Don't Simulate Twice: One-Shot Sensitivity Analyses via Automatic Differentiation

Arnau Quera-Bofarull
University of Oxford
Oxford, UK
arnau.quera-bofarull@cs.ox.ac.uk

Ayush Chopra
Massachusetts Institute of Technology
Cambridge, USA
ayushc@mit.edu

Joseph Aylett-Bullock
United Nations Global Pulse
New York, USA
joseph@unglobalpulse.org

Carolina Cuesta-Lazaro
Massachusetts Institute of Technology
Cambridge, USA
cuestalz@mit.edu

Anisoara Calinescu
University of Oxford
Oxford, UK
ani.calinescu@cs.ox.ac.uk

Ramesh Raskar
Massachusetts Institute of Technology
Cambridge, USA
raskar@mit.edu

Michael Wooldridge
University of Oxford
Oxford, UK
mike.wooldridge@cs.ox.ac.uk

ABSTRACT

Agent-based models (ABMs) are a promising tool to simulate complex environments. Their rapid adoption requires scalable specification, efficient data-driven calibration, and validation through sensitivity analyses. Recent progress in tensorized and differentiable ABM design (GradABM) has enabled fast calibration of million-size populations, however, validation through sensitivity analysis is still computationally prohibitive due to the need for running the model a large number of times. Here, we present a novel methodology that uses automatic differentiation to perform a sensitivity analysis on a calibrated ABM without requiring any further simulations. The key insight is to leverage gradients of a GradABM to compute exact partial derivatives of any model output with respect to an arbitrary combination of parameters. We demonstrate the benefits of this approach on a case study of the first wave of COVID-19 in London, where we investigate the causes of variations in infections by age, socio-economic index, ethnicity, and geography. Finally, we also show that the same methodology allows for the design of optimal policy interventions. The code to reproduce the presented results is made available on GitHub [1].

KEYWORDS

Differentiable Agent-based Models; Sensitivity Analysis; Automatic Differentiation; Computational Epidemiology

ACM Reference Format:
Arnau Quera-Bofarull, Ayush Chopra, Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Anisoara Calinescu, Ramesh Raskar, and Michael Wooldridge. 2023. Don't Simulate Twice: One-Shot Sensitivity Analyses via Automatic Differentiation. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 10 pages.

[1] https://github.com/arnauqb/one_shot_sensitivity

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

1 INTRODUCTION

Decision making in complex environments, such as in the midst of a public health crisis, is challenging. Data-driven modelling and simulation tools to explore this complexity and the effects of policies can be invaluable. Agent-based models (ABMs) are powerful tools for capturing the dynamics of complex systems, most recently widely deployed in response to the COVID-19 pandemic. For instance, these have helped decide lockdown strategies [19], prioritize vaccination schedules [14, 41] as well as inform mitigation strategies in refugee settlements [6]. Their utility requires scalable specification, data-driven calibration, and sensitivity analyses for validation – which can often be challenging. First, simulations in fields such as climate science and epidemiology are generally on the scale of millions of agents with large networks of interactions [4, 16, 32], and performing a single forward pass may take several days [5]. Second, this makes calibration subpar when procedures involve running the model numerous times while iteratively tuning parameters [40, 46]. Third, sensitivity analyses require re-running the model under different parameter configurations [9, 10] which suffer from both the high computational cost as well as sub-optimally calibrated parameters.

GradABM [13] is a recent tensorized and differentiable design of ABMs that has been shown to alleviate some of these concerns regarding scalability and data-driven calibration. GradABM follows a tensorized implementation where agents are represented as vectors, their interaction networks as adjacency matrices, and all discrete distributions (e.g. Bernoulli) are reparameterised with continuous approximations. This framework realizes simulations that scale to million-size populations in few seconds on commodity hardware, run on both CPUs and GPUs, and are end-to-end differentiable [13]. Ensuring differentiability enables gradient-based learning, which allows to merge with deep neural networks (DNNs) and integrate heterogeneous data sources, to improve calibration. A remaining challenge is to perform sensitivity analyses of the model outputs with respect to the input parameters, in an efficient

manner. Even with faster (tensorized) simulations, current methods that rely on evaluating the simulator multiple times may be prohibitive due to exponential cost scaling from joint regulation of multiple parameters. However, sensitivity analyses are essential to assess the confidence levels associated with a given model fit and to validate its utility for robust decision making. Making these computationally tractable is the focus of this work.

We introduce a gradient-based mechanism for conducting sensitivity analyses. The key insight is to repurpose the gradients of GradABMs to avoid having to rerun the simulations. Basically, parametric sensitivity analyses require running the model under different parameter configurations [9, 10]. These effectively seek to estimate numerically the partial derivatives of the model parameters with respect to the output. GradABMs provide this information directly by making use of automatic differentiation [7] in an exact way. Essentially, this allows us to conduct sensitivity analyses without requiring any new simulation runs.

For experiments, we utilize the highly-detailed June epidemiological ABM [5], which has been parameterised to simulate the spread of COVID-19 among all the people in England (~55M agents). We reformulate June as a GradABM (henceforth, call this implementation GradABM-June) and calibrate the model to the first wave of the pandemic in London to investigate differences in infection rates by age group, socio-economic index, ethnicity, and geography. This allows us to ask questions of the form: i) was the socio-economic status of an individual a risk factor for COVID-19 infection? (section 5.1), ii) what made the Asian population of London disproportionately vulnerable to infection? (section 5.2)

Further, beyond such retrospective analysis, our gradient-based mechanism can also be used to ask prospective questions of the form: what could the government have done differently when designing a lockdown strategy? (section 5.3). Exploring such optimal policy design is a core part of informing decision making and can be made possible by ABMs that simulate counterfactuals. However, their discovery presents similar challenges to those arising when calibrating traditional ABMs — namely the time-consuming and challenging nature of optimising complex models. GradABM-June produces (epidemiological) cost-effective policies before and after the national lockdown in London during the first wave of the pandemic.

2 RELATED WORK

In this section, we first contextualize the use of ABMs in policy decision making, commenting on the current limitations that they face. We then explain the importance of carrying sensitivity analyses in ABMs and the high computational cost associated with them. Lastly, we review other domains where gradients have been used to understand latent model behaviour.

2.1 ABMs for Decision Making

ABMs aim to model the behaviour of complex systems through the interactions of their most primitive components. For instance, in an epidemiological model, ABMs simulate the movement and interactions of a population, which enables the tracking of the spread of an infection at an individual level. In principle, this makes ABMs an ideal tool for policy design, since we can study very detailed counter-factuals.

ABMs have been used for decision making in many settings including: economics [15], climate change [11, 17], public health [48] and emergency response [28]. Recently, ABMs have been widely deployed for decision making in response to the COVID-19 pandemic. Uses include informing lockdown measures [19, 30], contact tracing policies [8, 29], and vaccine distribution mechanisms [34, 41]. Specifically, the June model used in this paper informed acute hospital capacity planning in the UK in collaboration with the NHS [5] and vaccination campaign planning [14], and it was deployed to model intervention strategies in refugee settlements [6].

Barriers remain to the widespread use of ABMs. A key challenge is to specify realistic simulation populations, calibrate with real-world data, and validate correctness of decisions [33]. This must often be performed rapidly to inform timely decision making. However, the complexity of the models, and their time consuming inference and calibration time hinders their applicability.

