Facilitating simulation development for global challenge response and anticipation in a timely way

Derek Groen *, Diana Suleimenova, Alireza Jahani, Yani Xue

Department Of Computer Science, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, Greater London, United Kingdom

A R T I C L E   I N F O

A B S T R A C T

An important subset of today's global crises, such as the 2015 migration crisis in Syria and the 2020 COVID pandemic, has a rapid and hard-to-extrapolate evolution that complicates the preparation of a community response. Simulation-based forecasts for such crises can help to guide the selection or development of mitigation policies or inform the efficient allocation of support resources. However, the time required to develop, execute and validate these models can often be intractably long, causing many of these forecasts to only become accurate after the damage has already occurred.

In this paper, we present a generic simulation development approach (or SDA) to tackle this challenge. It consists of three important phases: identifying anticipatory activities required for developing application-agnostic modelling tools, identifying activities required to adapt these models to address specific (global) challenges, and automating a large subset of the aforementioned activities using existing software tool. Here, a key aspect is to ensure that our models are reliable: this involves a range of tasks for validation, ensemble forecasting, uncertainty quantification and sensitivity analysis. To showcase the added value of a generic simulation development approach, we present and discuss two specific applications of this approach: one in the context of modelling conflict-driven migration and one in the context of modelling the spread of COVID-19.

1. Introduction

Global challenges are serious problems that can occur on a world-wide scale. These challenges can be long-lasting by nature, as in the strive for peace or fight against poverty, slowly evolving, as in global warming or environmental degradation, or sudden, as in the COVID-19 epidemic or the rapid onset of war. Simulations are helpful in anticipating and understanding the development of these challenges, as well as in identifying effective means to mitigate or prepare for them. In some cases, such as climate change, simulations are even essential to obtain a full understanding of the scale of the problem.

Now some global challenges can emerge and escalate in a matter of weeks (e.g., pandemics or armed conflicts), while production simulations often take many research years to develop, test and validate. This poses a particular problem, as often simulation-driven insights are only available after an acute global crisis has already inflicted much of its damage. We provide a few examples to illustrate this: (1) in the context of COVID-19 spread, the first comprehensive forecasting report in the UK [1] was released in mid-March 2020, leaving extremely little time for the government to intervene. This particular report presented results from CovidSim [2]: a C++ code that needed to be repurposed from

influenza to coronavirus spread before developers were able to make forecasts. (2) In the context of volcanic eruptions, the eruption of the Eyjafjallajökull in 2010 required the rapid use of an ash cloud model by the London Volcanic Ash Advisory Centre. Because a sophisticated forecasting infrastructure was available here, the centre was able to provide essential forecasts [3], although the crisis did lead to a range of modifications to the underlying model (called "NAME"), to make it more accurate on future occasions [4]. (3) Another example is the 2015 migration crises triggered by the war in Syria. In this case, no validation forecasting models had been published for conflict-driven migration in this context, even though the problem received media attention as early as in 2010.[1] Our local research team actually set out to develop such a model, established a prototype generic model in 2016 [5], and only managed to create a generalized approach for forecasting conflict-driven population displacement in 2017 [6], two years after the Syrian refugees fled from the conflict in large numbers.

In this paper, we aim to facilitate a more rapid simulation development process in response to rapid and hard-to-extrapolate global challenges such as pandemics and violent conflicts, and thereby increase the ability of researchers to deliver timely simulation insights

---

* Corresponding author.
  E-mail address: derek.groen@brunel.ac.uk (D. Groen).
[1] https://www.bbc.com/news/world-europe-11139345

in these situations. To do this, we present a generic conceptual framework called the Simulation Development Approach (SDA) in Section 2, where we identify the steps required to establish the underpinning generic models: ones that need to be provisioned and maintained to ensure effective and timely simulation development in response to sudden global challenges. In Section 3, we discuss the use of the SDA specifically in the context of anticipating and responding to global challenges, indicating clearly which steps are required as part of the global challenge response on a short time scale and which steps can be done in anticipation of a specific global challenge. In Section 4, we demonstrate the benefit of the SDA by applying it to two specific contexts: conflict-driven migration and the spread of COVID-19 in a local context. Lastly, we provide some closing thoughts in Section 5.

1.1. Background and related work

The notion of the 'simulation development process' is varied in literature (i.e. conceptual modelling [7], methodological process or framework [8], model or life cycle of simulation [9–11], model evaluation [12] and approach or steps for a successful simulation [13,14]). Despite these variations, they define the concept of a systematic and cyclic set of activities or phases of development. Specifically, these activities are the formulation of the real-world problem, the transformation of it into a model, the conversion of the model into a computerized simulation and the execution of experimental runs with analysis of the outcome [15].

The distinction between *model* and *simulation* is a formulated problem (model) prior to translation and deployment into a computational or computerized version (simulation). Thus, models are a representation of the real system through conceptual modelling, which is "a non-software specific description of the computer simulation model … describing the objectives, inputs, outputs, content, assumptions and simplifications of the model" [16]. Researchers derive a conceptual model from requirements to address the validity, reliability, credibility and reproducibility of computational solutions. An accurate formulation of requirements is a model design advantage providing the right information and simulation results. Hence, requirements are necessary for the rapid construction of models and execution of simulations.

The SDA we present in this paper can be applied to any type of simulation, although the two examples we present here concern global challenge simulations that (i) address a problem within a specific context, and (ii) have predefined assumptions, and are not self-learning as such. The first characteristic contrasts with digital twins, which can be used to address a wide range of problems and contexts at the expense of a more effort-intensive (and complex) simulation development procedure. The second characteristic contrasts simulation with artificial intelligence, or machine-learning based approaches. Although machine-learning tools are used to produce emergency forecasts by a range of communities, it has several important limitations [17]. For instance, it fundamentally needs historical reference data which, for example, in the case of a newly erupted armed conflict or disease, might not exist.

In a more applied context, there are several publications which relate to the work we present here. For instance, Kwakkel and Pruyt [18] examine the use of exploratory modelling and analysis for a range of complex systems, with the aim to provide forecasts that inform design decisions. Pruyt [19] independently also examined the simulation of an emergency intervention and development itself in the context of Ebola. In addition, the German Computational Immediate Response Center for Emergencies project[2] focuses on assessing the potential for the rapid use supercomputers in support of emergency-driven forecasting, while e.g. the Scientific Advisory Group for Emergencies in the UK has delivered forecast results directly to the UK government during the pandemic [20]. Related to this, there has been research specifically on enabling emergency access to large-scale computing resources [21], for instance to facilitate the modelling of storm surge events [22].

---

2. A generic Simulation Development Approach

Any effort to facilitate timely simulation development has to start with mapping the simulation development process itself. Suleimenova et al. [6] presented a "generalized" Simulation Development Approach (SDA) specifically for creating and validating simulations of conflict-driven migration, irrespective of the conflict of interest. In this section, we present an even more generic SDA; one that can be applied to a wide range of simulation development contexts irrespective of the application domain. We will do this step by step, moving gradually from a user perspective to a full developer perspective.

Before we do so however, it is useful to clarify a few concepts: (i) the SDA contains validation tasks, which are aimed to measure the degree to which a model is an accurate representation of the real world based on comparisons between computational results and experimental data [23,24]. When repeating simulations of a prior publication to test reproducibility, the results from the prior publication can be viewed as the "experimental data" to validate against. (ii) When we refer to sensitivity analysis, we measure to which extent variations in the numerical and physical parameters affect simulation outcomes. (iii) When performing uncertainty quantification [25], we run a given simulation a large number of times to account for probabilistic effects (aleatoric uncertainty), and vary the underlying parameters for each run within realistic ranges to account for epistemic uncertainty. In addition to these definitions, there is also a limitation of scope. Because ethical, societal, political and legal considerations are highly field-dependent (see Guillén and Teodoro [26] for ethical considerations on migration modelling), we have not incorporated them in the generic SDA. As a result, if those considerations have not been clearly accounted for yet when simulation development commences, the development and use of the simulation are likely to become delayed.

