TOWARDS MODELLING THE EFFECT OF EVOLVING VIOLENCE ON FORCED MIGRATION

Derek Groen
David Bell
Hamid Arabnejad
Diana Suleimenova
Simon J. E. Taylor
Anastasia Anagnostou

Department of Computer Science
Brunel University London
Uxbridge, UB8 3PH, nUK

ABSTRACT

We present a modelling approach to investigate the evolution of violent events on the forced displacement of people in affected countries. This work is performed in the context of the EU-funded HiDALGO Centre of Excellence, where we seek to establish more scalable and accurate models for migration prediction. Such a hybrid simulation approach is necessary as we need to understand how conflicts may evolve if we are to forecast the escape of people from future conflicts. To accomplish this, we couple a new model for conflict propagation with existing forced migration models. We explore the implications of our setup by studying the effect of different conflict progressions on the forced displacement of people, using an established model of the Mali 2012 conflict. We conclude that accurately predicting conflict evolutions is a key determinant in the outcomes of forced migration simulations, particularly if such forecasts are made over longer periods.

1 INTRODUCTION

Today, around 68.5 million people have been forcibly displaced across the world, according to the UNHCR (https://www.unhcr.org/uk/figures-at-a-glance.html). Around 80% of these people reside in developing countries, where humanitarian resources are limited and monitoring can be intermittent. Forecasting the movements of these people, who are frequently displaced due to violence in their home country, has the potential to inform governments and NGOs on how to allocate humanitarian support (Edwards 2008; Groen 2016). In addition, such forecasts may provide estimates of arrivals in regions where it is difficult to monitor or register the arrival of displaced persons, and may be able to inform stakeholders about the potential consequences of policy decisions such as border or road closures. A number of forced migration forecasting tools have emerged in recent years, for instance to examine migration in Syria (Hébert et al. 2018), arrivals in the DRC (Frydenlund et al. 2018), or inter-provincial movements in Vietnam (Nguyen et al. 2018). However, all such forecasts come with a vast range of uncertainties, and the wide range of possible reasons for migration makes the development of accurate, robust and widely re-applicable models scientifically very challenging (Willekens 2018).

We have previously proposed a Simulation Development Approach (SDA) to enable the creation and validation of forced displacement (or forced migration) simulations, specifically in the context of people escaping a civil war in search for safety abroad (Suleimenova 2017). We tested this approach across three different conflicts, and managed to predict more than 75% of the destinations correctly in these cases. Although this work enabled us to validate our approach against existing situations, and we have accelerated the simulation construction through automation (Suleimenova 2017b), there are a number of elements missing to enable the forecast of population displacement in future conflicts. For instance, we have not yet established the means to (a) predict how conflicts may evolve, and (b) how many people will flee a particular settlement once one or more violent event occurs.

In this paper we first analyse the simulation input data that is derived from conflict events. We identify several drivers of violent events, and provide a preliminary investigation of the relation of materials stocks with violent event occurrence, and characterize the geographical spread of these conflicts in an example situation (the Mali conflict). These findings will be used to develop a conflict evolution network model, and to couple it to our existing agent-based forced migration simulation, forming a hybrid simulation approach. We highlight the main challenges in developing a conflict evolution model, and propose a modified simulation development approach to study how variations in the conflict progressions could affect the destination choice of forcibly displaced people.

2   RELATED WORK

Whilst hybrid simulation has been around for some time, a more recent renewed interest in hybridization is motivating a revisiting of issues around simulation integration and coupling. Hybridization requires an additional focus on conceptual modelling (Zulkepli & Eldabi 2015) and this includes the points at which simulations should be integrated and connected. The combination of modelling paradigms can also symbiotically enhance their respective capabilities (Chahal and Eldabi 2008). Complex simulations of this type are able to help in understanding a phenomena and unintentional consequences of a complex adaptive nature (AlSubhi & Bell 2018). Hybrid techniques are further able to "generate a phenomenon without pre-supposing it is the best way to understand it, and this is best done in a simulation" (Bishai et al. 2014).