2.2 Sensitivity of ABMs

ABMs tend to be parameterised by a large number of variables. Several can be fixed by expert knowledge, while others require the model to be calibrated to real world data. Given the non-linearity of ABMs, small variations of input parameters can lead to a wide range of emergent behaviour. It is therefore crucial to understand which outputs are most sensitive to which parameters. Different methods have been proposed to carry this task [9, 42, 47], but they all require to run the ABM a large number of times, especially when the number of input and output quantities is large. Since most ABMs are slow to run, thorough sensitivity analyses of ABMs are often computationally prohibitive. An example of this is the uncertainty quantification study done by [18] on the CovidSim epidemiological model [19]. The study required the use of a supercomputer and thousands of evaluations of the CovidSim model to determine the sensitivity of the model outputs to the input parameters. In a real case scenario where the validity of a forecast is just as precious as the time needed to obtain it, this hinders the adoption of ABMs as a tool for real-time policy decision making. The key motivation for this work is that implementing an ABM as a GradABM allows to significantly improve the speed and cost of such analyses by just using the (directly accessible) gradients in the simulator to measure sensitivity of output to specific parameters.

2.3 Analysis with Gradients

Gradient-based methods can be used to identify saliency of input features by using the gradient to estimate their influence on the output. Such saliency maps are a popular visualization technique used to add a layer of interpretability to deep neural networks (DNNs). They are used to localize input regions essential to classifying an image in computer vision [12, 43, 45, 54]; to explain how agents choose actions in deep reinforcement learning [2, 3, 39, 52]; and to analyse the robustness of the trained DNN [35, 53]. The end-to-end differentiability of DNNs allows to quickly execute such analysis at scale, due to direct availability of gradients (from the computation graph) via automatic differentiation. The key focus of this work is to bring similar insights for validating ABMs through

GradABMs, which are amenable to gradient-based learning with automatic differentiation.

3 PRELIMINARIES

This section describes the epidemiological model June and summarizes the specification and calibration of a GradABM. We then present the training and inference pipelines for GradABM-June.

3.1 JUNE Epidemiological Model

The June epidemiological model is an ABM built with the highest granularity that a demographic census allows. Initially designed to model the spread of SARS-CoV-2 in England, June models the movement of a synthetic population in and between a variety of locations such as households, care homes, companies, and leisure venues. All locations are distributed according to real data, and the daily activities of the agents are calibrated from time use surveys. We refer to the main June paper [5] for a thorough description of the model.

The June model assisted NHS operations and was afterwards adapted to model refugee settlements in collaboration with United Nations agencies [6]. The original implementation of June following an object-oriented Python approach is available in GitHub [2]. We focus here on the most relevant part of the model for the purpose of this work: the transmission module, since it is the component that needs to be calibrated using the observed infection data.

Given a susceptible agent exposed to the virus at location L, the probability of that agent getting infected is given by

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

where ψ_s is the inherent susceptibility to infection of the agent, the summation is over all contacts with infected individuals at the location, $I_i(t)$ 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 are not a physical quantity that can be measured, they are typically calibrated using data on the number of cases or fatalities over time.

3.2 GradABM

GradABMs [13] are ABMs amenable to gradient-based learning with automatic differentiation. First, in contrast to conventional ABMs which follow an object-oriented design where agents are modeled as objects (which is appealing but often inefficient), GradABMs follow a tensorized design where agents are represented as vectors and their interaction networks with adjacency matrices. This allows to code ABMs using modern machine learning frameworks, which can run forward simulations rapidly with high parallelization through GPUs. Second, ABMs are stochastic simulators which may conventionally require sampling from discrete distributions during execution (e.g.: all interactions with an infected agent may not result in a new infection; COVID-19 tests may return false positives/negatives with some probability). In GradABM, all discrete distributions (e.g. Bernoulli) are reparameterised with continuous approximation (e.g. Gumbel-Softmax [31]), to enable end-to-end differentiability. This design allows scaling to realistic populations and leveraging automatic differentiation, which provides a wide array of practical benefits (e.g. integrating with deep neural networks).

A direct application of integrating the GradABM with neural networks is that we can use a deep neural network (CalibNN) to assist us in the calibration process. The CalibNN is tasked with predicting the optimal parameters θ, which are input to the GradABM, that minimize a loss function $\mathcal{L}$ that compares the GradABM output $\hat{y}$ to the calibration data y. Therefore, rather than learning θ directly, we are learning the CalibNN weights ϕ, such that optimization is done as

$$\phi_{t+1} = \phi_t - \gamma \times \frac{\mathrm{d}\mathcal{L}(y, \hat{y}; \theta)}{\mathrm{d}\phi} \tag{2}$$

where γ adjusts the learning rate. The full details of this framework are described in [13]. The key takeaway is that the tensorized and differentiable specification of GradABM makes simulation and calibration of large-scale ABMs possible without the need to use techniques that increase the model miss-specification error, such as surrogate modeling [49].

4 GRADABM-JUNE

In this section, we describe GradABM-June, which is the ported version of the June model into the GradABM framework in PyTorch. GradABM-June exhibits a computational speed-up of x10,000 with respect to the original June implementation, eliminating the computational constraints that June suffered from. Our objective is to demonstrate how we can use the gradients of GradABM-June to conduct fast sensitivity analyses. As a case study, we design an experiment centered on the first wave of COVID-19 in London. The first step is to specify the calibration of the GradABM-June model for this experiment, which we do next.

4.1 GradABM-June Calibration

To initialize GradABM-June, we need to specify the transmission input parameters (see section 3.1), the symptoms input parameters (e.g. mortality rate), and the active policies as a function of time. We take all the clinical parameters regarding the probabilities and the times to transition to different symptomatic stages from the Covasim model (see Table 1 of [32] and references therein). This fixes all the parameters regarding the progression of symptoms of the agents. As for the transmission model, we aim to calibrate the contact intensity parameters β_i, described in section 3.1, as well as the number of initial cases in the simulation. In Table 1, we show the complete list of the parameters to calibrate, as well as the range (in logarithmic space) within which they are allowed to vary.

GradABM-June can also simulate a wide range of policies, as in the original June model [5]. For simplicity, the only policy we consider here is social distancing (SD), which we vary at 3 specific dates corresponding to government guidelines that were active during the first wave. A SD policy is parameterised by α_i where i corresponds to each location, similarly to β_i. α_i is then a multiplicative factor to β_i that reduces the contact intensity at location i. Table 2 describes the lockdown parameters that we consider. These parameters should be interpreted as a heuristic implementation of

[2] https://github.com/idas-durham/june

Table 1: List of parameters that are calibrated in GradABM-June. All β's are the respective location contact intensities, and I_i is the fraction of initial number of cases.

Parameters to calibrate	
Name	Range [$\log_{10}$]
$\beta_{\text{household}}$	[0.0, 1.5]
$\beta_{\text{care home}}$	[0.0, 1.5]
β_{school}	[0.0 0.5]
β_{company}	[−1.0, 1.0]
$\beta_{\text{university}}$	[−1.0, 1.0]
β_{pub}	[−1.0, 1.0]
β_{grocery}	[−1.0, 1.0]
β_{gym}	[−1.0, 1.0]
β_{cinema}	[−1.0, 1.0]
$\beta_{\text{residence visits}}$	[−1.0, 1.0]
$\beta_{\text{carevisits}}$	[−1.0, 1.0]
I_i	[−4.0, −2.0]

the lockdown and are based on previous calibrations of the June model [49], where we also set $\alpha_i = 0$ in those locations that were completely closed during the first lockdown.