2.1. User perspective

A common approach to present simulation research is by focusing mostly on the simulation execution task and providing all the ingredients necessary for repeating the simulation in the article, through electronic supplements or via open-source tools. We sketch an SDA from this perspective in Fig. 1. Because this version of the SDA focuses only on repeatability, it effectively only involves a trivial selection of the situation (the one detailed in the paper), a collection of validation data (the results from the paper), and setting up and executing an identical simulation. Once the simulation is run, the researcher can quantify the uncertainties on their repeated runs (or perform sensitivity analysis as part of the same task), validate them against the original results, and evaluate the outcome.

A use case that is less commonly presented but more commonly applied is the reuse of an existing simulation in a slightly different context. Such an adaptation of the SDA is bound to involve some kind of modification or refinement of the simulation. For instance, one may introduce new rules, events, objects or boundary conditions. Another thing that changes is the situation selection task. From this perspective, the user will select a different situation and may articulate one or multiple counterfactuals. Now because a simulation is essentially the implemented counterpart of a conceptual model, this implies that we also likely need to modify it. As a result, we obtain the SDA presented in Fig. 2 for simulation reuse.

2.2. Developer perspective

Moving towards the developer perspective, we will want to specify exactly what components we are refining. In the case of reuse, we would take a model (or simulation) suited for an existing situation, and refine it to accommodate its application in a new situation. More generally, many of the models (and their implemented simulation
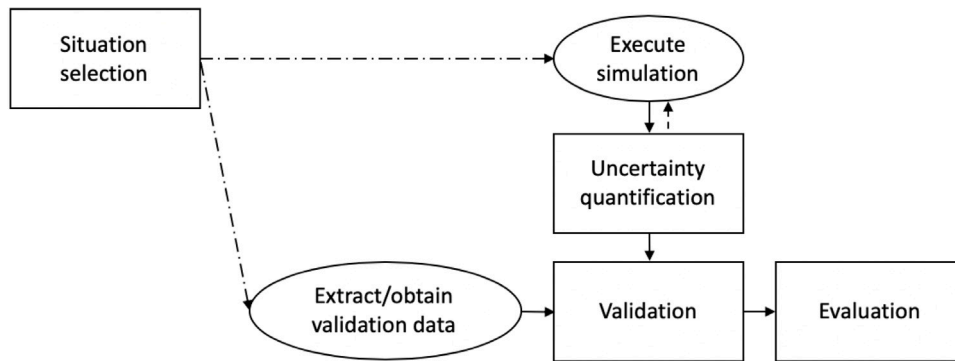
Fig. 1. SDA from a repeatability perspective.
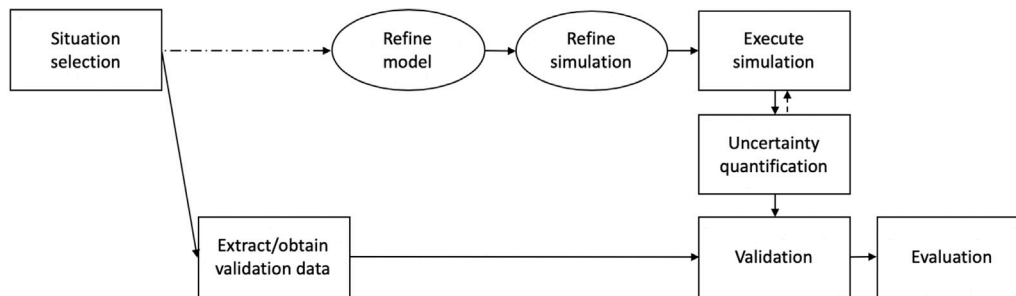

Fig. 2. SDA from a simulation reuse perspective.

counterparts) rely on a set of tools and techniques that are situation-agnostic. For example, a COVID-19 simulation of London could rely on an agent-based modelling code, an epidemic SEIR (Susceptible-Exposed-Infectious-Recovered) model, a compartmental model or a diffusion model, each of which could be used to model other COVID-19 spread situations as well. Indeed, each model type has its own advantages and limitations, and the initial choice for an appropriate situation-agnostic model is a non-trivial effort in its own right.

To incorporate the use and adaptation of situation-agnostic models, we add three tasks that are required to refine a model: (1) to extract or obtain input data (arguably this could already be necessary for simulation reuse), (2) to define or find a situation-agnostic model (this includes selecting the corresponding model class) and (3) to gather and curate situation-specific circumstantial evidence, which is needed to adapt the model to the specific situation. This evidence can be re-used at the simulation refinement stage, though one does require a situation-agnostic simulation that can be adapted (models normally do not automatically implement themselves). With the development work now explicitly resolved in the SDA, the refine simulation step also has a slightly broader scope now: it includes both the introduction of situation-specific parameters and rules, and the iterative testing, debugging and/or calibration of the situation-specific simulation code. With these changes, we then arrive at the high-level development perspective presented in Fig. 3.

The notion of having situation-agnostic components as well as situation-specific ones is important because situation-agnostic components do not have to be redeveloped when a new (crisis) situation emerges. To reflect this and add further detail to the tasks required, we present the SDA from a high- and low-level development perspective in Fig. 4. In this SDA, we added several tasks required to develop situation-agnostic models and simulations. One of these is problem definition, to determine what type of problems the model seeks to address. In this context, we define problems to be on a more general level (e.g., are we attempting to model traffic, storms, or the spread of airborne diseases for instance), while situations are more specific (e.g. modelling the COVID-19 pandemic in London in 2020). In other

words, a situation-agnostic model will be applicable to a range of situations that fall within the scope of the problem definition. In addition, these models will need a specification for what input data they might use, and against what metrics they could be validated. We represent this dependency with two-way arrows because new validation metrics or input data requirements may emerge as the situation-agnostic model is developed and improved. Last but not least, the situation-agnostic simulation should be verified, which technically means to check that the computational model accurately represents the underlying conceptual model and its solution [27]. Alternatively, in the context of this SDA, it means that we need to make sure that the implementation of the situation-agnostic simulation is behaving in a way that corresponds to the situation-agnostic (conceptual) model.

2.3. Accelerating Simulation Development

When examining our SDA from a high- and low-level development perspective (see Fig. 4), we distinguish between generic modelling tasks (those that do not need to be repeated when addressing a new situation) and simulation development and validation activities (those that do). Generic modelling tasks serve to provide a collection of tools and techniques that developers can integrate and adapt whenever a specific situation needs to be simulated (more on this in Section 3.1). Since the simulation development and validation tasks are situation-specific, accelerating these tasks delivers a measurable benefit towards rapid simulation development in response to a global challenge.

Within our SDA we distinguish ten simulation development and validation tasks, some of which can be accelerated and automated more readily than others. We summarize these simulation development and validation tasks, along with a non-exhaustive list of suggestions on how to accelerate and/or automate these tasks, in Table 1.