Integration and coupling of simulation models is particularly complex and covers a varied number of patterns. Four popular approaches include: 1) Multi-scale integration, 2) multi-paradigm integration, 3) Multi-platform/multi-architecture integration and 4) Multi-processing integration (including parallelization). Alowayyed et al. (2017) highlighted computational challenges when couple across a range of scales that include near-realistic representations. Multi-paradigm is more typical in the hybrid community, where a number of techniques (e.g. System Dynamics and Discrete Event Simulations) are applied in model implementation phase of a simulation study in order to gain further insights (Mustafee & Powell 2018). Multi-platform/multi-architecture is coupling across software platforms (typically between commercial software packages) or hardware architectures. Finally, multi-processing coupling covers both parametric studies with looser coupling and MPI style tight coupling.

Forecasting where conflicts occur help us to ensure new agents in our forced migration model are correctly created in (violent) places of origin. This is essential for making our migration models accurate. A number of external factors affect the occurrence of conflict events, and we are still in the process of exploring these. However, for example mining activity increases the likelihood of conflict incidence

(Berman et al. 2015), especially near active mines, and a spike of mineral prices increases conflict risk in areas that produce these commodities. Goldsmith (2015) found that Africa is not homogeneous when considering election-related civil violence, as election times tend to be more disruptive and violent in some regions. For instance, Mali is shown to have a significantly greater risk of riots around election periods. Weather events and rain also have had an impact on conflict, unsurprisingly this is curvilinear with greater impact seen when little or too much rain falls (Hendrix and Salehyan 2012).

3   THE HYBRID MODELLING APPROACH

In this paper, we use an adapted version of the Simulation Development Approach (SDA) presented by Suleimenova at al, (2017). The purpose of an SDA is to provide a recipe that not only describes a model, but covers a full process of constructing, refining and validating a model, and its implementation. In the original approach, we developed an SDA to validate against data from three historical conflicts, focusing exclusively on the forced migration modelling. In this paper, we attempt to forecast forced migration movements for future conflicts. To do so, we can no longer rely on historical conflict data, and need to integrate a *conflict evolution model* which estimates how violence in a civil war setting evolves over time. To incorporate this, we modify our existing SDA, and specifically replace some of the activities that take place in the "data source selection" and "model construction" phases. We provide a summary of our modifications in Figure 1.
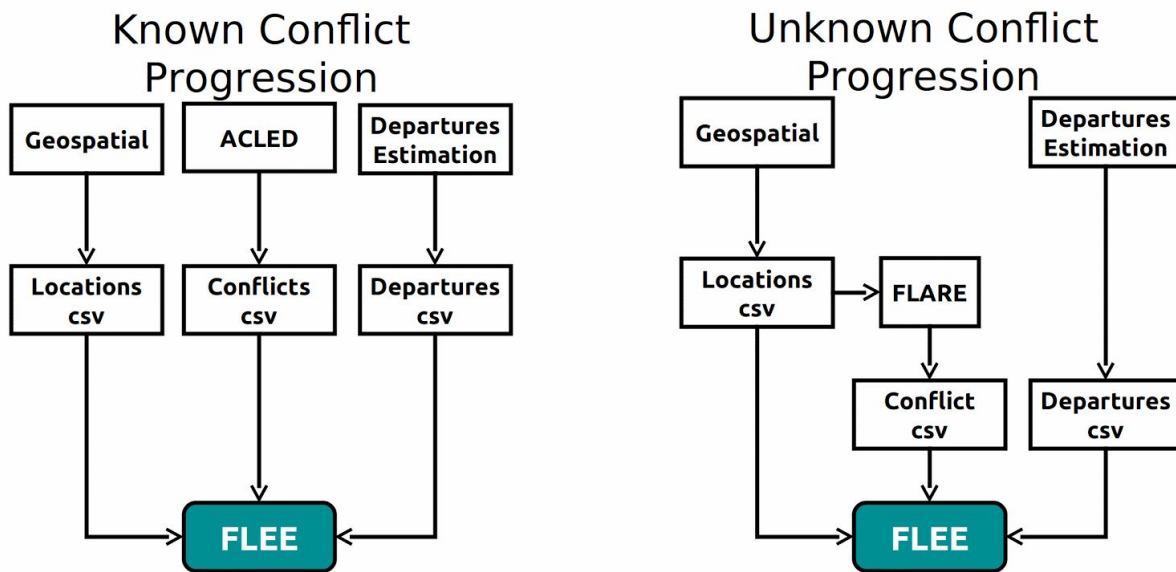

Figure 1: Overview of the envisioned SDA modifications required to use modelled conflict evolutions. The box "locations csv" also includes route information.