Table 2: SD parameters as a function of time.

Lockdown parameters			
SD parameter	16/03/2020	24/03/2020	11/05/2020
$\alpha_{\text{household}}$	0.8	0.5	0.6
$\alpha_{\text{care home}}$	0.8	0.5	0.6
α_{school}	0.8	0.1	0.1
α_{company}	0.8	0.2	0.4
$\alpha_{\text{university}}$	0.8	0.0	0.0
α_{pub}	0.8	0.0	0.0
α_{grocery}	0.8	0.3	0.3
α_{gym}	0.8	0.0	0.0
α_{cinema}	0.8	0.0	0.0
$\alpha_{\text{residence visits}}$	0.8	0.0	0.0
$\alpha_{\text{carevisits}}$	0.8	0.25	0.25

As for the calibration data, we consider the daily deaths time-series for each of the London local authority districts [26]. We do not consider any additional data sources, since quantities such as case data or hospitalisations where very uncertain during the first wave.

The original June model was calibrated using a surrogate model in the form of a Gaussian Process (GP), which was used to iteratively discard implausible regions of the parameter space, a process known as history matching [1, 49]. The procedure ends when the parameter space has been reduced to a region containing parameters that recreate the data. The training of the GP emulator using the June model required 100k CPU hours, due to the high computational cost of the June model and the required number of evaluations.

Here, we use CalibNN to calibrate GradABM-June in the specified setup. Each calibration conducted by CalibNN generates a set of parameters that recover the observed data. Given that the deaths data is not informative enough to break the degeneracy among all the β_i parameters, there are multiple sets of β_i combinations that can reproduce the same deaths curve. Ideally, one should perform a full Bayesian calibration to obtain the posterior distribution over the parameters to correctly account for uncertainty, but this is beyond the scope of this work. Instead, we run the calibration algorithm 5 times to obtain 5 different sets of parameters and we conduct future experiments on each of these sets. It is worth noting that each calibration required fewer than 50 model evaluations, showcasing the efficacy of the CalibNN pipeline. For the ensemble of 5 runs, we compute the mean deaths curve, which we plot in Figure 1, where we also show a comparison with ground truth data as well as with the previous calibration of the June model. The error of the data is taken to be $2\sigma_{\text{obs}}$, with $\sigma_{\text{obs}} = \sqrt{\bar{y}}$ to account for Poissonian counts error. For both June and GradABM-June the uncertainty bounds are computed by taking 2 standard deviations of the set of fitted models. The original June calibration [49] aimed to reproduce the stratified deaths by age across all the regions of England, so it is expected that GradABM-June provides a better fit, since we only calibrate to London.

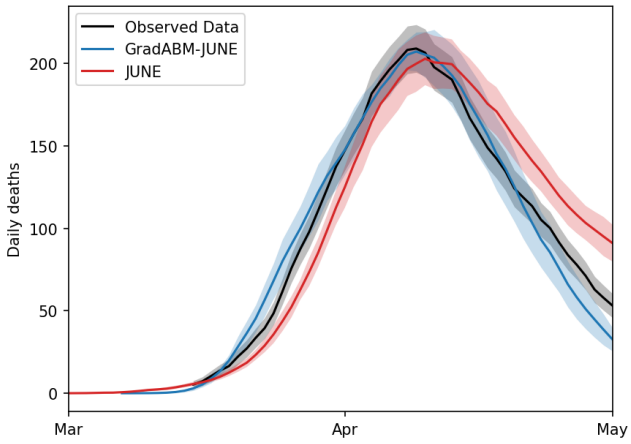

Figure 1: Daily deaths in London by date. Real data is shown in black, with the shadowed region denoting a 2σ Poisson error. The blue region corresponds to the GradABM-June average calibration of the 5 baseline runs. The red region denotes the calibration fit from the original June model to London from [49]. For both simulations, the error denotes 2σ, where σ is the standard deviation across the fitted models outputs. It is important to note that the June calibration [49] was performed for all England, so the London fit is a compromise across regions.

As an additional check, we compare in Figure 2 the obtained prevalence at the end of the first wave in GradABM-June with the Ward study [50], which measured the COVID-19 antibody prevalence at the end of the first wave in England. Although not directly comparable, since the prevalence in London may differ from the average England one, the good agreement between the two gives us confidence that our calibrated model runs capture the real world dynamics.

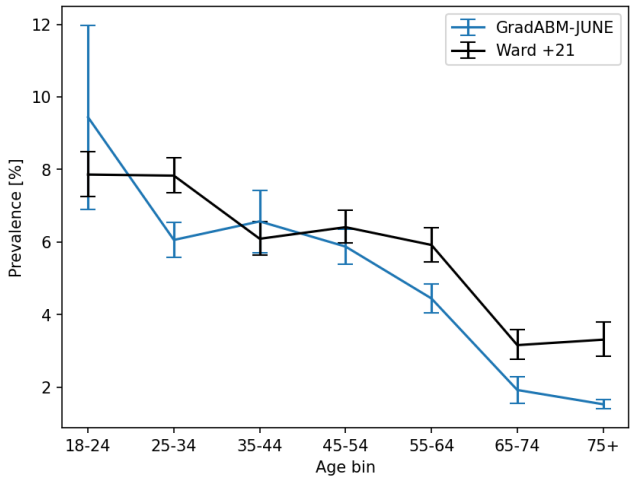

Figure 2: Cases prevalence for different age groups. Black error bars show estimates of prevalence from real data [50] of prevalence in England, blue error bars are the average and standard deviation of the 5 best fit models using GradABM-June in London. Both lines show a downward trend with age, with younger age groups being infected the most.

	Simulation	Calibration	Sensitivity Analysis
JUNE	50 hours	100k hours	5k hours
GradABM-June (CPU)	5 minutes	10 hours	10 minutes
GradABM-June (GPU)	5 seconds	20 minutes	10 seconds

Table 3: Approximate running times for simulation, calibration, and validation of the June model compared to the GradABM-June implementation, running on CPU and GPU.

4.2 GradABM-June Sensitivity Analysis

We now present the main methodological contribution of this work: fast sensitivity analysis using gradients. The automatic differentiation engine of machine learning frameworks such as PyTorch [38] stores the computational graph of the model execution. This computational graph contains a trace of all the operations that have been executed in the model, thus allowing to calculate derivatives in a very fast and exact way. Interpreting the gradient as the rate of change of the output with respect to a variation of the input parameter, the gradients directly give us the sensitivity of the former with respect to the latter. Hence, after one run of GradABM-June, we can directly query the computational graph to obtain the sensitivity of any output in GradABM-June to any input parameter. In contrast, a non-differentiable simulator would need to be run multiple times with slight variations of the input parameters to achieve the same task. This can be a great limitation if the particular ABM is computationally expensive [18]. In section 5.2, we present a series of sensitivity analyses carried on the calibrated GradABM-June runs, with the aim of understanding how COVID-19 affected different demographics groups during the first wave in London.