3. Simulation development for global challenges

Many of the global challenges today have a rapid and hard-to-extrapolate evolution that complicates the timely preparation of a community response. As a result, several new organizations have emerged
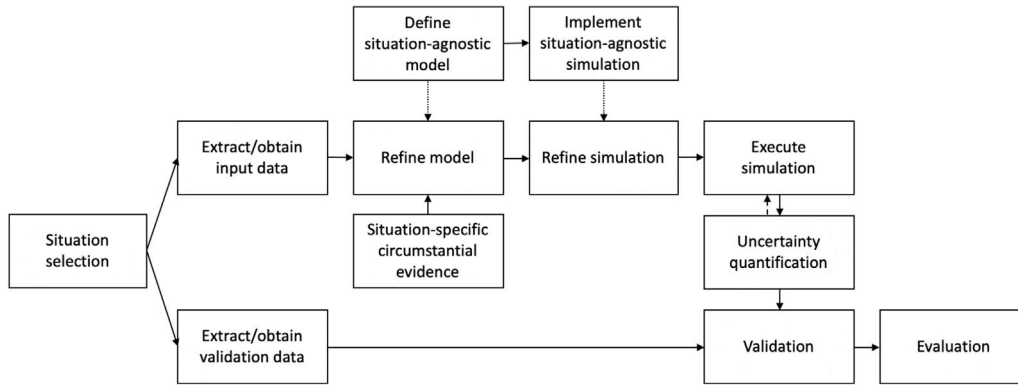
Fig. 3. SDA from a high-level simulation development perspective.
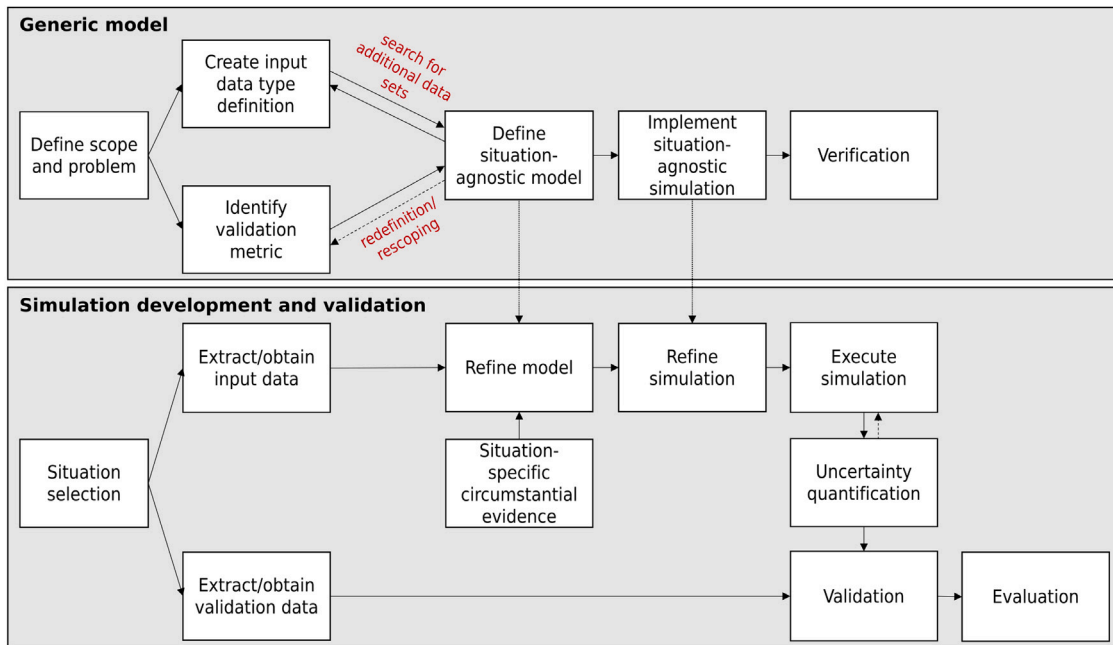

Fig. 4. SDA from a high- and low-level development perspective.

Table 1
Simulation development and validation tasks in the SDA, along with suggestions for ways to accelerate and automate them.

Task name	Ways to accelerate	Ways to automate
Situation selection	–	–
Extract/obtain input data	FAIR[a]	data extraction and web crawler scripts
Extract/obtain validation data	FAIR	data extraction and web crawler scripts
Refine model	human-readable input files	procedurally generated parameters
(Obtain) situation-specific evidence	MoU with stakeholders on report formats	document crawlers
Refine simulation	best practice adaptation guides	data-driven simulation
Execute simulation	parallel execution using high performance computing and/or job farming	adopt a workflow manager
Uncertainty quantification (UQ)	adopt a UQ toolkit	adopt an automation toolkit
Validation	shorthand commands for validation	continuous validation
Evaluation	MoU with stakeholders on reporting format	procedurally generated reports

[a] https://www.go-fair.org/fair-principles/

to support anticipatory actions, such as START[3] and the Anticipation Hub.[4]

Accurate forecasting simulations can inform both the response to global challenges and the anticipatory actions to prevent or mitigate them. For instance, simulation forecasts can help to guide the selection or development of mitigation policies, to inform the efficient allocation of humanitarian resources, or to justify to funding bodies that immediate funding is required. However, to fulfil any of these roles it is critical that these simulations are developed, validated, executed and disseminated in time.

Within this section, we specifically discuss the use of the simulation development approach in the context of anticipating global challenges, as well as responding to them. We also highlight a few key challenges around data collection and data use in this context.

3.1. Anticipatory context

We define the anticipatory context as the situation where a type of global challenge has been identified and recognized, but the actual events that would trigger a response have not (yet) occurred. During this period we can perform a range of anticipatory actions in the context of the SDA. These include in order of descending importance: problem definition, generic model development, infrastructure development, and anticipatory forecasts. The first two actions are explicitly captured in Fig. 4.

First, for problem definition, we sketch a representative range of global challenge scenarios that are being anticipated. For example, one could choose several flood scenarios with different intensities in several regions of Pakistan, five different trajectories for future pandemics in Western Europe, or four possible ways how a conflict could escalate within a given country. This problem definition should clarify what needs to be modelled, how and to what extent it will be validated, and it should inform the input data requirements, as well as the validation metrics for the situation-agnostic model. Problem definition is essential because it directly steers all other anticipatory actions in the SDA.

Second, the focus of generic model development is to actually develop this situation-agnostic model once the problem has been defined, and implement it as a flexible simulation. The purpose of this situation-agnostic simulation is to accelerate simulation development during the response context by providing a forecasting tool that can be rapidly adapted to specific situations. The design of a generic model may mismatch with the envisioned input data sets and validation metrics. For instance, developers may need to identify additional input data sets to improve the accuracy or completeness of the model, or they may need to redefine validation metrics if the simulation produces different output metrics than envisioned.

Third, infrastructure development includes efforts to accelerate and automate the SDA as a whole, as discussed in Section 2.3. It may also include more basic activities such as ensuring the availability of sufficient computing and storage capacity for doing on-demand forecasts or assembling a crisis team that is able to redirect development efforts at short notice.

Fourth, anticipatory forecasts involve choosing specific situations within the scope of the problem definition that are believed to be likely to happen and then performing simulation development and forecasting for those situations. Now the SDA for forecasting purposes is slightly different to the SDA for validation purposes, and we present it in Fig. 5 (leaving out the generic model development aspects for simplicity). The essential difference is that direct validation is not possible, because observations only emerge after the forecast has been performed. Once the observations are available, after the forecast has been used, the forecasting results can be validated and evaluated.

3.2. Response context

We define the response context as the situation where an actual crisis event has occurred, and a situation-specific forecast is urgently required. It is worth noting here that saturation and turning points of time-dependent forecasting curves are notoriously hard to forecast, as their accurate estimation often requires data points that are not available in the 'early' phase of the process. Nevertheless, even then, a situation-specific forecast may still deliver important additional insights that can inform decision-making.

Developers may be able to rely on efforts undertaken in the anticipatory context, such as clear definitions of the forecasting problem, generic models and simulation tools, available computational and human infrastructure, and/or relevant anticipatory forecasts.

In a response context forecasting is one under time pressure, and generic or anticipatory simulation development tasks will be avoided if at all possible. Because of this, Fig. 5 presents an accurate and reasonably complete SDA from a forecasting perspective in a response context. Here the key objective is to establish an accurate and relevant forecast, delivered on time to the responding organizations.

3.3. Data collection and global challenge simulation development

Within the SDA, there are a number of tasks that rely on external data. These naturally include data extraction tasks, the task of obtaining situation-specific evidence, the model and simulation refinement tasks as well as the validation task. Estimating the effort required to obtain and apply such external data brings with it additional uncertainties for a variety of reasons. For example, data sources may be (a) unavailable for the specific task, (b) more difficult to find than expected, (c) producing data that is less complete or more biased than expected, (d) producing data that is in an unexpected or inconsistent format or (e) producing data that is not widely accepted as a ground truth. Fortunately, many simulation approaches can still be applied in the face of imperfect data, although their forecasting accuracy may be reduced and additional effort may be required to mitigate data issues. Another, more specific issue, is the use of incomplete, biased or noisy validation data. Model outputs that are compared against such imperfect validation data produce error rates that, while still informative, do not fully correspond to the mismatch between simulation and reality. In these cases, it is particularly important not to put too much stock in validation performance, and avoid automatically optimizing or calibrating models to achieve a low (and likely inaccurate) validation error score.