First and foremost, in our modified SDA we perform a forecast, and can therefore no longer rely on conflict information obtained from the Armed Conflict Locations and Events Database (ACLED, http://www.acleddata.com). Instead, we will rely on a *conflict evolution model*, named Flare which aims to forecast where violent events are likely to occur. The Flare code currently requires three inputs, a list of locations, routes interconnecting them (derived from geospatial data), and an initial location (or list of locations) where a conflict is known to have erupted. Using these inputs, Flare will then make a prediction as to which location will be affected by conflict during which period, and write this information to a Conflict.csv file, which contains a list of locations, along with days on which violence has occured in these places.

To enable the processing of conflict information by the Flee agent-based migration modelling code, we have adapted the Data Manager to support an additional data format (conflict CSV files). This conflict information is read in at the start of every Flee simulation, and subsequently stored in the simulation state. During each iteration, the Flee code checks whether a given location is a *conflict zone* using this state, and then modifies the appropriate waitings whenever the status of a given location has changed.

To support easy access by different stakeholders (policy makers, researchers, etc.) we also ported Flee on a Science Gateway. Science Gateways are used by numerous scientific communities worldwide aiming to support knowledge sharing by providing access to specialist resources through a user-friendly interface. We used the FutureGateway Framework to support the Flee Science Gateway application which in a very high level consists of three main components: i) the API Server front-end that accepts API calls, ii) the FutureGateway Database that maintains job information and iii) the API Server that coordinates the job submissions (Salomoni et al. 2018). The Flee Science Gateway application is currently hosted by the EGI Applications on Demand service (https://www.egi.eu/services/applications-on-demand/) site.

4   INPUT DATA ANALYSIS OF ARMED CONFLICT EVENTS

Agent behaviour is uncovered from a robust investigation of the domain of study. We undertake this behaviour mining first by attempting to contrast the ACLED datasets with a range of potential event drivers (e.g. political, environmental or financial volatility in the area).   Several possible drivers of conflict have been previously explored and discussed earlier, including: 1) Mineral mining and pricing, 2) elections, 3) weather events and 4) food security (Vanhille Campos et al. 2019). This section focuses on the mining price changes that have seen to have a particular impact on conflict occurrence. We compare ACLED data conflict occurrence (Figure 2) with the MSCI World Metals and Mining Index (USD) (see https://www.msci.com/documents/10199/fb6ddbdf-8df3-4fdf-96b4-e578e2ac776c, Figure 3).
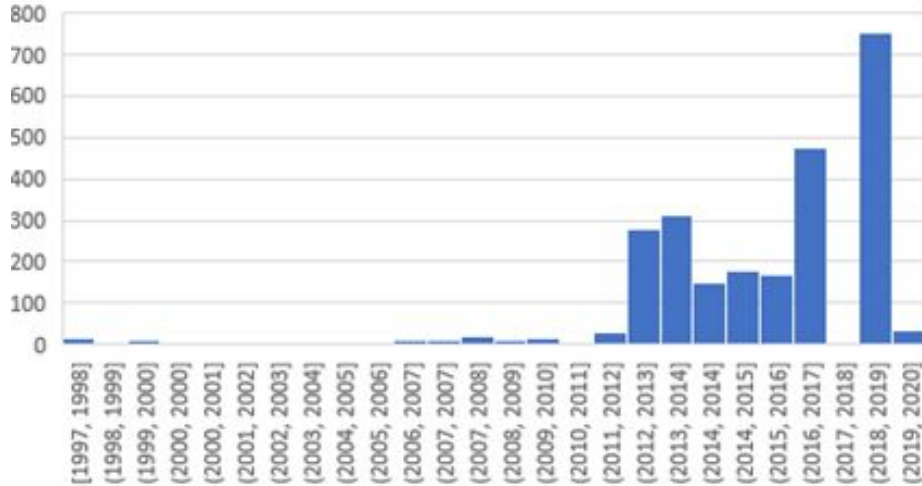
Figure 2: Yearly Conflict Summary.