4.3 GradABM-June Optimal Policy Design

The last bit of utility that we can extract from the gradients is a method to design optimal policies. Let us suppose we want to implement a social distancing policy on the 16th of March 2020, corresponding to the first time-stamp in Table 2. To fully contain the virus, the most effective policy would be to just reduce all β_i as much as possible, but drastic policies come at a great economical and social cost. Instead, let us suppose that we want to implement a policy that moves us a step of length c in the β parameter space, where c is not too large. That is, after the policy, $\vec{\beta}' = \vec{\beta} - \vec{c}$. The optimal direction of $\vec{c}$ is directly given by the gradient of the number of cases with respect to β_i. In section 5.3, we design an experiment where we compare this optimal policy to the lockdown described in Table 2

4.4 Performance Benchmarking

To summarize the benefits of GradABM-June, we show in Table 3 the differences in running times for the tasks of simulation, calibration, and sensitivity analysis of June compared to GradABM-June (running on CPU or GPU). Even when memory requirements would stop from running GradABM-June on a GPU, we still see massive gains in the CPU version. Of particular importance is the amortization of the computational graph, which allows us to perform sensitivity analysis at a very low computational cost. This extends prior results in [13] which demonstrate how GradABM leads to faster simulations and more efficient calibration (fewer CPU hours), due to the tensorized implementation and end-to-end differentiability.

5 RESULTS

As a first study, we query the computational graph to compute the sensitivity of the number of cases at the end of the simulation with respect to the contact intensity at each location (β_i) . In Figure 3, we plot the gradient $\frac{\mathrm{d}(\text{cases})}{\mathrm{d}\beta_i}$ averaged over the 5 calibrated runs, with the error bars denoting 2 standard deviations. It is worth noting that this plot is more informative than simply plotting the number of infections in each location, since the gradient takes into account the impact of input parameters on secondary infections. For instance, locations such as pubs or household visits may not produce lots of infections directly, but can play a critical role in spreading the virus to previously unaffected parts of the population. These effects are captured by the gradients.

5.1 Social Inequalities and COVID-19

One of the key components of the June model, which is present in the new GradABM-June version, is the level of detail of its synthetic population. In particular, each agent is assigned an age, sex, ethnicity, and socio-economic status following the English census data. We use the ethnicity categorisation employed by the English census [20], which divides the population into five groups: White, Mixed, Asian, Black, and Other. As for the socio-economic

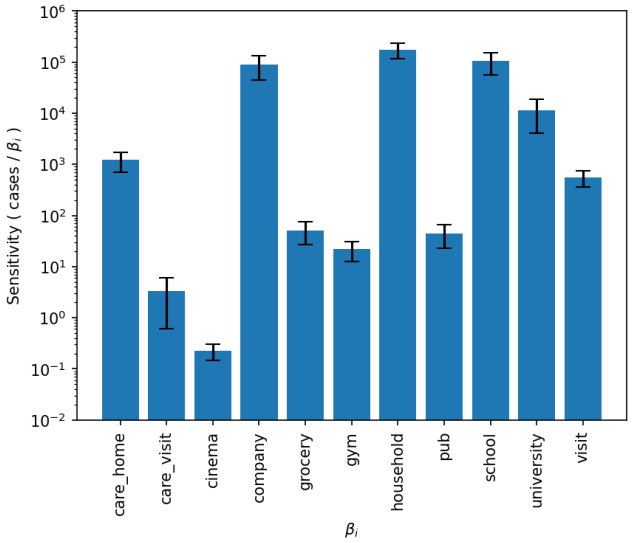

Figure 3: Value of the gradient of the number of cases with respect to each input parameter β_i.

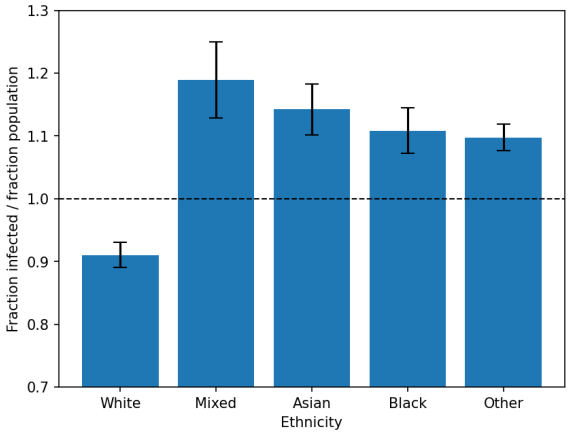

Figure 4: Fraction of infected individuals by ethnicity, divided by the fraction of the population they represent. Non-white ethnic groups are disproportionately more affected than White.

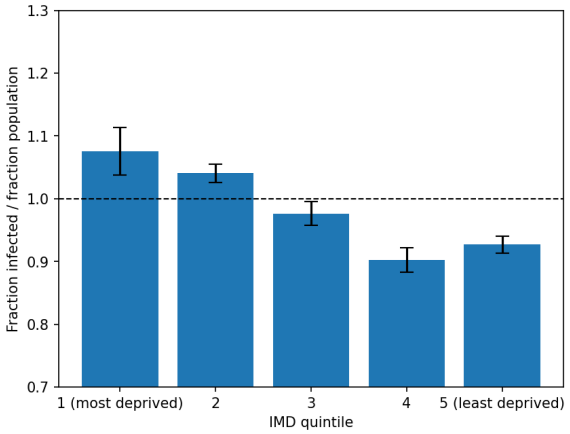

Figure 5: Fraction of each IMD quintile in the infected population normalised by the fraction of IMD agents in the general population. Agents living in the most deprived areas are significantly over-represented in the infected population.

status, each agent in GradABM-June is assigned a quintile of Index of Multiple Deprivation (IMD) according to their area of residence [37]. Lower indices correspond to lower socio-economic status, while higher indices correspond to those least deprived.

Epidemiological data from England shows that certain demographic groups suffer from higher attack rates to COVID-19. Notably, [27, 36, 51] find that non-white ethnic groups are at a higher risk of death compared to the white population. Furthermore, people living in the most deprived areas are significantly more vulnerable than their less deprived counterparts. To examine whether our calibrated models exhibit the same behaviour, we study the attack rate for each demographic group,

$$f^d = \frac{n_i^d / n_i}{n^d / n}, \tag{3}$$

where n_i^d is the number of infected people in demographic group d, n_i is the total number of infected people, n^d is the number of people in demographic group d, and n is the total number of people. Intuitively, if $f^d > 1$ then the demographic group d is overly represented among the infected population, and conversely for $f^d < 1$.

We find that the GradABM-June calibrated models naturally reproduce the demographic imbalances that are present in the epidemiological data. In Figure 4, we plot the f^d for each of the 5 ethnic groups. Consistent with the findings of [51], non-white ethnic groups are at a higher risk of infection. Similarly, in Figure 5, we plot f^d, where d now denotes each of the IMD quintiles. We observe that lower IMD agents are significantly more likely to be infected than agents with higher IMDs, in accordance with [51].

5.2 Sensitivities of Demographic Groups to Infection Locations

Given that we recover the broad social imbalances in the infection data, we now study the specific mechanism in the model which generates these inequalities. Even with a calibrated model, it is hard to analyse the entire latent structure. However, we can use the information in the gradients to circumvent this through a sensitivity analysis. The differences in sensitivity across demographic groups provides insight into the infection channels that are most relevant for each group.