4. Example applications

Although we have not previously created a comprehensive generic description of the SDA, we have used the concept internally for two types of global challenges since it was first introduced in 2017 [6]. Here we discuss these applications, in conflict-driven migration and disease spread, and explain how we applied the SDA concepts to help facilitate more rapid simulation development in a global challenge response context.

Of course, forecasting under time pressure has risks associated with it. For instance, models may become less detailed, less deeply scrutinized and/or less accurate than intended. In addition, there is an increased risk for human mistakes, as well as unknown side effects that may only manifest itself after results have been reported. We argue that these risks should be clearly acknowledged and weighed against the expected benefit of the simulation-driven forecast relative to the existing foresight. In the case of conflict-driven migration, our approach has a track record [6,28] in creating reasonably accurate numerical arrival forecasts without the need for training data sets. In the case of disease spread, our model was developed because at the time there was no alternative approach to predict expected intensive care admissions for a specific hospital during a pandemic wave.
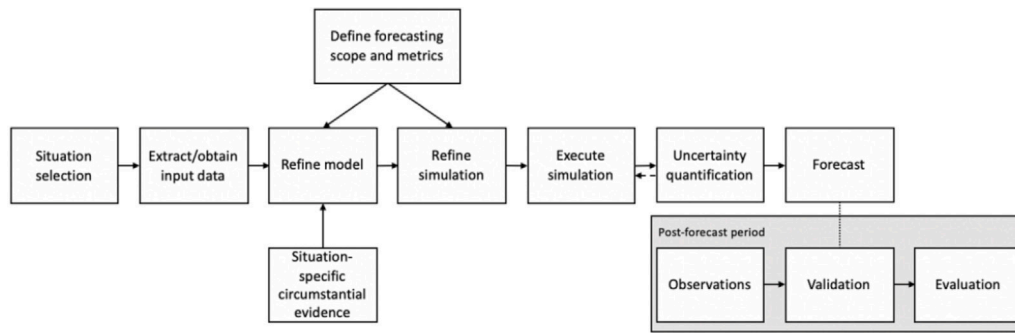

Fig. 5. SDA from a forecasting perspective.

4.1. Conflict-driven migration

Armed conflicts are commonplace nowadays, and the number of forcibly displaced people now exceeds 100 million as a result of that.[5] Forecasts that predict where persons displaced by violence may arrive, before their actual arrival, can inform the preparation of refugee camps by humanitarian organizations, or help use their (often limited) aid budget more effectively.

Within our group, we have developed an agent-based modelling code, named Flee (not an acronym), which is specifically suited for modelling conflict-driven migration (see Suleimenova et al. [6] for a detailed description of the code and the associated SDA). The code relies on the representation of persons as autonomous agents, with the spatial environment represented as a graph where camps, towns and conflict zones are represented as vertices. We validated this code initially against three African conflicts (in Mali, Burundi and Central African Republic), followed by a second validation study in the context of South Sudan where we tested several automation approaches [29], as well as a sensitivity analysis study across four conflicts [30].

In November 2020, we conducted a trial in simulation construction of Flee in the context of the Tigray conflict in Ethiopia, in collaboration with Save The Children.[6] Here, domain experts from Save the Children gave essential input about the scope and requirements of the forecast, and provided descriptions of how the conflict could possibly evolve. We were initially given six weeks to develop a prototype simulation of conflict-driven migration in and around the Tigray region in Ethiopia. This led to the submission of the first forecasting report on December 18th 2020, followed by three more reports in 2021. In this example, we needed to adapt an existing solver (Flee) for this new context, so we performed this work from a high-level simulation development perspective. In this paper, we focus on the simulation development aspects, but the scientific results of these runs are discussed by Suleimenova et al. [28].

We present our initial SDA, which we followed for the first forecasting report for Tigray, in Fig. 6A. In this figure, we provide a time estimate for each of the steps and provide a list of subtasks below each of the steps along with coarse time estimates. When preparing the initial report, we were struggling with the time deadline mainly because four tasks turned out to be particularly time intensive: First, creating the location graph involved a large amount of manual work. Second, we needed to generate viable conflict scenarios, but that also were detailed enough to be used by Flee. This required us to develop a dedicated script to perform this. Third, due to the high agent counts, simulations took relatively long to complete on local resources. And fourth, the additional executions required for uncertainty quantification likewise required a large amount of time. In the end, we did meet

the six week deadline, but the initial report lacked important detail in the area of uncertainty quantification (see Fig. 7). One aspect worth discussing is to what extent a larger team size could have helped us to perform the simulation development more quickly. As can be seen in Fig. 6, every phase of the SDA has major dependencies on the outputs of the previous phase. However, the tasks performed within each phase could have been done in parallel if we had a larger team. For instance, one person could be assigned to work on demographic data while another could focus on location graph extraction. Performing a single simulation development task, such as location graph extraction, with multiple persons is also an option, but the effectiveness of that is not guaranteed as it will depend on the nature of the task, the simulated situation, and the skill sets of the team members.

To improve our ability to compile future reports, we accelerated these four activities by incorporating automation techniques (see Fig. 6B). For the location graph automation, we used the techniques presented by Schweimer et al. [31], while for automating simulation execution and uncertainty quantification we used several components of the VECMA toolkit (nowadays known as the SEAVEA toolkit) [32]. The hardest part to automate was the work to generate conflict progressions. Because we were dealing with hypothetical scenarios, we asked Save The Children to describe these scenarios. We then created a script with randomization techniques to generate variable conflict progressions that were in accordance with the desired scenario type. This script allowed us to rapidly generate for instance 100 different conflict progressions under the assumption that a conflict would flare up in the West of Tigray. This automation approach is not ideal, however, as the script may not be easy to reuse for future crises situations.

