How Policy Decisions Affect Refugee Journeys in South Sudan: A Study using Automated Ensemble Simulations

Diana Suleimenova[1] and Derek Groen[1,2]

[1]Department of Computer Science, Brunel University London, 313 Kingston Lane, Uxbridge, Middlesex, UB8 3PH, United Kingdom
[2]Centre for Computational Science, University College London, London, WC1H 0AJ, United Kingdom
Correspondence should be addressed to *diana.suleimenova@brunel.ac.uk*

Abstract: Forced displacement has a huge impact on society today, as more than 68 million people are forcibly displaced worldwide. Existing methods for forecasting the arrival of migrants, especially refugees, may help us to better allocate humanitarian support and protection. However, few researchers have investigated the effects of policy decisions, such as border closures, on the movement of these refugees. Recently established simulation development approaches have made it possible to conduct such a study. In this paper, we use such an approach to investigate the effect of policy decisions on refugee arrivals for the South Sudan refugee crisis. To make such a study feasible in terms of human effort, we rely on agent-based modelling, and have automated several phases of simulation development using the FabFlee automation toolkit. We observe a decrease in the average relative difference from 0.615 to 0.499 as we improved the simulation model with additional information. Moreover, we conclude that the border closure and a reduction in camp capacity induce fewer refugee arrivals and more time spend travelling to other camps. While a border opening and an increase in camp capacity result in a limited increase in refugee arrivals at the destination camps. To the best of our knowledge, we are the first to conduct such an investigation for this conflict.

Keywords: Refugee Modelling, Agent-Based Modelling, Automation Toolkit, Policy Decisions, Validation, Sensitivity Analysis

Introduction

1.1 A civil war makes people vulnerable and leads them to migrate, in search for a secure and stable location. The choice of destination determines whether fleeing individuals are internally displaced people (IDPs) or refugees. IDPs seek safety within their own country and do not cross borders to neighbouring nations while refugees have been forced to flee their home countries due to war or violence (UNHCR 2010). There are more than 68 million people forcibly displaced worldwide, of which 24 million are refugees (UNHCR 2018a). These fleeing individuals are the unfortunate victims of civil wars and internal conflicts, who make decisions to migrate at the times of distress. To understand the causes of forced displacement, researchers establish three concerns faced by migrants, namely, the choice to stay or flee, the choice to flee internally or across borders, and the choice of destination (Salehyan 2014). Their decisions are often based on economic and political push-pull factors in sending and receiving countries. Especially, Schmeidl (1997) states that economic and political instabilities, poverty, violence and insecurity in the origin countries push people to flee. In contrary, economically and politically stable and safe countries pull forcibly displaced people to receiving areas. Thus, we can consider the economic and political conditions, security, the challenges and expenses of moving internally or across borders as causes of forced displacement.

1.2 Unfortunately, forced displacement has enduring consequences on the refugees, as well as on both sending and receiving countries. For instance, civil war and violence within the origin countries may spread across borders. Similarly, receiving nations may interfere in internal conflicts and wars occurring in sending countries to prevent

further increasing migrant arrivals (Gleditsch et al. 2008). In addition, receiving countries may face external costs by hosting refugees as residents and have to share available resources. Some countries may not have enough support and required refugee aid in terms of shelter, food and safety. Hence, hosting refugees can have positive or negative consequences (Martin 2005).

1.3 According to Jacobsen (1996, p. 674), refugee-receiving governments base their decision whether to host or refuse refugee arrivals on "the costs and benefits of accepting international assistance, relations with the sending country, political calculations about the local community's absorption capacity, and national security considerations". Thus, both sending and receiving governments make decisions based on human rights, economic, political and humanitarian factors. Their institutional decisions turn into policies that can manage, resolve and overcome the consequences of forced displacement. Similarly, aid organisations and non-governmental organisations (NGOs) take part in the decision-making to achieve their objectives, such as facilitating efficient allocation of human resources for refugees in camps. However, the literature lacks in identifying effective policies to overcome and assist in forced migration. It is also seldom clear how policy decisions affect refugee journeys and camp arrival rates, particularly those in other countries.