Figure 3: Metal and Mining Index (2004-2019). Source: msci.com

The World Metals and Mining index is clearly falling during the year 2012-16 and rising from 2016-19 – with spikes around 09/10 and 18/19. The number of conflict occurrences reduce and increase respectively during these periods and give an indication of potential relationship. Earlier conflict reporting appears to be sparse in these data sets. Consequently, we have included rules that adjust the conflict probability during these periods when baselining the model. Future price change indication can be found on international trading exchanges that trade futures (e.g. The London Metal Exchange – http://www.lme.com). When considering the price of Gold, it can be seen (in Figure 4) that over the next two years (Gold future closing price for different contract lengths) the price is estimated to increase.
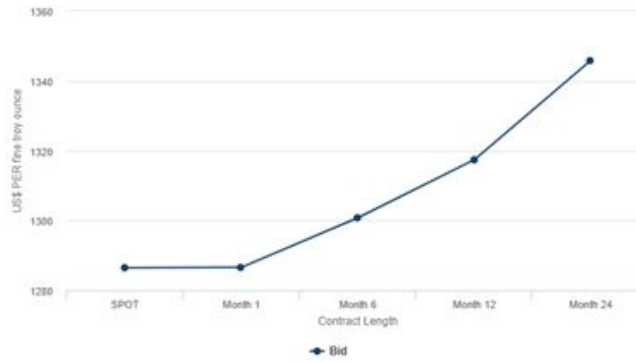
Figure 4: Gold Futures Contract prices (May 2019 – +24 months).

Conflict more generally is likely to happen in locations where conflict has occurred previously or where nearby conflict has spread. Harari and La Ferrara (2018) carried out a cellular analysis of ACLED data that supports this geospatial conclusion and further explores linked relationships such as mineral deposits and ethnic fragmentation. When analysing the ACLED conflict data (www.acleddata.com) for Mali it is clear that conflict occurs in the same locations (darker patterns) – see Figure 5. Circle size depicts a larger number of fatalities. Geospatial analysis encompasses many of the driver of conflict as they also have geospatial relationships, e.g. mine locations, weather effects or ethnic fragmentation.

Figure 5: Mali conflicts.

5   sDESIGNING FLARE: A CONFLICT EVOLUTION MODEL

In the context of this paper, we require a conflict evolution model to predict where violent events are likely to occur next, given a historical occurrence of violence. At time of writing, we are still working on

an implementation of such a model, which we name FLARE. The aim of Flare is to use a range of data sources and algorithms to construct an accurate conflict evolution forecast. Thanks to the existence of ACLED, validating such a model is relatively straightforward, as we will be able to directly validate algorithms by attempting "forecasts" in historical periods, and then validating them against the conflict events as recorded by ACLED in Africa and Asia.

Developing the conflict evolution algorithm itself is challenging, as many factors could be accounted for (see e.g., Section 4), and these factors may have different levels of impact depending on the conflict setting. As an example, the effect of mineral mining and pricing could have a much stronger effect on conflict evolution in mineral-rich countries such as Congo or Mali than on countries that rely less on mining income. To capture these challenges step-by-step, we have chosen to perform our algorithm implementation in four stages of increasing complexity:

- Stage 1 - Fully artificial evolution: in this stage, we prototype algorithms which only rely on generic growth parameters driven by high-level heuristics. This will result in very simplistic algorithms, that are unlikely to be particularly accurate, but that are useful in prototyping coupled workflows, as we do in this paper.
- Stage 2 - Extrapolated evolution: in this stage, we reserve a section of the ACLED data to be used for training. We then extract correlations from the training part of the ACLED data set in regards to historical conflicts, and use these correlations to construct a rule set for forecasting conflict evolution. Importantly, our validation setting is constrained as using a larger sample size for training results in less data to validate against. We also still need to identify mechanisms to prevent overtraining.
- Stage 3 - Evolution driven by external factors: in this stage, we use data sources capturing external factors (e.g., as identified in Section 4) to construct an algorithm for conflict propagation. This may include for example price levels, GDP levels, weather events or food security indicators, but explicitly excludes historical conflict data.
- Stage 4 - Composite algorithms: here we will develop algorithms combining advances from the three stages presented above.