4.1.1. Current status and next steps

With the automation in place, we found ourselves able to cut down the simulation development time from about 47 days to 14 days (see Fig. 6C). This makes our SDA still slightly too slow in an actual crisis situation, as ideally we would want to be able to complete simulation development and generate a forecast within a week of a conflict erupting. To reach this goal, we will first need to make a more flexible and easy-to-reuse version of the scenario-based conflict generation script. In addition, two new bottlenecks have emerged in the SDA: the refine simulation step and the uncertainty quantification (UQ) step. We are attempting to accelerate simulation refinement by preparing a new 3.0 release of Flee which features a much wider range of user-configurable parameters. This will accelerate simulation refinement because more modifications can then be done without rewriting the source code. As for UQ, although we have usable automated UQ scripts, the next step here could be to integrate them more tightly with FabSim3 [33], so that UQ is done on the fly whenever forecasting runs are done. If these two improvement efforts are successful, then we would become able to develop forecasts within an estimated time of approximately 7 days.
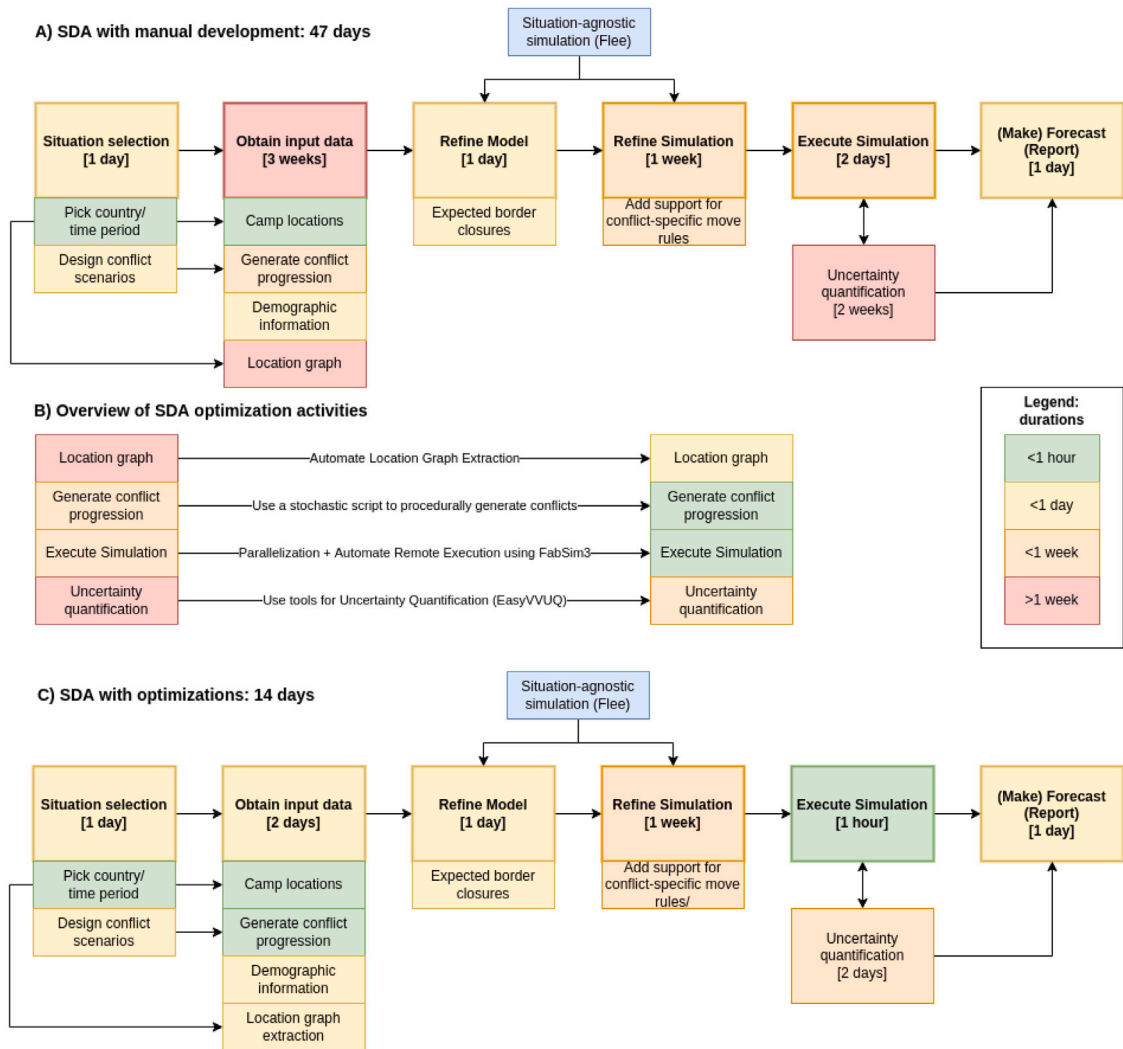

Fig. 6. SDA for forecasts of conflict-driven migration in Tigray along with time estimations based on our development experience, without automation (A) and with automation in place (C). (B) overview of optimization/automation steps performed. Note that time spent during the "execute simulation" phase is mostly spent waiting, and could be partially overlapped with other tasks.

4.2. Spread of COVID-19

The SARS-CoV-2 virus emerged in late 2019 and quickly spread worldwide to cause a pandemic of COVID-19 disease. At the time of writing, over 6 million people have been confirmed to have died from COVID-19, with many more suffering from long-term health problems. During the initial phase of the pandemic, in early 2020, there were several national-level forecasting models available that helped inform governments about the effectiveness of non-pharmaceutical interventions. However, reliable forecasts on the level of a hospital catchment area were generally not available, leading to uncertainty amongst hospital management boards about how to allocate intensive care capacity and how to adjust their long-term care strategy in response to the pandemic.

In March 2020, we developed a localized COVID-19 model, named the Flu And Coronavirus Simulator (or FACS) [34], after several UK National Health Service (NHS) Hospital Trusts in London approached us with this need. Similar to Flee, this model represents persons as autonomous agents that are scheduled to visit a variety of locations, such as hospitals, schools, offices and shops, for each simulated day. Mahmood et al. [34] provide an overview of the core assumptions and modelling approaches available in FACS, as we used it in 2020 (the code has since been heavily updated and a new paper is in preparation).

Using this prototype model, we made a total of fifteen forecasting reports, between April and December 2020, for hospitals in Brent, Harrow, Ealing and Hillingdon boroughs in London, UK. Here, domain experts from the NHS formulated the requirements for each report, provided feedback about the retrospective quality of our forecasts, and provided corrective feedback on underlying model assumptions when these were deemed to be unrealistic from their perspective. After April, a number of other simulation codes emerged that could address this problem [35,36], though to this day the FACS code remains relatively quick to deploy and produces different forecasts than these alternatives.

We present our initial SDA, which we followed for the initial forecasting reports for the NHS Trusts, in Fig. 8A. At that stage, very little work was automated and in a single-person development setup, we estimate that the simulation would have taken 53 days to develop. In practice, we managed to deliver the first report within approximately a month due to the kind support from a range of colleagues in obtaining the disease information and vetting the location graph, as well as doing very minimal uncertainty quantification during that early stage, and not (yet) having to take into account vaccinations.

During the course of the project, we identified five areas where we could accelerate our simulation development. These are presented in Fig. 8B. Three of them are already covered in our earlier example, but two additional ones are unique to this application. First, we developed
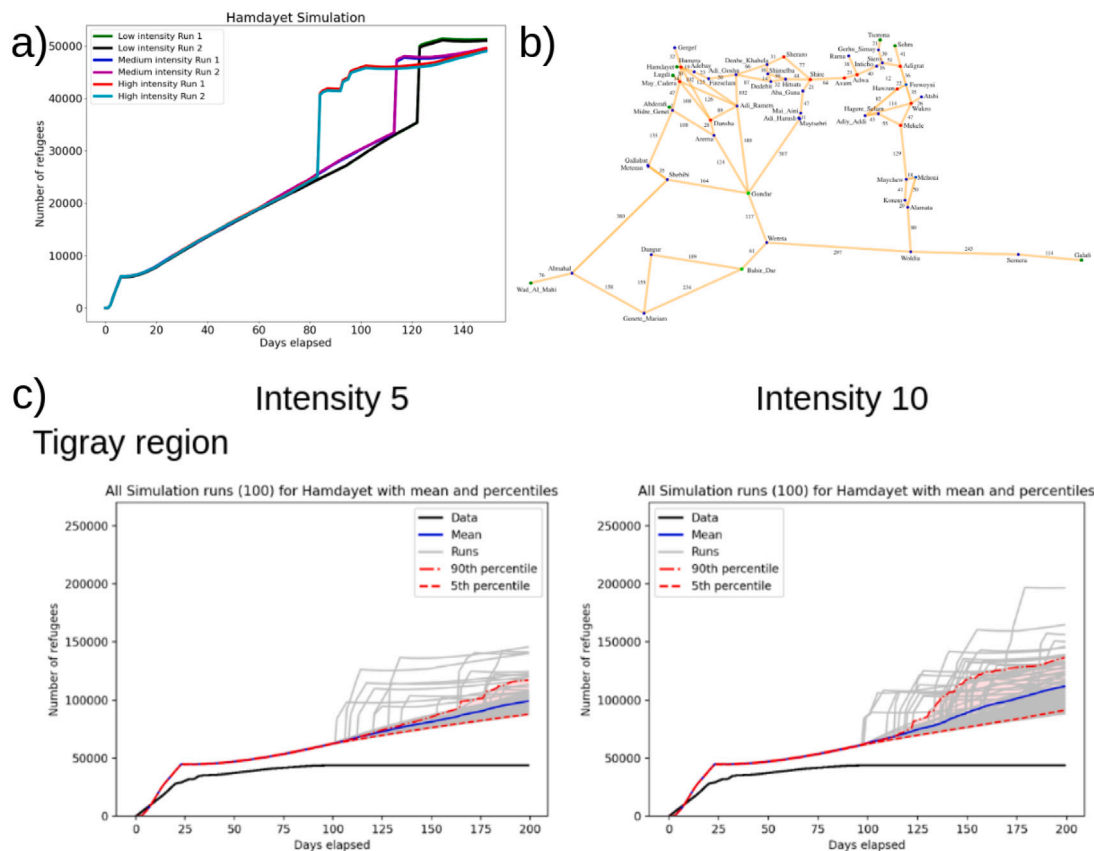
Fig. 7. Sample forecast presentation of refugee arrivals in the Hamdayet camp in Sudan from the first report (a), and the third report (c). The initial version of the location graph is given in (b). Reprinted and reformatted from an internal report with permission from Save The Children.

a shared disease definition, so that all groups can rely on common knowledge about the infectious characteristics of COVID-19 and its variants. By storing this information in a .YML (Yet another Markup Language) file, people are able to scrutinize these assumptions and adapt them easily for their own variant simulation workflows. Disease-specific characteristics only vary to a limited extent between locations in the case of COVID, so this greatly accelerated that particular step for later reports.

