniel Ward, and Sebastian M Schmon. Investigating the impact of model misspecification in neural simulation-based inference. *arXiv preprint arXiv:2209.01845*, 2022.

Ayush Chopra, Ramesh Raskar, Jayakumar Subramanian, Balaji Krishnamurthy, Esma S Gel, Santiago Romero-Brufau, Kalyan S Pasupathy, and Thomas C Kingsley. Deepabm: scalable and efficient agent-based simulations via geometric learning frameworks-a case study for covid-19 spread and interventions. In *2021 Winter Simulation Conference (WSC)*, pp. 1–12. IEEE, 2021.

Ayush Chopra, Alexander Rodriguez, Jayakumar Subramanian, Balaji Krishnamurthy, B Aditya Prakash, and Ramesh Raskar. Differentiable Agent-based Epidemiology. *AAMAS 2023*, 2022a.

Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Balaji Krishnamurthy, B Aditya Prakash, and Ramesh Raskar. Differentiable Agent-based Epidemiology. *arXiv preprint arXiv:2207.09714*, 2022b.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL `https://proceedings.neurips.cc/paper/2019/file/7ac71d433f282034e088473244df8c02-Paper.pdf`.

Joel Dyer, Patrick Cannon, J Doyne Farmer, and Sebastian Schmon. Black-box Bayesian inference for economic agent-based models. *arXiv preprint arXiv:2202.00625*, 2022a.

Joel Dyer, Patrick Cannon, J Doyne Farmer, and Sebastian M Schmon. Calibrating agent-based models to microdata with graph neural networks. *arXiv preprint arXiv:2206.07570*, 2022b.

Jakob Grazzini, Matteo G Richiardi, and Mike Tsionas. Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, 77:26–47, 2017.

Ryan P Kelly, David J Nott, David T Frazier, David J Warne, and Chris Drovandi. Misspecification-robust Sequential Neural Likelihood. *arXiv preprint arXiv:2301.13368*, 2023.

Jeremias Knoblauch, Jack Jewson, and Theodoros Damoulas. An optimization-centric view on bayes' rule: Reviewing and generalizing variational inference. *Journal of Machine Learning Research*, 23(132):1–109, 2022.

Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte carlo gradient estimation in machine learning. *The Journal of Machine Learning Research*, 21(1):5183–5244, 2020.

Corrado Monti, Marco Pangallo, Gianmarco De Francisci Morales, and Francesco Bonchi. On learning agent-based models from data. *arXiv preprint arXiv:2205.05052*, 2022.

James Paulin, Anisoara Calinescu, and Michael Wooldridge. Understanding flash crash contagion and systemic risk: A micro–macro agent-based approach. *Journal of Economic Dynamics and Control*, 100:200–229, 2019.

Donovan Platt. Bayesian estimation of economic simulation models using neural networks. *Computational Economics*, pp. 1–52, 2021.

Jonathan K Pritchard, Mark T Seielstad, Anna Perez-Lezaun, and Marcus W Feldman. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular biology and evolution*, 16(12):1791–1798, 1999.

A Quera-Bofarull, A Chopra, J Aylett-Bullock, C Cuesta-Lazaro, A Calinescu, R Raskar, and M Wooldridge. Don't simulate twice: one-shot sensitivity analyses via automatic differentiation. Association for Computing Machinery, 2023.

Arnau Quera-Bofarull. arnauqb/gradabm-june: Aamas zenodo release, February 2023. URL `https://doi.org/10.5281/zenodo.7623959`.

François Rozet and Felix Divo. francois-rozet/zuko: Zuko 0.1.4, February 2023. URL `https://doi.org/10.5281/zenodo.7625673`.

Simon Tavaré, David J Balding, Robert C Griffiths, and Peter Donnelly. Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518, 1997.

Elske van der Vaart, Mark A Beaumont, Alice SA Johnston, and Richard M Sibly. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling*, 312:182–190, 2015.

I. Vernon, J. Owen, J. Aylett-Bullock, C. Cuesta-Lazaro, J. Frawley, A. Quera-Bofarull, A. Sedgewick, D. Shi, H. Truong, M. Turner, J. Walker, T. Caulfield, K. Fong, and F. Krauss. Bayesian emulation and history matching of JUNE. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2233):20220039, October 2022. doi: 10.1098/rsta.2022.0039.

Daniel Ward, Patrick Cannon, Mark Beaumont, Matteo Fasiolo, and Sebastian M Schmon. Robust Neural Posterior Estimation and Statistical Model Criticism. In *Advances in Neural Information Processing Systems*, 2022.

A APPENDIX

A.1 THE JUNE MODEL

The JUNE model (Aylett-Bullock et al., 2021) is a one-to-one agent-based epidemiological model that uses the English census to create a synthetic population at a very high resolution. The model has been used in a wide variety of settings such as to study the first and second wave of SARS-CoV2 in England (Vernon et al., 2022) and to mitigate disease spread in refugee settlements (Aylett-Bullock et al., 2021). The model has been ported to the GRADABM framework to allow for much faster performance and gradient-based calibration (Chopra et al., 2022b; Quera-Bofarull, 2023). JUNE has a wide variety of configurable parameters regarding disease transmission and progression, vaccination, and non-pharmaceutical interventions. Given a susceptible agent exposed to an infection at location L, the probability of that agent getting infected is given by

$$p = 1 - \exp\left(-\psi_s \ \beta_L \ \Delta t \sum_{i \in g} \mathcal{I}_i(t)\right) \tag{6}$$

where ψ_s is the inherent susceptibility to infection of the agent, the summation is over all contacts with infected individuals at the location, $\mathcal{I}_i(t)$ is the time-dependent infectious profile of each infected agent, Δt is the duration of the interaction, and, finally, β_L is a location-specific parameter that models the difference in the nature of interactions for each location. Since the parameters β_L are not physical parameters that can be measured, they are typically calibrated using data on the number of cases or fatalities over time. In this work, we focus on calibrating the β_L parameters for our three modelled locations: households, schools, and companies.

A.2 NEURAL NETWORK ARCHITECTURE AND TRAINING

To obtain a posterior density, we use a Neural Spline Flow (NSF) using ZUKO(Rozet & Divo, 2023) as q_ϕ, consisting of 3 transformations, each parameterized by a 3-layer fully connected neural network with 128 units in every layer. We did not explore the impact of different flow models or parameterisations on the calibration results. We train the NSF model following the procedure described in Section 3 until the validation loss consisting of the forecasting and regularisation term (see Equation 2) does not significantly change across epochs. At each epoch, we estimate

$$D_{\mathrm{KL}}(q_\phi \| \pi) \approx \frac{1}{R} \sum_{r=1}^{R} \left(\log q_\phi(\boldsymbol{\theta}^{(r)}) - \log \pi(\boldsymbol{\theta}^{(r)})\right),$$

where $\boldsymbol{\theta}^{(r)} \overset{iid}{\sim} q_\phi$ and $R = 10^4$. Likewise, the forecast loss (scoring rule) is estimated by following Equation 4 with $T = 10$. We obtain the best results by setting a regularisation weight of $w = 10^{-2}$.