1.4 For instance, South Sudan has more than 4 million forcibly displaced people, of which 2.4 million are refugees hosted by neighbouring Sudan, Uganda, Kenya and Ethiopia (UNHCR 2018b). These people escape a civil war that lasted for decades prior to its independence from Sudan and still continues to this date as local authorities fail to provide basic needs (Reid 2018). Moreover, the Sudanese government closed their border with South Sudan and worsened the situation as it disabled free movement of residents and obstructed the exchange of goods between both countries. To understand the South Sudan conflict extensively, we focus on how policy decisions, such as camp or border closure, camp capacity change and forced redirection, can have an effect on the distribution of refugee arrivals across camps in neighbouring countries.

1.5 Forecasting forced displacement is challenging as many forced population data sets are small and incomplete, data sources have too little information; statistical methods are outdated and do not consider refugee arrival estimations (Edwards 2008; Disney et al. 2015). Yet, forced population predictions are essential to save refugee lives, to investigate the consequences of a nation closing its border for forced population, and to help complete incomplete data collections on refugee movements (Groen 2016). Improvements in data collection may be a possible solution to overcome data issues, but we require an enhanced logical framework to capture forced displacement thoroughly.

1.6 In this paper, we investigate the effect of policy decisions in the South Sudan conflict on the camp arrival rates of displaced persons. Policy changes may occur at short notice in crisis situations, and correctly predicting arrivals of displaced people in these cases is essential to prevent shortages in food, water and shelter in refugee camps.

1.7 The recent study by Gilbert et al. (2018) examine the role and applicability of agent-based modelling (ABM) when experimenting with policy decisions, where ABM provides an understanding and knowledge to governments, stakeholders and policymakers by modelling complex systems, such as human movement in this paper. Our ABM approach examines the effects of policy implications under various scenarios. It also provides a new perspective and helps researchers and other organisations in forecasting refugee movements, and inform policy decisions related to forced displacement. This was previously much more difficult and ineffective due to incomplete data and outdated statistical analysis.

1.8 To enable this investigation, we extend the simulation development approach (SDA) to support counterfactual scenarios. The original SDA, initially proposed by Suleimenova et al. (2017a), adopts an ABM approach to estimate how refugees reach destination camps. As part of the Verified Exascale Computing for Multiscale Applications (VECMA) project, we seek systematic approaches to validate and analyse the sensitivity of our simulations, to investigate output variability and to generate more actionable results (Groen et al. 2019b). To do this, we incorporate integrated sensitivity analysis and exploration of policy decisions in our SDA. This helps us to better understand refugee behaviour and better assist policymakers with their decision-making process.

1.9 Moreover, we add value by automating parts of the simulation development process, from construction to execution. Here, we propose an automated policy exploration toolkit, together with the sensitivity analysis, which is an essential step towards enabling users to create refugee arrival forecasts within days of a new conflict erupting. To showcase our approach, we present simulation results for seven runs for South Sudan and discuss how changes in the policy of forced displacement affect the refugee arrival rates in camps. Forecasting refugee arrival rates in camps is crucial since governments and NGOs can use this information to better allocate humanitarian resources and provide humanitarian protection to forcibly displaced people (Groen 2016).

1.10 In the remainder of this paper, Section 2 discusses computational modelling techniques of forced displacement, describes a generalized simulation development approach for refugee modelling and presents an automated toolkit for policy explorations. We apply our proposed approach to the South Sudan conflict. We then discuss

simulation results in application to policy decisions of camp or border closure, camp capacity changes and forced redirection (Section 3). Finally, Section 4 concludes and offers final remarks on refugee modelling and policy decisions.

Computational Modelling of Refugee Movements

2.1 In recent years, there has been a gradual increase in the use of computational techniques, both machine-learning based (Sfyridis et al. 2017; Quinn et al. 2018) and simulation-based (Hassani-Mahmooei & Parris 2012; Sokolowski et al. 2014; Hebert et al. 2018), to provide migration forecasts. In the case of simulations, one of the more widely adopted approaches is ABM, a computational approach that provides an opportunity to model complex systems with individual heterogeneity. It consists of agents that represent animals, humans, organisations or any other types of entities interacting with each other and within their environment. Particularly, the use of ABM allows to model how agents and their environment vary across time and space. Agents are autonomous and often unique, meaning that each agent is distinct in terms of size, location and other attributes (Macal & North 2014).