The second optimization was actually performed in the summer of 2022 (Incorporate non-pharmaceutical intervention strategy task), after we had sent all the reports to the NHS. We had initially hard-coded all the interventions in FACS, which meant that the code was catered specifically for outbreaks in parts of London. When we needed to repurpose FACS for use in Turkey, Romania and Lithuania as part of the STAMINA project,[7] this became highly impractical. Instead, we developed a flexible YML-based system for defining non-pharmaceutical interventions (see Fig. 9 for an example). Although the underlying task is still manual, we did manage to reduce the development time from approximately 3 days to under a day. This was because the use of structured, human-readable input files instead of hard-coded measures led to fewer bugs, and less time spent debugging as a result.

4.2.1. Current status and next steps

With these optimizations in place, we are now able to make forecasts using FACS within approximately 14 days. In an ideal case, we would like to make rigorous optimizations to perform forecasts within 5 days, which is the timescale that many government advisory groups operate on. However, as we can see in the current SDA (see Fig. 8C), there are two tasks that prevent us from doing so: obtaining input data

(in particular the location graph), and uncertainty quantification (UQ). UQ can further be accelerated by aggressively scheduling many jobs on a supercomputer, using for instance advanced job packing tools like QCG-PilotJob [37], and by having resources for it readily available (using reservations or urgent computing [21] if necessary).

The extraction of location data is more complicated to accelerate. The COVID-19 spread application normally relies on the location extraction of 100,000s of buildings, a task which is now performed using automated extraction tools that integrate with OpenStreetMap.[8] However, the building annotation in OpenStreetMap is inconsistent across different locations, and sometimes even within individual towns, which leads to extraction errors and artefacts in the input files. Detecting these artefacts is a process that could arguably be automated, but correcting these artefacts is likely to remain a manual process that may require inspection of satellite imagery, searching online resources or even physically investigating relevant locations. Therefore, unless we either restrict the geographic scope of the application and/or undertake a large annotation exercise on OpenStreetMap, it is unlikely that we can accelerate the location extraction by much more. Yet we can resolve this bottleneck by building up a database of extracted locations as an anticipatory measure, instead of extracting buildings only after a crisis erupts. This would effectively eliminate this location extraction from the SDA in a global challenge response context, and allow us to develop simulations within approximately 5 to 6 days, at the expense of somewhat increased anticipatory effort.

5. Discussion

In this work, we presented a generic simulation development approach (SDA), and showcased its application to two global challenge
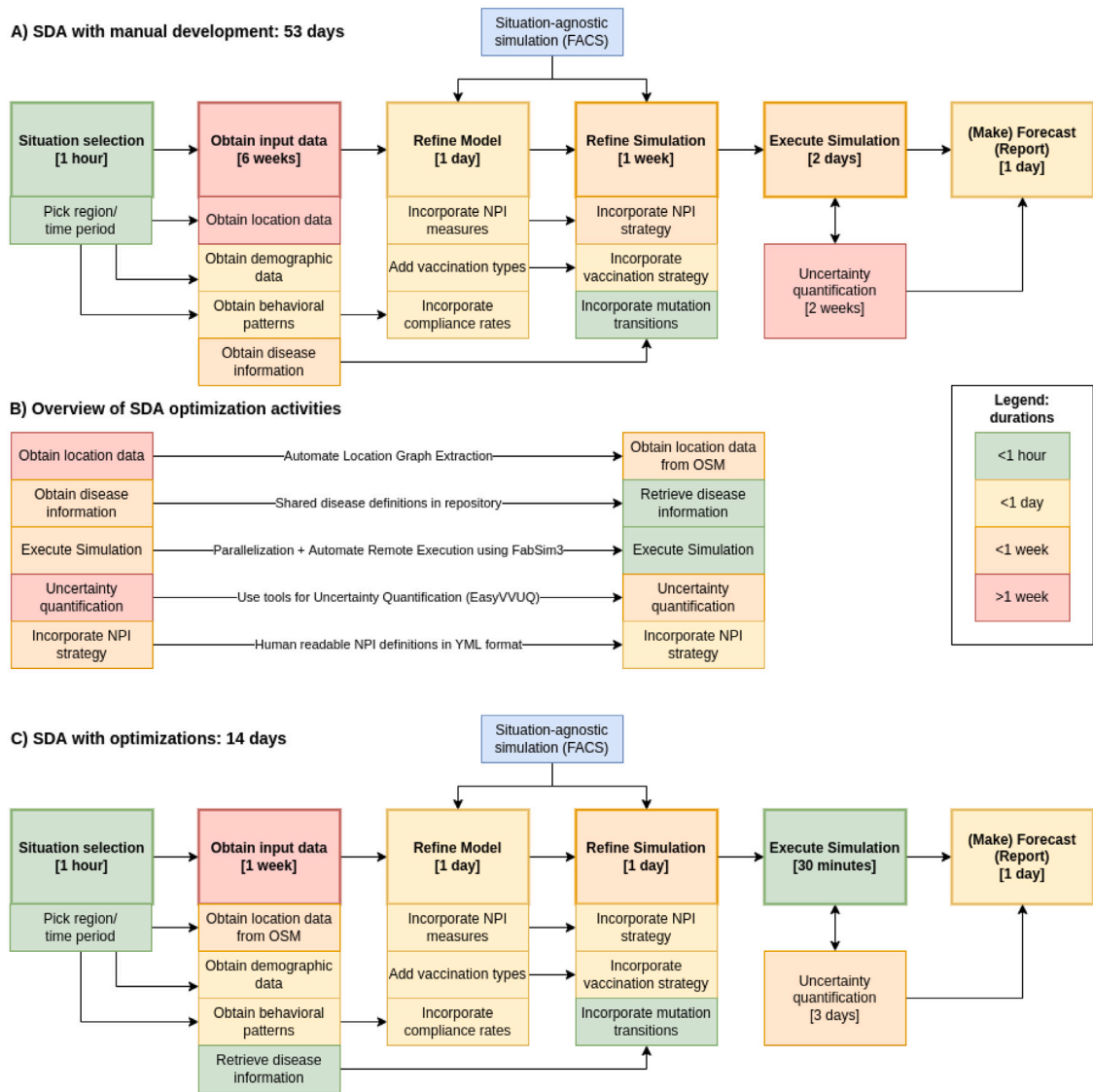
Fig. 8. SDA for forecasts of COVID-19 spread in the Brent Borough in London, along with time estimations based on our development experience, without automation (A) and with automation in place (C). (B) Overview of optimization/automation steps performed. The acronym NPI stands for non-pharmaceutical intervention. Note that time spent during the "execute simulation" phase is mostly spent waiting, and could be partially overlapped with other tasks.

```
44    23/3/2020:
45      partial_closure:
46        school: 1.0 # Full closure, except for key workers.
47        shopping: 0.7 # Non-essntial shops (70%) closed.
48      closure: ["leisure"]
49      social_distance: 0.5
50      work_from_home: 0.55
51      traffic_multiplier: 0.25 # Traffic at 25% of normal levels.
52      hospital_protection_factor: 0.8 # Hygiene -> -80% transmission.
```

Fig. 9. YML-defined non-pharmaceutical interventions for FACS: we used this method to speed up simulation development for the *Incorporate non-pharmaceutical intervention strategy* task.

problems. We note that simulations can be used for a wide range of purposes [38], for instance, to check the validity of a new theory or to impute missing data values. The SDA as we present it however is purpose-specific, in that is intended for use in anticipation of global challenges as well as in the response. At the same time, our SDA is generic in terms of application type, i.e. it can be used for any type of model that is suitable for this context.