In what follows, we analyse the sensitivity of attack rates (f^d) for each demographic group to the input parameters. Given that the most sensitive parameters are β_i with $i \in$ {household, school,

company, university} (see Figure 3), we restrict our analysis to these, for clarity.

(1) **Differences among age groups**
We first look at differences in sensitivity among age groups. In Figure 6, we plot the sensitivity of f^d for 8 different age groups. The results are quite intuitive. On the one hand, children are very susceptible to β_{school} and, on the other hand, increasing β_{company} reduces their representation in the number of infected individuals. Children tend to live in larger households than adults, so they are significantly more sensitive to the value of $\beta_{\text{household}}$. It is also interesting to observe that the fraction of young adults infected is anti-correlated with β_{school}, which highlights the fact that older adults are more likely to have children or work at a school than young ones. Lastly, the 18-24 age group is, unsurprisingly, the most affected by $\beta_{\text{university}}$.

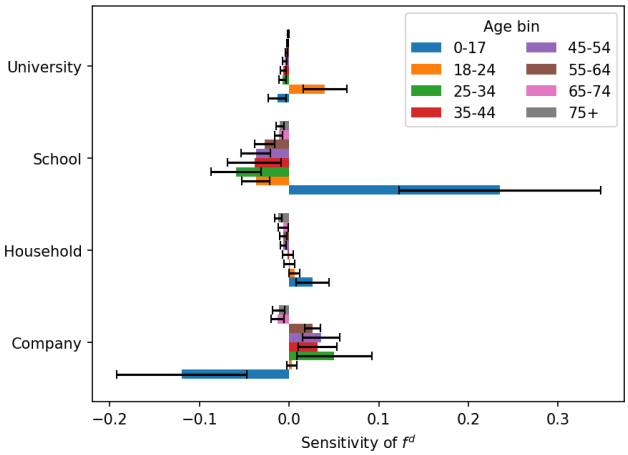

Figure 6: Sensitivity of f^d for different age groups. Error bars computed by running the sensitivity analysis over the baseline runs.

(2) **Differences among ethnic groups**
Next, we look at different ethnic groups. We plot the sensitivity of f^d for each ethnic group in Figure 7. We observe a very distinct pattern between the white and non-white population. First, the Asian ethnic group is very sensitive to $\beta_{\text{household}}$, which is consistent with the fact that larger households are more common in this group [23]. Conversely, most of the people who live alone are white [21], so the fraction of infected white people is anti-correlated with $\beta_{\text{household}}$. Similarly, non-white ethnic groups tend to have more children [21], making them significantly more sensitive to β_{school} than the white ethnic group. Lastly, the sensitivity to β_{company} has multiple components. The digital population of June takes into account unemployment, which is more prevalent among non-white ethnic groups [24]. However, our implemented lockdown (Table 2) does not model key-worker status in the population (although this is included in the original implementation of the June model [5]), which would be skewed by ethnicity [25] and would likely change this pattern. Finally, the Asian ethnic group is over-represented in higher education [44], as we can see from their sensitivity to $\beta_{\text{university}}$.

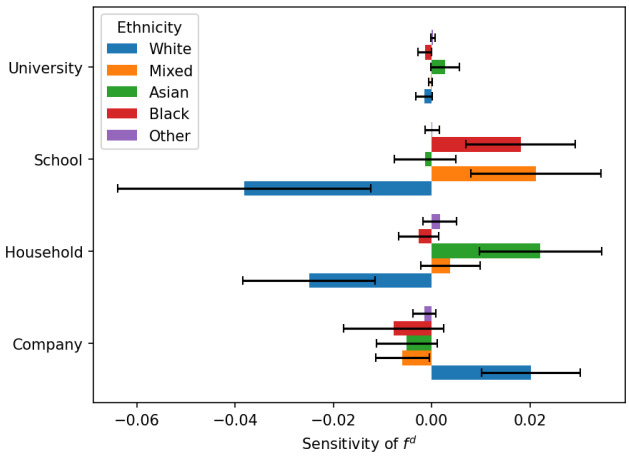

Figure 7: Sensitivity of f^d for different ethnic groups. Error bars computed by running the sensitivity analysis over the baseline runs.

(3) **Differences among socio-economic groups**
We now focus our attention to the differences among socio-economic groups. In Figure 8, we plot the sensitivity of f^d for each IMD quintile. We do not observe any significant differences in the rates of infection among the IMD quintiles. A more careful modeling of the lockdown which would take into account the demographic differences among the furloughed and key worker population which is included in the original implementation of the June model [5]) would likely reveal more structure in the sensitivity pattern. It is likely that the imbalances observed in Figure 5, are directly correlated with differences in ethnic groups, rather than purely to socio-economic status.

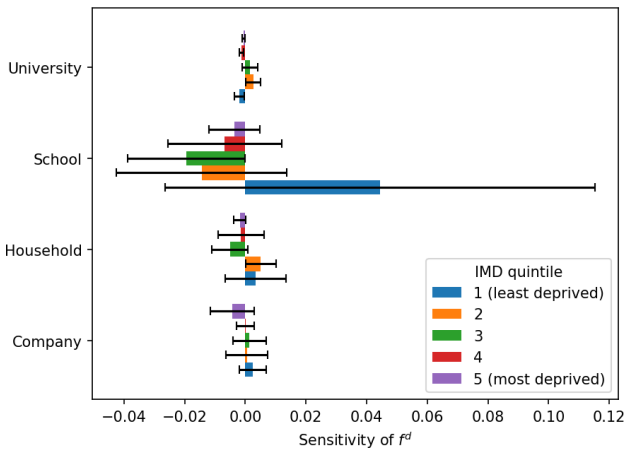

Figure 8: Sensitivity of f^d for different socio-economic groups. Error bars computed by running the sensitivity analysis over the baseline runs.

(4) **Differences among geographic groups**
Lastly, we study the sensitivity of the fraction of cases in each London local authority district. In Figure 9, we show a geographical map highlighting the differences among districts. We observe that West London is significantly sensitive to $\beta_{\text{household}}$. This is consistent with its demographic being predominantly of Asian origin, as we discussed previously. Conversely, East London is more sensitive to β_{school}, given its high birth rates [22].

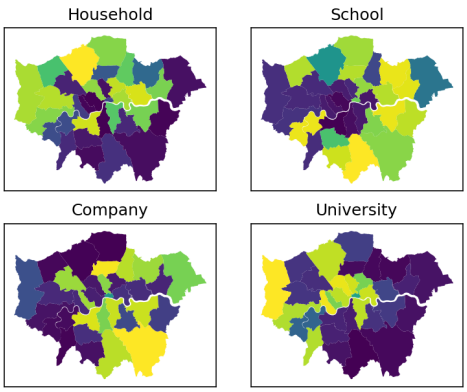

Figure 9: Sensitivity of f^d for different London districts. Higher sensitivity corresponds to lighter colors.