2.2 Recent forced displacement studies favour the use of ABM to study the influence of natural disasters and climate change on displacements (Kniveton et al. 2011; Hassani-Mahmooei & Parris 2012; Entwisle et al. 2016). Johnson et al. (2009) determined how refugees behave and interact with military groups in neighbouring camps, while Collins & Frydenlund (2016) investigated how refugees form groups depending on their travelling speed towards safer environments using an ABM. Moreover, Lin et al. (2016) analysed economic and social factors influencing refugee decisions to migrate by adopting an ABM. Anderson et al. (2006, 2007) proposed an ABM approach exploring policy decisions for sending and hosting governments and organisations interacting with refugees. Comparatively, Sokolowski & Banks (2014) presented an ABM Environment Matrix based on an early warning model of forced displacement (Schmeidl 1997; Schmeidl & Jenkins 1998) and applied to the Syrian conflict.

2.3 In the context of refugee arrival predictions, Suleimenova et al. (2017a) propose a generalized simulation development approach forecasting the distribution of refugee arrivals across destination camps. To understand the significance and generalisation of the proposed approach, the authors for the first time successfully modelled three African countries experiencing refugee emergencies, namely Burundi, Central African Republic and Mali, using a single approach. Their generalized approach relies on an ABM, where refugees are agents, and each time step represents one day since the validation data has a granularity of a single day and cannot be used to validate patterns on an intra-day timescale. The simulation starts by inserting a number of refugees (obtained from the United Nations High Commissioner for Refugees (UNHCR) database) in their conflict locations (extracted from the Armed Conflict Location and Event Data Project (ACLED)) that can be presented using a network-based ABM model. Each refugee can traverse from zero to more links during each simulation step. The probability of an agent's movement depends on the move chance, where the move chance of 1.0 represents agents in conflict and between locations, 0.001 for refugee camps and 0.3 for all other locations. Suleimenova et al. (2017a) provide a detailed flowchart of algorithm assumptions and agent parameters used by a simulation code - Flee, which can be found at `https://github.com/djgroen/flee-release`. It is optimised for its simplicity and flexibility, and it can be adapted to most scenarios involving escaping refugees.

Description of a generalized simulation development approach with the FLEE code

2.4 To facilitate rapid and consistent simulation development, Suleimenova et al. (2017a) suggest a generalized SDA, which enables rapid construction, execution and validation of refugee counts in conflict scenarios. We present a revised generalized SDA in Figure 1, which contains the same six phases of the original SDA (refer to Suleimenova et al. (2017a)), including situation selection, data extraction, model construction, model refinement, simulation execution and analysis, but also enhanced to fit the focus of this paper. Specifically, this revised version incorporates changes in policy decisions (e.g. camp and border closures, camp capacity changes and forced redirection) in the refinement phase and introduces both an ensemble of simulation executions and sensitivity analysis of simulation runs in the simulation execution phase.

2.5 Currently, the construction and execution of simulations are mostly done manually, which is both inefficient and time-consuming. For instance, an extraction of input data, construction of network maps and initial models for Burundi and CAR required 2-3 weeks of manual work. While refugee predictions need quick construction and execution as there is a prediction urgency of refugee crises or multiple conflict scenarios to simulate on

short time period (Suleimenova et al. 2017b). Hence, we automate several phases of the SDA, namely, the construction, an instantiation and execution of ensemble runs using a unified approach. In the next section, we describe the automation of each phase of a generalized SDA.
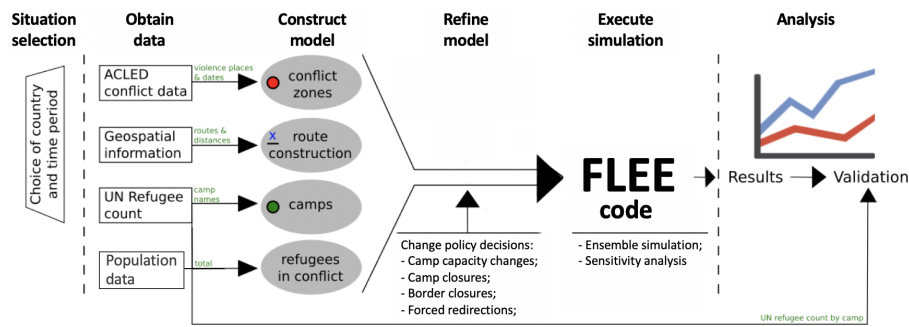

Figure 1: A generalized simulation development approach forecasting the distribution of refugee arrivals across destination camps. We use the same assumptions as given in Suleimenova et al. (2017a) for our simulations (except where this is mentioned otherwise for individual runs).