At time of writing, we have implemented a Stage 1 prototype, using a simple, highly stochastic algorithm. The purpose of this prototype is not to create an accurate forecast, but to learn how different conflict evolutions could affect camp arrivals in the forced migration simulations. We perform this analysis using a hybrid/coupled simulation case study, which we present in the next section.

6   HYBRID/COUPLED SIMULATION CASE STUDY: THE 2012 NORTH MALI CONFLICT

To exemplify our approach, we apply a first prototype of our modified SDA to a model of the 2012 North Mali Conflict, as previously applied by Groen (2016) en Suleimenova et al. (2017). The simulation is hybrid in the sense that Flare relies on a stochastic network-based algorithm, while Flee uses an agent-based modelling algorithm. As a base, the simulations use the refined location graph for Mali presented by Suleimenova et al. (2017), as well as the other default parameter choices in that paper (including the duration of 300 days). However, we do run the code with a newer version of Flee (one that can handle CSV inputs), and for reasons of simplicity we chose to represent Mbera and Fassala in Mauretania as a single (camp) location, and not perform rescaling on our results.

The workflow in this case study follows the SDA for unknown conflict progressions, and has been automated using a specifically developed plugin for FabSim3 (Groen et al., 2016). It consists of the following steps:

- Twenty executions of the Flare code, each generating a different type of conflict progression. Each run takes around 10 seconds, requires one core, and executions can be done concurrently..
- Conversion of Flare output to Flee input.
- 20 executions of the Flee code, one for each Flare output. Currently, these Flee runs take around 20 minutes using one core, and can be done in parallel using up to 24 cores (with an efficiency of around 70%).
- Collation of the data into a single representation.
- Graph visualization of arrivals, including the uncertainty introduced by the Flare conflict code.

A simplified version of this workflow can be installed and run as part of the VECMA tutorial on migration, which is available at: https://www.vecma-toolkit.eu/tutorials/.

Within Flare we adopted a very simple Stage 1 algorithm (See Section 5). Here the initial conflict locations are Kidal, Timbuktu and Menaka, as used in previous works. During each day of the simulation, any adjacent location has a y% probability to become a conflict location, where y is equal to the number of adjacent conflict locations, divided by two. There is a 0% chance of conflict locations appearing if conflict does not take place in an adjacent location, and any conflict location will retain its conflict status for the remainder of the simulation.

We perform our Flee simulations in two locations: on a local desktop, which we use for code development and testing, but also on the 1.3 PFlop/s Eagle supercomputer in Poznan (https://www.top500.org/site/50183). We use the latter resource to show that our FabSim3-based workflow implementation can seamlessly switch between local and remote execution, and that we can maintain this workflow in the case that our simulations become more detailed (and expensive).