5.3 Optimal Cost-Effective Policy Design

As outlined in section 4.3, we can use Figure 3 to design an optimal policy, since it tells us the direction in the parameter space of maximum sensitivity of the cases respect to the β_i. We construct a study where we compare 4 different scenarios. We take each of the baseline runs that fit the data and we let them run until the 15th March 2020. On the 16th March 2020 we apply a different policy to each of the scenarios:

(1) An optimal cost-effective lockdown policy, consisting of taking a step of unit size ($c = 1$) along the direction of maximum variance.
(2) A naive cost-effective lockdown where $\alpha_i = 1/\sqrt{N_\beta}$, where N_β is the number of β parameters.
(3) The lockdown used in Table 2.
(4) No lockdown is implemented.

In Figure 10, we plot the cumulative number of cases for each of the 4 different scenarios. We observe that our designed optimal policy achieves a similar reduction in the total number of cases to the hard lockdown that was implemented, despite it only being a much smaller displacement of the β_i values.

6 CONCLUSIONS

We introduce a new way to conduct sensitivity analysis of ABMs: using the gradients obtained by automatic differentiation engines. This new method enables performing sensitivity analyses of large-scale ABMs at almost zero computational cost.

As a case study, we have ported the June [5] epidemiological code to the GradABM [13] framework, which allows for rapid simulation and calibration thanks to its tensorized implementation and

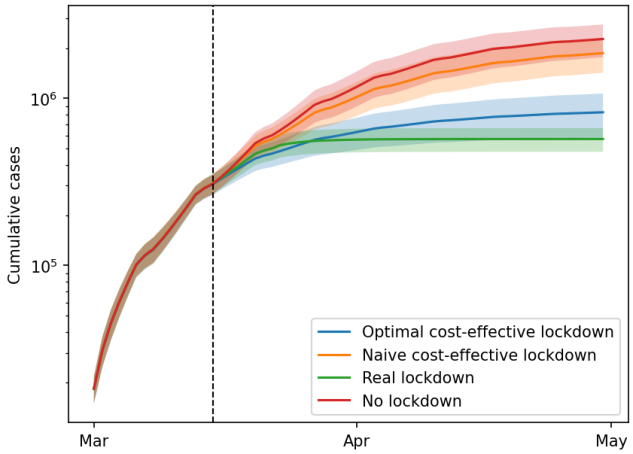

Figure 10: Comparison of the cumulative number of cases for 4 scenarios: 1. [blue] Optimal cost-effective lockdown obtained using the gradients, 2. [orange] Naive cost-effective lockdown, where social-distancing is applied homogeneously, 3. [green] Parameterisation of the real lockdown (Table 2), and 4. [red] No lockdown.

automatic differentiation. We have then calibrated GradABM-June to reproduce the first wave of COVID-19 in London, obtaining a good calibration fit much more rapidly than with the previously employed methods [49]. After obtaining a set of calibrated models, we have used the new proposed methodology to perform sensitivity analyses on them at no extra computational cost. This has been possible thanks to the exploitation of the automatic differentiation-capabilities of machine learning frameworks such as PyTorch [38], which allow for the amortization of the computational graph.

GradABM-June is able to reproduce the demographic and social inequalities present in the COVID-19 epidemiological data. Thanks to the easiness at which we can conduct sensitivity analysis in GradABM-June, we have studied how different demographic groups differ in their sensitivity to different infection locations, shedding light on opened questions regarding the imbalance of representation in the infected population for certain demographic groups. In particular, this work supports the hypothesis that the differences in attack rates across ethnic groups in London can be, at least partially, explained by considering sociological factors such as household overcrowding or the presence of children in the household. Lastly, we have also shown how the gradients obtained through automatic differentiation can be used to quickly design policies that can have maximum impact around a baseline scenario. In particular, we have designed, caveated by our model uncertainty, an alternative lockdown on March 2020 that has a comparable effect to the more severe version that was actually implemented.

Overall, this work brings ABMs one step closer to becoming the reference tool in simulation-based policy decision making which are conventionally often hindered by the prohibitive cost and difficulty of analysis. Interesting directions of future work include generalizing the idea to other application domains such as economics and biology, and going beyond the gradient to explore the utility of higher order differentiation.

ACKNOWLEDGMENTS

We would like to thank Nicholas Bishop and Imran Hashmi for helpful discussions about the project. This research was supported by a UKRI AI World Leading Researcher Fellowship awarded to Wooldridge (grant EP/W002949/1). M. Wooldridge and A. Calinescu acknowledge funding from Trustworthy AI - Integrating Learning, Optimisation and Reasoning (TAILOR) (https://tailor-network.eu/), a project funded by European Union Horizon2020 research and innovation program under Grant Agreement 952215. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1 and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

REFERENCES

[1] Ioannis Andrianakis, Ian R. Vernon, Nicky McCreesh, Trevelyan J. McKinley, Jeremy E. Oakley, Rebecca N. Nsubuga, Michael Goldstein, and Richard G. White. 2015. Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. *PLOS Computational Biology* 11, 1 (Jan. 2015), e1003968. https://doi.org/10.1371/journal.pcbi.1003968 Publisher: Public Library of Science.

[2] Raghuram Mandyam Annasamy and Katia Sycara. 2019. Towards better interpretability in deep q-networks. In *Proceedings of the AAAI conference on artificial intelligence* (Honolulu, HI, USA), Vol. 33. AAAI Press, 4561–4569. Issue: 01.

[3] Akanksha Atrey, Kaleigh Clary, and David Jensen. 2020. Exploratory Not Explanatory: Counterfactual Analysis of Saliency Maps for Deep Reinforcement Learning. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rkl3m1BFDB

[4] Robert L. Axtell. 2016. 120 Million Agents Self-Organize into 6 Million Firms: A Model of the U.S. Private Sector. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 806–816.

[5] Joseph Aylett-Bullock, Carolina Cuesta-Lazaro, Arnau Quera-Bofarull, Miguel Icaza-Lizaola, Aidan Sedgewick, Henry Truong, Aoife Curran, Edward Elliott, Tristan Caulfield, Kevin Fong, Ian Vernon, Julian Williams, Richard Bower, and Frank Krauss. 2021. June: open-source individual-based epidemiology simulation. *Royal Society Open Science* 8, 7 (July 2021), 210506. https://doi.org/10.1098/rsos.210506 Publisher: Royal Society.

[6] 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. 2021. 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 (Oct. 2021), e1009360. https://doi.org/10.1371/journal.pcbi.1009360

[7] Atılım Günes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. 2017. Automatic differentiation in machine learning: a survey. *The Journal of Machine Learning Research* 18, 1 (Jan. 2017), 5595–5637.

[8] Martin Bicher, Claire Rippinger, Christoph Urach, Dominik Brunmeir, Uwe Siebert, and Niki Popper. 2021. Evaluation of Contact-Tracing Policies against the Spread of SARS-CoV-2 in Austria: An Agent-Based Simulation. *Medical Decision Making* 41, 8 (Nov. 2021), 1017–1032. https://doi.org/10.1177/0272989X211013306 Publisher: SAGE Publications Inc STM.