Automation of simulation development using FabFlee

2.6 Manual routine tasks in model construction and simulation execution can be simplified using automation tools. Automation is essential in simulation development since it provides time efficiency to modellers, reduces the probability of human error, simplifies and accelerates process activities and delivers a highly transparent and customised programme to users. Suleimenova et al. (2017b) comprehensively discuss existing automation tools, as there is an extensive number of languages, open-source software and automation tools that facilitate the development of computational research. Groen et al. (2019a) perform an analysis of added value for a range of coupling tools, including several automation tools. In both works, FabSim is recognised as a toolkit that helps to curate and simplify simulation research at the simulation deployment, execution and optimisation stage. Based on these findings, we chose to use the FabSim3 toolkit, which is an improved version of FabSim. Among other things, FabSim3 simplifies organising input and output files, user and machine configurations, and application executions (Groen et al. 2016). Currently, FabSim3 contains an integrated test infrastructure, more flexible customisation options using a plug-in system, and in-code documentation and examples to improve usability. It is distributed under a BSD 3-clause license.

2.7 A FabSim3-based FabFlee toolkit is one of the plug-in applications, which predicts the distribution of incoming refugees across destination camps under a range of different policy situations (Groen et al. 2019b). FabFlee is a partially automated implementation of our SDA, and provides an environment for researchers and organisations to construct and modify refugee simulations, instantiate and execute multiple runs for different policy decisions, as well as to validate and visualise the obtained results against the existing data. In Figure 2, we present the SDA phases with automated functionalities from model construction to analysis. Specifically, we aim to construct the initial model using a comma-separated values (csv) formats, refine the model with a new set of parameters or policy range decisions, execute an automated ensemble of runs and analyse the obtained results with the use of automated plotting tools. In the next section, we provide a detailed description of automation applied to each phase of the SDA.
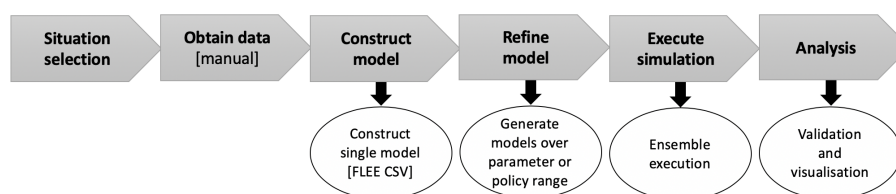

Figure 2: Phases of our simulation development approach, given in arrow boxes, and automation implemented in FabFlee for each phase, described in the ovals.

Model construction

2.8 To start with, we simplify the model construction phase by creating reader modules for csv formats for input data. Three formats of csv files, namely *locations.csv*, *routes.csv* and *closures.csv*, are integrated with FLEE's input interface. The initial idea of introducing these files is described in Suleimenova et al. (2017b). For this paper, we revised their outline to reduce data collection time and implemented these csv file formats in the model construction phase. We create these csv files manually according to the formats demonstrated in Tables 1, 2 and 3 for the conflict scenario and store them under the base conflict data in a specified conflict directory (e.g. `https://github.com/djgroen/FabFlee/tree/master/conflict_data/SSudan`).

name	county	country	latitude	longitude	location_type	conflict_date*	population/capacity
					conflict		population of location
					town	-	-
					camp	-	camp capacity
					forwarding_hub	-	-

Table 1: locations.csv contains all locations with the properties required for simulation construction, such as name and geographical information of locations, and populations (for non-camp locations) or capacities (for camp locations). *Note:* conflict_data is given as an integer, counting the number of days after the simulation start. The value of 0 indicates the start, while -1 indicates the end date of the simulation.

location1	location2	distance [km]	forced_redirection*
			0
			1
			2

Table 2: routes.csv specifies distances between two locations. *Note:* forced_redirection refers to redirection from source location (can be town, camp or forwarding_hub) to destination location (mainly camp). The value of 0 indicates no redirection, 1 indicates redirection from location2 to location1 and 2 corresponds to redirection from location1 to location2.

closure_type*	name1	name2	closure_start*	closure_end*
location				
country				

Table 3: closures.csv provides camp closure event specifying locations names or border closure event requiring country names to name1 and name2 respectively. *Note:* closure_type can be two types: location corresponding camp closure and country referring to border closure. closure_start and closure_end are given as integers, counting the number of days after the simulation start. The value of 0 indicates the start, while -1 indicates the end of the simulation.