We pres   ent example forecasts of camp population growth for the two largest camps in the conflicts, Fassala-Mbera and Abala, respectively in Figures 6 and 7. In these forecasts, we account for the uncertainty introduced by the stochasticity of our simple conflict evolution model, which are represented using 95% confidence intervals. In our results, we find that the introduction of different conflict evolutions results in a very large uncertainty in the growth of camp populations, particularly in the middle segment of our simulations. This behaviour occurs even though our conflict evolution model has quite restrictive assumptions (e.g., conflicts can only spread to neighbouring locations). In the startup phase of our simulations, the uncertainty is very small, as new agents require some time to arrive in camps.
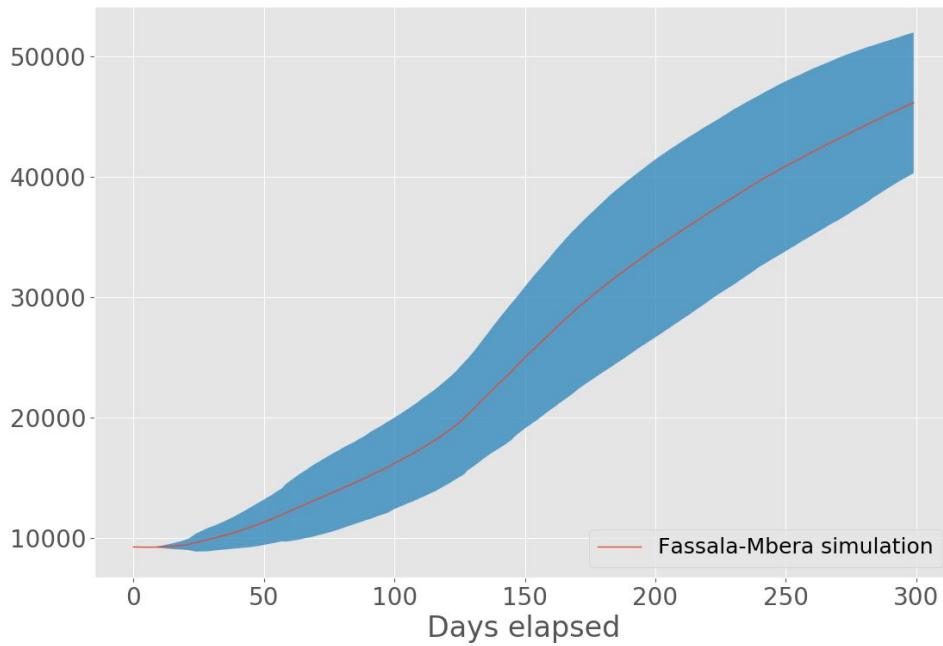
Figure 6: Camp population (value on y-axis) in Fassala and Mbera, the largest camp, as a function of the number of days elapsed. The mean forecast is indicated with a red line, and a 95% confidence interval (two standard deviations) accounting for different conflict progressions is indicated with the blue region.
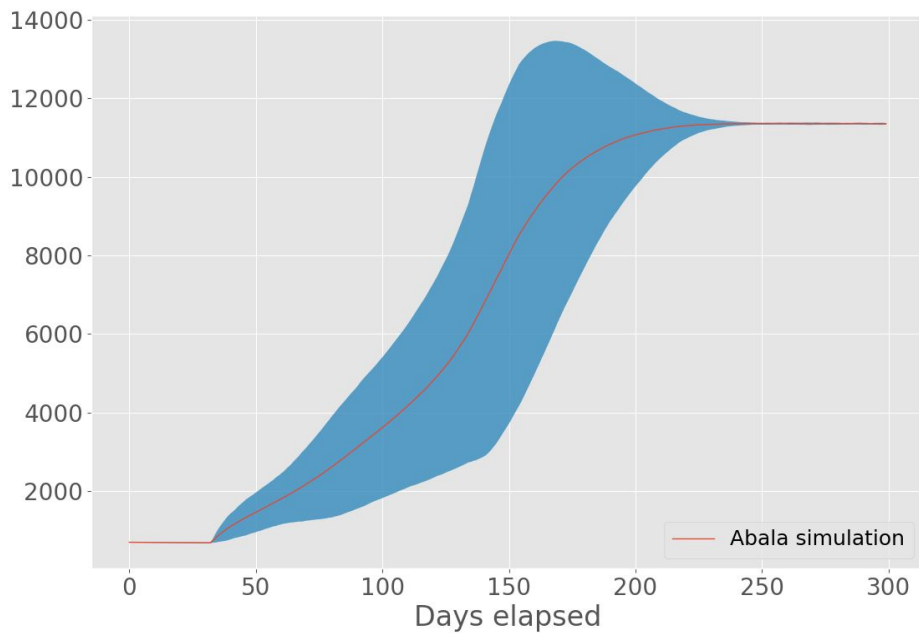

Figure 7: Camp population (value on y-axis) in Abala, the second largest camp, as a function of the number of days elapsed. See Figure 6 for an explanation of the lines and symbols.