[9] Emanuele Borgonovo, Marco Pangallo, Jan Rivkin, Leonardo Rizzo, and Nicolaj Siggelkow. 2022. Sensitivity analysis of agent-based models: a new protocol. *Computational and Mathematical Organization Theory* 28, 1 (March 2022), 52–94. https://doi.org/10.1007/s10588-021-09358-5

[10] J. Cariboni, D. Gatelli, R. Liska, and A. Saltelli. 2007. The role of sensitivity analysis in ecological modelling. *Ecological Modelling* 203, 1 (April 2007), 167–182. https://doi.org/10.1016/j.ecolmodel.2005.10.045

[11] Juana Castro, Stefan Drews, Filippos Exadaktylos, Joël Foramitti, Franziska Klein, Théo Konc, Ivan Savin, and Jeroen van den Bergh. 2020. A review of agent-based modeling of climate-energy policy. *WIREs Climate Change* 11, 4 (2020), e647. https://doi.org/10.1002/wcc.647 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.647.

[12] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. 2018. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 839–847.

[13] Ayush Chopra, Alexander Rodríguez, Jayakumar Subramanian, Arnau Quera-Bofarull, Balaji Krishnamurthy, B. Aditya Prakash, and Ramesh Raskar. 2023. Differentiable Agent-based Epidemiology. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems* (London, UK) *(AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC.

[14] Carolina Cuesta-Lazaro, Arnau Quera-Bofarull, Joseph Aylett-Bullock, Bryan N. Lawrence, Kevin Fong, Miguel Icaza-Lizaola, Aidan Sedgewick, Henry Truong, Ian Vernon, Julian Williams, Christina Pagel, and Frank Krauss. 2021. Vaccinations or Non-Pharmaceutical Interventions: Safe Reopening of Schools in England. https://doi.org/10.1101/2021.09.07.21263223 Pages: 2021.09.07.21263223.

[15] Herbert Dawid and Michael Neugart. 2011. Agent-based Models for Economic Policy Design. *Eastern Economic Journal* 37, 1 (Jan. 2011), 44–50. https://doi.org/10.1057/eej.2010.43

[16] Christophe Deissenberg, Sander van der Hoog, and Herbert Dawid. 2008. EURACE: A massively parallel agent-based model of the European economy. *Appl. Math. Comput.* 204, 2 (Oct. 2008), 541–552. https://doi.org/10.1016/j.amc.2008.05.116

[17] Jlenia Di Noia. 2022. Agent-Based Models for Climate Change Adaptation in Coastal Zones. A Review. https://doi.org/10.2139/ssrn.4180554

[18] Wouter Edeling, Hamid Arabnejad, Robbie Sinclair, Diana Suleimenova, Krishnakumar Gopalakrishnan, Bartosz Bosak, Derek Groen, Imran Mahmood, Daan Crommelin, and Peter V. Coveney. 2021. The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nature Computational Science* 1, 2 (Feb. 2021), 128–135. https://doi.org/10.1038/s43588-021-00028-9 Number: 2 Publisher: Nature Publishing Group.

[19] Neil M. Ferguson. 2020. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. https://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/

[20] Office for National Statistics. 2011. DC2101EW (Ethnic group by sex by age) - Nomis - Official Census and Labour Market Statistics. https://www.nomisweb.co.uk/census/2011/dc2101ew

[21] Office for National Statistics. 2011. Families and households. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/families-and-households/latest

[22] Office for National Statistics. 2014. Births and Fertility Rates, Borough - London Datastore. https://data.london.gov.uk/dataset/births-and-fertility-rates-borough

[23] Office for National Statistics. 2019. Overcrowded households. https://www.ethnicity-facts-figures.service.gov.uk/housing/housing-conditions/overcrowded-households/latest

[24] Office for National Statistics. 2019. Unemployment. https://www.ethnicity-facts-figures.service.gov.uk/work-pay-and-benefits/unemployment-and-economic-inactivity/unemployment/latest

[25] Brigid Francis-Devine. 2020. Coronavirus: Which key workers are most at risk? https://commonslibrary.parliament.uk/coronavirus-which-key-workers-are-most-at-risk/

[26] UK government. 2022. England Summary | Coronavirus (COVID-19) in the UK. https://coronavirus.data.gov.uk

[27] Ewen M. Harrison, Annemarie B. Docherty, Benjamin Barr, Iain Buchan, Gail Carson, Tom M. Drake, Jake Dunning, Cameron J. Fairfield, Carrol Gamble, Christopher A. Green, Chris Griffiths, Sophie Halpin, Hayley E. Hardwick, Antonia Ho, Karl A. Holden, Joe Hollinghurst, Peter W. Horby, Clare Jackson, Srinivasa Vittal Katikireddi, Stephen Knight, Ronan A. Lyons, James MacMahon, Kenneth A. Mclean, Laura Merson, Derek Murphy, Jonathan S. Nguyen-Van-Tam, Lisa Norman, Piero L. Olliaro, Manish Pareek, Roberta Piroddi, Riinu Pius, Jonathan M. Read, Clark D. Russell, Naveed Sattar, Catherine A. Shaw, Aziz Sheikh, Ian P. Sinha, Olivia Swann, David Taylor-Robinson, Daniel Thomas, Lance Turtle, Peter JM Openshaw, J. Kenneth Baillie, Malcolm G. Semple, and Isaric4c Investigators. 2020. Ethnicity and Outcomes from COVID-19: The ISARIC CCP-UK Prospective Observational Cohort Study of Hospitalised Patients. https://doi.org/10.2139/ssrn.3618215

[28] Glenn I. Hawe, Graham Coates, Duncan T. Wilson, and Roger S. Crouch. 2012. Agent-based simulation for large-scale emergency response: A survey of usage and implementation. *Comput. Surveys* 45, 1 (Dec. 2012), 8:1–8:51. https://doi.org/10.1145/2379776.2379784

[29] Robert Hinch, William J. M. Probert, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, Luca Ferretti, Daniel Montero, James Warren, Nicole Mather, Matthew Abueg, Neo Wu, Olivier Legat, Katie Bentley, Thomas Mead, Kelvin Van-Vuuren, Dylan Feldner-Busztin, Tommaso Ristori, Anthony Finkelstein, David G. Bonsall, Lucie Abeler-Dörner, and Christophe Fraser. 2021. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLOS Computational Biology* 17, 7 (July 2021), e1009146.

https://doi.org/10.1371/journal.pcbi.1009146 Publisher: Public Library of Science.

[30] Nicolas Hoertel, Martin Blachier, Carlos Blanco, Mark Olfson, Marc Massetti, Marina Sánchez Rico, Frédéric Limosin, and Henri Leleu. 2020. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine* 26, 9 (Sept. 2020), 1417–1421. https://doi.org/10.1038/s41591-020-1001-6 Number: 9 Publisher: Nature Publishing Group.

[31] Eric Jang, Shixiang Gu, and Ben Poole. 2017. Categorical Reparameterization with Gumbel-Softmax. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rkE3y85ee

[32] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, Lauren George, Michał Jastrzębski, Amanda S. Izzo, Greer Fowler, Anna Palmer, Dominic Delport, Nick Scott, Sherrie L. Kelly, Caroline S. Bennette, Bradley G. Wagner, Stewart T. Chang, Assaf P. Oron, Edward A. Wenger, Jasmina Panovska-Griffiths, Michael Famulare, and Daniel J. Klein. 2021. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology* 17, 7 (July 2021), e1009149. https://doi.org/10.1371/journal.pcbi.1009149 Publisher: Public Library of Science.