2.9 After the generation of *locations.csv, routes.csv* and *closures.csv* files, we follow the FabFlee workflow diagram (see Figure 3). As a start, we load a base conflict data which includes csv files and the source data of a conflict scenario using *load_conflict* command. This, in turn, duplicates all existing files from a base conflict directory to a working directory, namely active conflict data. The load command also generates a text file (i.e. *commands.log.txt*) that records command logs of commencing activities. Moreover, to refine the model, we examine policy implications through parameter explorations for policy decisions related to a refugee emergency. We have developed several parameter exploration commands to modify a range of parameters illustrated in Table 4.
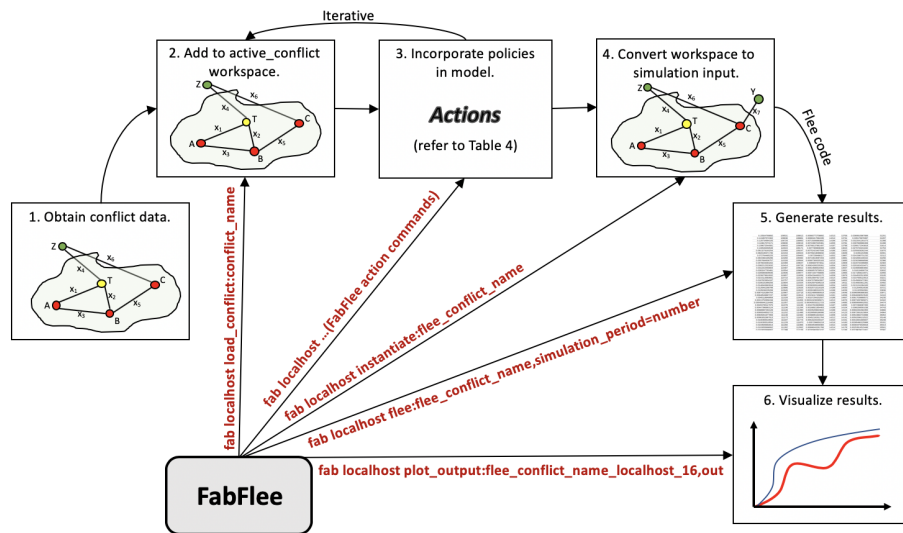
Figure 3: FabFlee workflow diagram demonstrating steps to explore policy decisions

Actions	FabFlee command
change camp capacity	change_capacities:camp_name=capacity
add a new location	add_camp:camp_name,region,country,lat,lon
delete an existing location	delete_location:location_name
camp closure	close_camp:camp_name,country,closure_start,closure_end
border closure	close_border:country1,country2,closure_start,closure_end
forced redirection	redirect:source,destination,redirect_start,redirect_end

Table 4: FabFlee functions for policy decision exploration.

2.10 Following the refinement phase, we duplicate parameter changes of the model by running the instantiate command. The instance is then saved in a new directory, which can include run name, version and date of instantiation on users insert choice. Now that we have our simulation input, we can proceed with the fifth phase of our SDA and run execution command triggering the FLEE code and producing results. Next, we visualise and validate the obtained results with graphs for each camp in a neighbouring country by running *plot_output* command.

2.11 To create a clean slate for future work, we can clear the active conflict directory using

fab localhost clear_active_conflict,

upon which we can reload the conflict and change other parameters (and instantiate and run a new simulation). Indeed, phases four to six in Figure 3 can be iterative and produce additional results as we extend our policy and parameter exploration. Similarly, we can conduct sensitivity analysis for each instantiated model by running *test_sensitivity* function (see Table 5 for more details).