Our approach distinguishes anticipatory activities, required for developing application-agnostic modelling tools, from activities that need to be undertaken when a crisis hits and a forecast is imminently

required to inform the crisis response. We show that the SDA can be used to systematically capture tasks required in both contexts and to help researchers and responders identify which steps become bottlenecks in these situations. These bottlenecks can then be addressed in various ways, e.g. through workflow automation, developing additional pre- and post-processing tools or optimizing computations for faster execution.

We have demonstrated the added insights provided by the SDA in two real-world contexts: conflict-driven migration modelling, in collaboration with a non-governmental organization, and local COVID-19 spread modelling, in collaboration with a NHS hospital trust. In both cases, we have used the SDA to understand the full development process, identified the main bottlenecks and optimized these time-consuming steps (using e.g. automation). Through this exercise, we were able to accelerate simulation development by a factor of 3 to 3.5 in both cases, although further acceleration is required in both contexts to make the simulation development rapid enough to support a direct crisis response.

We learned several major lessons when defining the SDA in a global challenge context, and applying it to real-world problems. First, by developing the SDA we are able to see the role of high performance computing (a highly active field) from the perspective of simulation development at large. A faster executing simulation mainly speeds up simulation development because it proportionally reduces the time to perform uncertainty quantification. This is somewhat ironic because a substantial fraction of high performance computing simulations are actually performed and presented without any degree of uncertainty quantification, leaving the reader of such papers to guess whether the results are robust or spurious.

Second, in both our exemplars we find that obtaining input data is a primary bottleneck in our SDAs, and in both cases we accelerate this task by using automated extraction tools. Generating data-derived initial conditions becomes complex when the underlying data is incomplete, biased or inconsistent, and aside from dedicating more effort to pre-processing tools one could also consider annotating the data with corrections and imputations. The latter is particularly helpful when many applications depend on the same data source (as is for instance the case with OpenStreetMap).

Third, we give concrete examples of how work performed in the anticipatory context will lead to saved time in the response context. However, it becomes a problem to justify anticipatory effort when global challenges do not occur regularly. As the report on the Eyjafjallajökull eruption in 2010 notes: "When time passes and the last event becomes an increasingly distant memory, it is harder to draw stakeholders to the table to participate in possibly costly exercises and contingency planning". [3]. Within the academic community, we may be able to strengthen our anticipatory work by acknowledging this bias and adjusting our research agendas accordingly where possible.

Through our discussions both with NGOs and healthcare providers, it has become clear that simulation-based forecasts can (i) provide additional relevant information on future developments and (ii) help estimate the impact of preventative or mitigating actions in case emergencies arise. This information can be of use in the human-driven decision-making process when handling emergency situations, but cannot and should not drive emergency decision-making directly. The human experience, contextual knowledge, interconnection and ability to scrutinize results are absolutely fundamental in the decision-making process. In addition, there are major ethical, practical, moral and legal hazards that are associated with fully automated decision making in high-risk settings [26].

In terms of general experiences around the migration case study, i.e., which areas of the process worked well and which ones did not, an in-depth analysis of our group activities has been performed by Nandi [39], who monitored our group's activities for over a year. For the COVID-19 study, we unfortunately did not have such in-depth analysis. However, in general we found that the most challenging aspect

there was to align our research group activities with other COVID-19 modelling efforts in the UK, due to misaligned objectives and sometimes in-transparent structures for research collaboration. In addition, the severe funding reductions in epidemiological research in the UK post-pandemic complicated our efforts along with many other COVID modelling groups. What did work well for us in the COVID-19 use case were (i) the interactions with local NHS trusts, who communicated clearly, reviewed our work rigorously and were generally responsive, (ii) the willingness of existing research consortia to adapt their research to the emergence of COVID-19 and [this occurred both in the HiDALGO and STAMINA EU-funded projects] and (iii) the willingness of many colleagues to voluntarily contribute to the development and testing of the code, particularly early in the pandemic.

In terms of future research directions, we believe that priorities include to (i) establish a much larger scale automated validation environment for simulations and other forecasting tools to be used in this context, (ii) clarify and sensibly address any challenges around moral, ethical, political, bureaucratic or financial obstacles that could prevent the useful application of these tools, and (iii) further accelerate the SDA for our two case studies, such that reliable forecasts can be made possible within a single week.

Lastly, it is our view that end-to-end simulation development research in general (not only our proposed Simulation Development Approach) directly benefits the computational science community, and warrants a higher priority than it has today. We hope that future work on this topic will lead to new conceptual tools and methodologies that find widespread uptake, and someday supersede the SDA.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

References

[1] N. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, et al., Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, Imp. Coll. London 10 (77482) (2020) 491–497.

[2] N. Ferguson, G. Nedjati Gilani, D. Laydon, Covid-19 CovidSim microsimulation model, 2020, URL https://github.com/mrc-ide/covid-sim.

[3] G. Petursdottir, U. Reichardt, D. Bird, A. Donovan, G. Gísladóttir, A. Hauksdottir, G. Johannesdottir, F. Sigmundsson, E. Thordardottir, G. Ulfarsson, Eyjafjallajökull eruption in 2010, in: Science for Disaster Risk Management 2020: Acting Today, Protecting Tomorrow, EUR 30183 EN, Publications Office of the European Union, 2020, http://dx.doi.org/10.2760/571085.

[4] S. Millington, R. Saunders, P. Francis, H. Webster, Simulated volcanic ash imagery: A method to compare NAME ash concentration forecasts with SEVIRI imagery for the Eyjafjallajökull eruption in 2010, J. Geophys. Res.: Atmos. 117 (D20) (2012).

[5] D. Groen, Simulating refugee movements: Where would you go? Procedia Comput. Sci. 80 (2016) 2251–2255.

[6] D. Suleimenova, D. Bell, D. Groen, A generalized simulation development approach for predicting refugee destinations, Sci. Rep. 7 (1) (2017) 1–13.

[7] S. Robinson, Conceptual modeling for simulation, in: J. Cochran (Ed.), Encyclopedia of Operations Research and Management Science, Wiley, New York, 2011.

[8] I. Nikolic, A. Ghorbani, A method for developing agent-based models of socio-technical systems, in: International Conference on Networking, Sensing and Control, The Netherlands, 2011, pp. 44–49.

[9] O. Balci, Validation,verification and testing techniques throughout the life cycle of a simulation study, Ann. Oper. Res. 53 (1994) 121–173.

[10] J. Banks, J. Carson, B. Nelson, D. Nicol, Discrete-Event System Simulation, fifth ed., Prentice Hall, Upper Saddle River, NJ, 2009.

[11] J. Augusiak, P. van den Brink, V. Grimm, Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach, Ecol. Model. 280 (2014) 117–128.

[12] V. Grimm, S. Railsback, Individual-Based Modeling and Ecology, Princeton University Press, Princeton, NJ, 2005.

[13] A. Law, Simulation Modeling and Analysis, fourth ed., McGraw-Hill, New York, 2007.

[14] D. Sturrock, Tutorial: tips for successful practice of simulation, in: C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, A. Uhrmacher (Eds.), Proceedings of the 2012 Winter Simulation Conference, IEEE, Piscataway, NJ., 2012, pp. 354–361.