[33] Alexander Lavin, Hector Zenil, Brooks Paige, David Krakauer, Justin Gottschlich, Tim Mattson, Anima Anandkumar, Sanjay Choudry, Kamil Rocki, Atılım Güneş Baydin, Carina Prunkl, Brooks Paige, Olexandr Isayev, Erik Peterson, Peter L. McMahon, Jakob Macke, Kyle Cranmer, Jiaxin Zhang, Haruko Wainwright, Adi Hanuka, Manuela Veloso, Samuel Assefa, Stephan Zheng, and Avi Pfeffer. 2021. Simulation Intelligence: Towards a New Generation of Scientific Methods. http://arxiv.org/abs/2112.03235 arXiv:2112.03235 [cs].

[34] Qingfeng Li and Yajing Huang. 2022. Optimizing global COVID-19 vaccine allocation: An agent-based computational model of 148 countries. *PLOS Computational Biology* 18, 9 (Sept. 2022), e1010463. https://doi.org/10.1371/journal.pcbi.1010463 Publisher: Public Library of Science.

[35] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2018. Towards Deep Learning Models Resistant to Adversarial Attacks. In *International Conference on Learning Representations*. OpenReview. https://openreview.net/forum?id=rJzIBfZAb

[36] Christopher A. Martin, David R. Jenkins, Jatinder S. Minhas, Laura J. Gray, Julian Tang, Caroline Williams, Shirley Sze, Daniel Pan, William Jones, Raman Verma, Scott Knapp, Rupert Major, Melanie Davies, Nigel Brunskill, Martin Wiselka, Chris Brightling, Kamlesh Khunti, Pranab Haldar, and Manish Pareek. 2020. Socio-demographic heterogeneity in the prevalence of COVID-19 during lockdown is associated with ethnicity and household size: Results from an observational cohort study. *eClinicalMedicine* 25 (Aug. 2020). https://doi.org/10.1016/j.eclinm.2020.100466 Publisher: Elsevier.

[37] Communities & Local Government Ministry of Housing. 2015. MHCLG Open Data : English Indices of Deprivation 2015 - LSOA Level. https://opendatacommunities.org/data/societal-wellbeing/imd/indices

[38] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[39] Nikaash Puri, Sukriti Verma, Piyush Gupta, Dhruv Kayastha, Shripad Deshmukh, Balaji Krishnamurthy, and Sameer Singh. 2020. Explain Your Move: Understanding Agent Actions Using Specific and Relevant Feature Attribution. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJgzLkBKPB

[40] Santiago Romero-Brufau, Ayush Chopra, Alex J. Ryu, Esma Gel, Ramesh Raskar, Walter Kremers, Karen S. Anderson, Jayakumar Subramanian, Balaji Krishnamurthy, Abhishek Singh, Kalyan Pasupathy, Yue Dong, John C. O'Horo, Walter R. Wilson, Oscar Mitchell, and Thomas C. Kingsley. 2021. Public health impact of delaying second dose of BNT162b2 or mRNA-1273 covid-19 vaccine: simulation agent based modeling study. *BMJ* 373 (May 2021), n1087. https://doi.org/10.1136/bmj.n1087 Publisher: British Medical Journal Publishing Group Section: Research.

[41] Nuru Saadi, Y-Ling Chi, Srobana Ghosh, Rosalind M. Eggo, Ciara V. McCarthy, Matthew Quaife, Jeanette Dawa, Mark Jit, and Anna Vassall. 2021. Models of COVID-19 vaccine prioritisation: a systematic literature search and narrative review. *BMC Medicine* 19, 1 (Dec. 2021), 318. https://doi.org/10.1186/s12916-021-02190-3

[42] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 181, 2 (Feb. 2010), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

[43] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 618–626. https://doi.org/10.1109/ICCV.2017.74

[44] Universities and Colleges Admissions Service. 2021. Entry rates into higher education. https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/higher-education/entry-rates-into-higher-education/latest

[45] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian J. Goodfellow, and Rob Fergus. 2014. Intriguing properties of neural networks. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1312.6199

[46] Farzaneh S. Tabataba, Bryan Lewis, Milad Hosseinipour, Foroogh S. Tabataba, Srinivasan Venkatramanan, Jiangzhuo Chen, Dave Higdon, and Madhav Marathe. 2017. Epidemic Forecasting Framework Combining Agent-Based Models and Smart Beam Particle Filtering. In *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, New York, USA, 1099–1104. https://doi.org/10.1109/ICDM.2017.145 ISSN: 2374-8486.

[47] Guus ten Broeke, George van Voorn, and Arend Ligtenberg. 2016. Which Sensitivity Analysis Method Should I Use for My Agent-Based Model? *Journal of Artificial Societies and Social Simulation* 19, 1 (2016), 5.

[48] Melissa Tracy, Magdalena Cerdá, and Katherine M. Keyes. 2018. Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annual Review of Public Health* 39, 1 (2018), 77–94. https://doi.org/10.1146/annurev-publhealth-040617-014317 _eprint: https://doi.org/10.1146/annurev-publhealth-040617-014317.

[49] 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. 2022. Bayesian emulation and history matching of JUNE. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380, 2233 (Oct. 2022), 20220039. https://doi.org/10.1098/rsta.2022.0039 Publisher: Royal Society.

[50] Helen Ward, Graham S. Cooke, Christina Atchison, Matthew Whitaker, Joshua Elliott, Maya Moshe, Jonathan C. Brown, Barnaby Flower, Anna Daunt, Kylie Ainslie, Deborah Ashby, Christl A. Donnelly, Steven Riley, Ara Darzi, Wendy Barclay, and Paul Elliott. 2021. Prevalence of antibody positivity to SARS-CoV-2 following the first peak of infection in England: Serial cross-sectional studies of 365,000 adults. *The Lancet Regional Health. Europe* 4 (May 2021), 100098. https://doi.org/10.1016/j.lanepe.2021.100098

[51] Elizabeth J. Williamson, Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, Amir Mehrkar, David Evans, Peter Inglesby, Jonathan Cockburn, Helen I. McDonald, Brian MacKenna, Laurie Tomlinson, Ian J. Douglas, Christopher T. Rentsch, Rohini Mathur, Angel Y. S. Wong, Richard Grieve, David Harrison, Harriet Forbes, Anna Schultze, Richard Croker, John Parry, Frank Hester, Sam Harper, Rafael Perera, Stephen J. W. Evans, Liam Smeeth, and Ben Goldacre. 2020. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 584, 7821 (Aug. 2020), 430–436. https://doi.org/10.1038/s41586-020-2521-4 Number: 7821 Publisher: Nature Publishing Group.

[52] Zhao Yang, Song Bai, Li Zhang, and Philip H. S. Torr. 2018. Learn to Interpret Atari Agents. https://doi.org/10.48550/ARXIV.1812.11276

[53] Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. 2019. Adversarial examples: Attacks and defenses for deep learning. *IEEE transactions on neural networks and learning systems* 30, 9 (2019), 2805–2824. Publisher: IEEE.

[54] Jianming Zhang, Sarah Adel Bargal, Zhe Lin, Jonathan Brandt, Xiaohui Shen, and Stan Sclaroff. 2018. Top-down neural attention by excitation backprop. *International Journal of Computer Vision* 126, 10 (2018), 1084–1102. Publisher: Springer.