Sensitivity test	FabFlee command
refugee move speed	test_sensitivity:flee_conflict_name,simulation_period=number, name=MaxMoveSpeed,values=50-100-200
refugee awareness level	test_sensitivity:flee_conflict_name,simulation_period=number, name=AwarenessLevel,values=0-1-2

Table 5: FabFlee functions for sensitivity test analysis

Forecasting the distribution of refugees in South Sudan

2.12 To understand the significance and practicality of a generalized and automated SDA, we construct a new model of the South Sudan conflict, which involves almost 2 million refugees fleeing to destination camps (UNHCR

2018a). For many years, Sudan experienced a civil war from which South Sudan declared independence on the 9th July 2011. However, the authorities of South Sudan failed to deliver the basic needs (Reid 2018), and in December 2013, a conflict between the government and rivals broke out.

2.13 Specifically, the civil war in South Sudan started on 15 December 2013, following fierce fighting between rival units of the Sudan Peoples' Liberation Movement (SPLM) and the Sudan People's Liberation Army (SPLA) in the capital, Juba (UNHCR 2015). Subsequently, South Sudan's president Salva Kiir announced that former vice president Riek Machar had attempted a coup. Machar escaped from Juba and became the leader of an armed opposition movement, namely the 'SPLM/A in Opposition'. Violence and fighting spread to other parts of the Jonglei, Upper Nile and Unity states, as well as other regions of South Sudan (ICG 2014). This forced people to flee internally and across neighbouring countries.

2.14 Our South Sudan model has a simulation period of 604 days starting from the 15th December 2013 to the 10th August 2015, during which 2.4 million refugees were known to escape the country. We run the simulation for 10 camps (listed in Table 6) in neighbouring countries, namely Sudan, Uganda and Ethiopia. Overview of the geographical network model for South Sudan demonstrated in Figure 4.

Countries	Camp names
Ethiopia	Tierkidi, Kule, Pugnido and Jewi
Kenya	Kakuma
Sudan	Khartoum and West Kordofan
Uganda	Adjumani, Rhino and Kiryandongo

Table 6: List of camps in neighbouring countries of South Sudan

Setup of simulation execution for South Sudan

2.15 After selecting our conflict country and the simulation period, we then extract data from the sources according to the SDA. Next, we construct our initial model for South Sudan with default settings using the discussed three csv file formats, namely *locations.csv*, *routes.csv* and *closures.csv*. The initial constructed model, which is the third phase of SDA, is then refined with additional information obtained from reports (fourth phase of SDA). In Figure 5, we demonstrate the layout of our simulation tests for the South Sudan conflict. This includes refinements to determine how policy decisions, such as camp and border closures, changes in camp capacities and redirection between camps, can affect the distribution of refugee counts and simulation results. Using our approach, we also automatically create and perform sensitivity analysis study for each of our scenarios. Bearing in mind, we set our default setting to the refugee move speed is equal to 200 km per day and the awareness of surrounding is 1 link.