In addition, in the case of Abala, the uncertainty gradually decreases after the camp has reached full capacity, as new agents are deterred from entering the camp. When we run simulations repeatedly

with the same conflict progressions (not shown here), camp populations may also vary due to the stochastic nature of the Flee code. However, this 95% confidence interval of these variations is substantially smaller, and always within a few hundred agents of the mean.

7   DISCUSSION

In this paper we presented a new, adapted simulation development approach to incorporate the evolution of conflict events into agent-based models of people escaping a civil war. We analyze the existing dataset from ACLED, and note that several external factors potentially have a relationship to the occurrence and spread of violent events. In addition, we demonstrated how a conflict evolution model could be integrated with the Flee agent-based model for forced population displacement, and showed that we can measure the uncertainty caused by a stochastic model for evolving violent events on the output of the agent-based model (populations in camps). We find that a random variation in conflict evolutions has a severe effect on the agent-based model output. In our opinion, this proves that any errors or uncertainty in the forecasting of conflict evolutions will have to be accurately propagated to any forced migration model that is derived from it. It also underlines that we need accurate and robust conflict evolution algorithms, if we want to make actionable migration forecasts in such a setting. The automation and HPC access tools we applied (FabSim3 and the FabFlee plugin we developed for it) are part of the VECMA toolkit, which is available at http://www.vecma-toolkit.eu (Groen 2019).

In terms of future directions, our main priority will be to establish a much better algorithm to predict the evolution of violent events (e.g., a Stage 2 or Stage 3 one), and to revisit some of the basic assumptions we have made in that regard in our migration simulations. This activity involves exploring ACLED and other datasets in search for key correlations, and to better characterize how the evolution of conflicts is affected for example by different time periods, or different geographical constraints. The Flare code is highly modular in its structure, and we are therefore easily able to apply and test conflict evolution algorithms proposed by external parties. Another thing we would like to do is to establish an environment where we can automatically validate the accuracy of such algorithms. Lastly, there is huge work to be done around the main migration agent-based model. Within VECMA this entails the coupling to other relevant models and performing a high quality, global, sensitivity analysis, while within HiDALGO we will seek to improve the scalability of the code, and establish integrations with a range of external models and data sources.

REFERENCES

Al Subhi, N.N. and D. Bell, 2018, " Hybrid user experience (UX) modelling in a heritage setting." In 2018 Winter Simulation Conference (WSC) (pp. 1298-1309). IEEE.

Berman, N., M. Couttenier, D. Rohner and M. Thoenig,, 2017. "This mine is mine! How minerals fuel conflicts in Africa. American Economic Review", 107(6), pp.1564-1610.

Bishai, D., Paina, L., Li, Q., Peters, D.H. and Hyder, A.A., 2014. "Advancing the application of systems thinking in health: why cure crowds out prevention". Health research policy and systems, 12(1), p.28.

Edwards, S., 2008. Computational tools in predicting and assessing forced migration. Journal of Refugee Studies, 21(3), pp.347-359.

Frydenlund, E., Foytik, P., Padilla, J.J. and Ouattara, A., 2018, December. Where are they headed next?: modeling emergent displaced camps in the DRC using agent-based models. In Proceedings of the 2018 Winter Simulation Conference (pp. 22-32). IEEE Press.

Goldsmith, A.A., 2015. "Elections and civil violence in new multiparty regimes: Evidence from Africa". Journal of Peace Research, 52(5), pp.607-621.

LiLitt

Groen, D., Bhati, A.P., Suter, J., Hetherington, J., Zasada, S.J. and Coveney, P.V., 2016. FabSim: facilitating computational research through automation on large-scale and distributed e-infrastructures. Computer Physics Communications, 207, pp.375-385.