[15] B. Heath, R. Hill, F. Ciarallo, A survey of agent-based modelling practices (January 1998 to July 2008), J. Artif. Soc. Soc. Simul. 12 (4, 9) (2009).

[16] S. Robinson, Conceptual modelling for simulation part I: Definition and requirements, J. Oper. Res. Soc. 59 (2008) 278–290.

[17] P.V. Coveney, E.R. Dougherty, R.R. Highfield, Big data need big theory too, Phil. Trans. R. Soc. A 374 (2080) (2016) 20160153.

[18] J.H. Kwakkel, E. Pruyt, Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty, Technol. Forecast. Soc. Change 80 (3) (2013) 419–431.

[19] E. Pruyt, From modelling uncertain surprises to simulating black swans, in: Proceedings of the 33-rd International Conference of the System Dynamics Society, Cambridge, Massachusetts, USA, 2015.

[20] L. Freedman, Scientific advice at a time of emergency. SAGE and Covid-19, Political Q. 91 (3) (2020) 514–522.

[21] S.H. Leong, D. Kranzlmüller, Towards a general definition of urgent computing, Procedia Comput. Sci. 51 (2015) 2337–2346.

[22] B. Blanton, J. McGee, J. Fleming, C. Kaiser, H. Kaiser, H. Lander, R. Luettich, K. Dresback, R. Kolar, Urgent computing of storm surge for North Carolina's coast, Procedia Comput. Sci. 9 (2012) 1677–1686.

[23] T. Simmermacher, G. Tipton, J. Cap, R. Mayes, The role of model V&V in the defining of specifications, in: Model Validation and Uncertainty Quantification, Volume 3: Proceedings of the 33rd IMAC, A Conference and Exposition on Structural Dynamics, 2015, Springer, 2015, pp. 257–263.

[24] R.G. Sargent, Verification and validation of simulation models: An advanced tutorial, in: 2020 Winter Simulation Conference, WSC, 2020, pp. 16–29, http://dx.doi.org/10.1109/WSC48552.2020.9384052.

[25] C.J. Roy, W.L. Oberkampf, A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, Comput. Methods Appl. Mech. Engrg. 200 (25–28) (2011) 2131–2144.

[26] A. Guillén, E. Teodoro, Embedding ethical principles into AI predictive tools for migration management in humanitarian action, Soc. Sci. 12 (2) (2023) 53.

[27] W.L. Oberkampf, C.J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, 2010.

[28] D. Suleimenova, W. Low, D. Groen, An agent-based forced displacement simulation: A case study of the Tigray crisis, in: International Conference On Computational Science, Springer, 2022, pp. 83–89.

[29] D. Suleimenova, D. Groen, How policy decisions affect refugee journeys in South Sudan: a study using automated ensemble simulations, JASSS (2020).

[30] D. Suleimenova, H. Arabnejad, W. Edeling, D. Groen, Sensitivity-driven simulation development: a case study in forced migration, Phil. Trans. R. Soc. A 379 (2197) (2021) 20200077.

[31] C. Schweimer, B.C. Geiger, M. Wang, S. Gogolenko, I. Mahmood, A. Jahani, D. Suleimenova, D. Groen, A route pruning algorithm for an automated geographic location graph construction, Sci. Rep. 11 (1) (2021) 1–11.

[32] D. Groen, H. Arabnejad, V. Jancauskas, W. Edeling, F. Jansson, R.A. Richardson, J. Lakhlili, L. Veen, B. Bosak, P. Kopta, et al., VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations, Phil. Trans. R. Soc. A 379 (2197) (2021) 20200221.

[33] D. Groen, H. Arabnejad, D. Suleimenova, W. Edeling, E. Raffin, Y. Xue, K. Bronik, N. Monnier, P.V. Coveney, Fabsim3: an automation toolkit for verified simulations using high performance computing, Computer Physics Communications 283 (2023) 108596.

[34] I. Mahmood, H. Arabnejad, D. Suleimenova, I. Sassoon, A. Marshan, A. Serrano-Rico, P. Louvieris, A. Anagnostou, S. JE Taylor, D. Bell, et al., FACS: a geospatial agent-based simulator for analysing COVID-19 spread and public health measures on local regions, J. Simul. 16 (4) (2022) 355–373.

[35] J. Aylett-Bullock, C. Cuesta-Lazaro, A. Quera-Bofarull, M. Icaza-Lizaola, A. Sedgewick, H. Truong, A. Curran, E. Elliott, T. Caulfield, K. Fong, I. Vernon, J. Williams, R. Bower, F. Krauss, June: open-source individual-based epidemiology simulation, Royal Soc. Open Sci. 8 (7) (2021) 210506, http://dx.doi.org/10.1098/rsos.210506, URL https://royalsocietypublishing.org/doi/abs/10.1098/rsos.210506.

[36] J. Ozik, J.M. Wozniak, N. Collier, C.M. Macal, M. Binois, A population data-driven workflow for covid-19 modeling and learning, Int. J. High Perform. Comput. Appl. 35 (5) (2021) 483–499.

[37] D. Groen, H. Arabnejad, V. Jancauskas, W. Edeling, F. Jansson, R.A. Richardson, J. Lakhlili, L. Veen, B. Bosak, P. Kopta, et al., VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations, Phil. Trans. R. Soc. A 379 (2197) (2021) 20200221.

[38] J.M. Epstein, Why model? J. Artif. Soc. Soc. Simul. 11 (4) (2008) 12.

[39] A. Nandi, A process based framework for evaluating humanitarian technologies: Case study of an agent-based simulation for forced displacement, 2022, Available at SSRN 4286445.

Dr. Derek Groen is a Reader in Computer Science at Brunel University London, and a Visiting Lecturer at University College London. He has a Ph.D. from the University of Amsterdam (2010) in Computational Astrophysics, and was a Post-Doctoral Researcher at UCL for five years prior to joining Brunel as Lecturer. Derek has a strong interest in high performance simulations, multiscale modelling and simulation, and so-called VVUQ (verification, validation and uncertainty quantification). In terms of applications he is a lead developer on the Flee migration modelling code and the Flu And Coronavirus Simulator (FACS) COVID-19 model. He has also previously worked on applications in astrophysics, materials and bloodflow. Derek has been PI for Brunel in two large and recent Horizon 2020 research projects (VECMA on uncertainty quantification, and Hi-DALGO on global challenge simulations) and he is currently the technical manager of the UK-funded SEAVEA project which develops a VVUQ toolkit for large-scale computing applications (seavea-project.org).

Dr. Diana Suleimenova is a Lecturer in Computer Science. She received her Ph.D. in 2019 and has publications in Nat. Comput. Sci., Phil. Trans. R. Soc. and Adv. Theory and Simul. She has developed VVUQ approaches for a variety of domains, including the Flee migration model [K4]. She is co-Chair of the Multiscale Modelling and Simulation workshop and was an invited speaker at several international organizations. Her work has led to a joint project with Columbia University on applying Markov-Chain Monte Carlo techniques and a two-way collaboration with the Save The Children on developing and analysing the sensitivities of new migration models for the Tigray Conflict in Ethiopia.

Dr. Alireza Jahani specializes in Agent-Based Modelling, Multi-Agent Systems, Machine Learning, and Knowledge Management. Before joining Brunel University London, he was the Deputy of Technical and Information Services and Assistant Professor at Faculty of Information Technology, Mehralborz University (MAU), Iran. He obtained his Ph.D. in Computer Science from University Putra Malaysia (UPM) in 2014 and he received his Master of Science in e-Commerce from Iran University of Science and Technology in 2008. Alireza is the main developer of DFlee, which is an extension of the Flee code to facilitate the modelling of disaster evacuation.

Dr. Yani Xue is a Lecturer at Brunel University London. She received a Ph.D. degree in Computer Science from Brunel University London in 2021, and currently works on many-objective optimization in the context of migration modelling through the EU-funded ITFLOWS project. Her main research interests include multi/many-objective optimization, evolutionary computation, search-based software engineering, engineering applications, simulation, and high performance data analytics.