Description of the base scenarios

2.16 After constructing the initial South Sudan model (*ssudan_default*), we executed and obtained the initial results. Next, we determined level 1 refugee registrations from the source data and included them to improve the initial model. We named the second model as *ssudan_reg* and executed to observe changes in the results. We further refined the *ssudan_reg* model using additional information obtained from publicly available online reports. Specifically, the UNHCR (2014) report declares that refugees arrived at Ethiopian camps on foot, due to the lack of roads. To accommodate this fact, we modified our simulation assumptions, and we incorporated specific "off-road links" from conflict zones to Ethiopian camps in a modified simulation setup named *ssudan_links*. To reflect the fact that off-road routes are likely to result in slower travel speeds, we multiplied the coordinate point-by-point distances by 2 for all walking routes. We also incorporated additional information in regards to later camp openings and closures, which was derived from the UNHCR reports (run *ssudan_ccamp*).
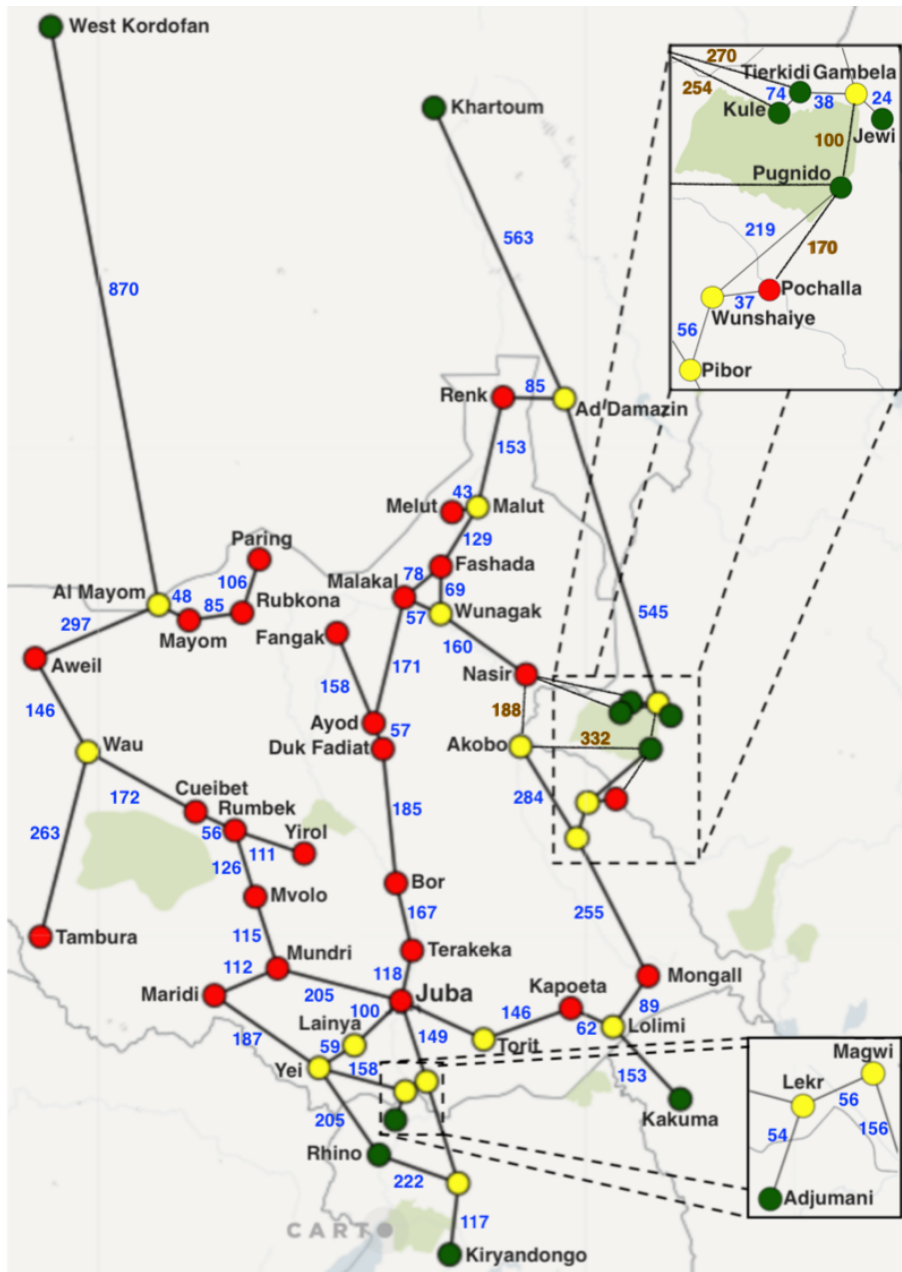
Figure 4: Overview of the geographic network model for South Sudan. This includes conflict zones (red circles), refugee camps (dark green circles) and other major settlements (yellow circles). Interconnecting roads walking routes are given with lines, with adjacent numbers used to indicate their length in kilometres (blue for roads and brown for walking routes). Background maps are courtesy of carto.com created using OpenStreetMap data.
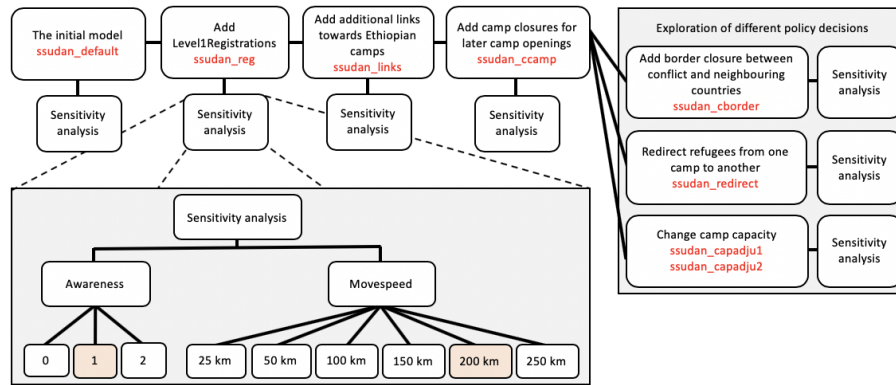
Figure 5: Setup of simulation execution for South Sudan. For each execution, we perform ensemble runs for sensitivity analysis. The structure of these ensembles is given in the bottom grey panel.