Groen, D., Richardson, R., Wright, D., Jancauskas, V., Sinclair, R., Karlshoefer, P., Vassaux, M., Arabnejad, H., Piontek, T., Kopta, P., Bosak, B., Lakhlili, J., Hoenen, O., Suleimenova, D., Edeling, W., Crommelin, D., Nikishova, A., Coveney, P., 2019, Introducing VECMAtk - verification, validation and uncertainty quantification for multiscale and HPC simulations. In International Conference on Computational Science (pp. 479-492). Springer, Cham.

Hébert, G.A., Perez, L. and Harati, S., 2018. An agent-based model to identify migration pathways of refugees: The case of syria. In Agent-Based Models and Complexity Science in the Age of Geospatial Big Data (pp. 45-58). Springer, Cham.

Hendrix, C.S. and I. Salehyan 2012. "Climate change, rainfall, and social conflict in Africa". Journal of peace research, 49(1), pp.35-50.

Nguyen, H.K., Chiong, R., Chica, M. and Middleton, R.H., 2018, November. Agent-based Modeling of Inter-provincial Migration in the Mekong Delta, Vietnam: A Data Analytics Approach. In 2018 IEEE Conference on Big Data and Analytics (ICBDA) (pp. 27-32). IEEE.

Salomoni et al. 2018 "INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures" Journal of Grid Computing 16(3):381-408 https://link.springer.com/article/10.1007/s10723-018-9453-3

Suleimenova, D., Bell, D., and Groen, D. 2017. "A generalized simulation development approach for predicting refugee destinations". Scientific Reports, Vol. 7, Article number: 13377

Suleimenova, D., Bell, D., and Groen, D. 2017b. Towards an automated framework for agent-based simulation of refugee movements. In 2017 Winter Simulation Conference (WSC) (pp. 1240-1251). IEEE.

Vanhille Campos, C., Suleimenova, D., Groen, D., 2019, A coupled food security and refugee movement model for the South Sudan conflict. Proceedings of the International Conference on Computational Science (in press).

Willekens, F., 2018. Towards causal forecasting of international migration. Vienna Yearbook of Population Research, 16, pp.1-20.

AUTHOR BIOGRAPHIES

DEREK GROEN is is a Lecturer in Simulation and Modelling at Brunel University London, and a visiting Lecturer at the Centre for Computational Science, University College London. He is principal investigator for Brunel in the EU-funded HiDALGO and VECMA HPC projects, and has published over 50 peer-reviewed papers. He is Chair of the Multiscale Modelling and Simulation Workshop at the International Conference on Computational Science, and specializes in multiscale computing, agent-based simulation, and high-performance computing. His email address is derek.groen@brunel.ac.uk.

DAVID BELL is a Reader in the Department of Computer Science at Brunel University London where he carries out research into digital service design – modeling and simulating user interactions, decision making and behavioral change. Research domains have included healthcare, heritage and cyber security. Dr. Bell also spent 15 years working in the IT industry (primarily within Investment Banking) in roles from developer to technology director. He is the Education Chair for the ACM SIGSIM where he is involved in the promotion and support of doctoral simulation researchers. His email address is david.bell@brunel.ac.uk.

HAMID ARABNEJAD is a Post-Doctoral Research Fellow in the Department of Computer Science at Brunel University London. His research interests are related to scheduling and data management in grid and cloud computing, workflow and performance optimization, multiscale computing, and agent-based simulation.His e-mail address is hamid.arabnejad@brunel.ac.uk.

DIANA SULEIMENOVA is a Ph.D. researcher in the Department of Computer Science at Brunel University London, and the founding developer of the FabFlee automation tool. Her current research focuses on forced migration and refugee movements models. Her email address is diana.suleimenova@brunel.ac.uk.

SIMON J. E. TAYLOR is a Professor of Computing and the Director of the Modelling and Simulation Research Group in the Department of Computer Science, Brunel University London. He leads major projects in industry and Africa. His email address is simon.taylor@brunel.ac.uk.

ANASTASIA ANAGNOSTOU is a Lecturer and the Deputy Director of the Modelling & Simulation group at the Department of Computer Science, Brunel University London. Her email address is anastasia.anagnostou@brunel.ac.uk.