Results

3.1 In Figure 6, we demonstrate the averaged relative difference for four simulations (*ssudan_default, ssudan_reg, ssudan_links* and *ssudan_ccamp*). Despite the same levels before day 200, the average relative difference for these runs persistently lessens respectively from 0.615 to 0.499 over the simulation period and the refinement of the South Sudan model as we incorporated additional details. Overall, *ssudan_ccamp* is the most refined with the lowest average relative difference in the aggregate level. We calculate the average relative difference using the equation below:

$$E(t) = \frac{\sum_{x \in S}(|n_{sim,x,t} - n_{data,x,t}|)}{N_{data,all}} \tag{1}$$

where, the number of refugees found in each camp x of the set of all camps S at time t is given by $n_{sim,x,t}$ based on the simulation predictions, and by $n_{data,x,t}$ based on the UNHCR data. The total number of refugees reported in the UNHCR data is given by $N_{data,all}$ (Suleimenova et al. 2017a).

3.2 Moreover, we perform a range of sensitivity analysis tests to identify the important input variables in an awareness level and agents' movespeed of the simulation outputs. To begin with, we executed 10 replicas of *ssudan_ccamp* with default settings to determine the range of the output due to the probabilistic nature of the simulations. Over these 10 executions, the average relative difference ranged between 0.495 and 0.502. In addition, we perform a sensitivity analysis for each run by varying the level of agent awareness range and a speed limit of refugees. Here, the awareness range represents the level of knowledge of refugees about nearby locations. They may know only the distance to the adjacent locations in the graph (path distance only), or also the type of location for adjacent locations (1 link away), or also the location type of locations adjacent to those (2 links away). We present the results of this analysis in Table 7. For the most refined scenarios, the averaged relative difference is lowest when agents are aware of locations 1 link away, though the difference is marginal compared to simulations with an awareness range of 2 links away. Our simulations are clearly sensitive to the maximum refugee move speed parameter, and in particular move speeds below 100km/day result in significantly higher validation errors. This parameter sensitivity is in line with our simulations of previous conflicts (Suleimenova et al. 2017a).
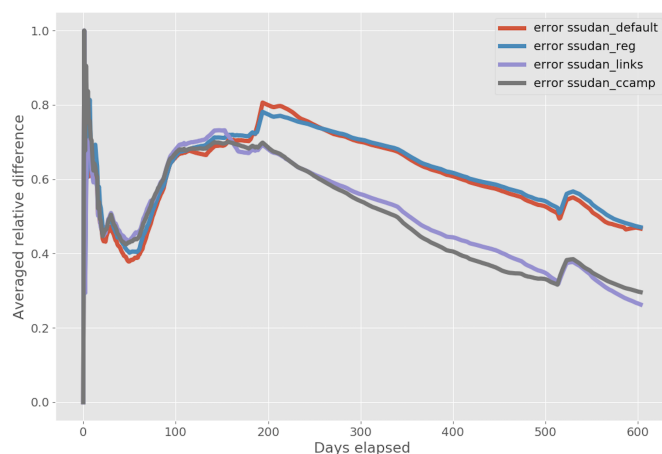
Figure 6: Overview of the averaged relative differences for ssudan_default (red line), ssudan_reg (blue line), ssudan_links (violet line) and ssudan_ccamp (grey line) simulations.

Run type	ssudan_default (least refined)	ssudan_reg ...	ssudan_links ...	ssudan_ccamp (most refined)
normal (default)				
1 link away, 200km/day	0.615	0.621	0.509	0.499
awareness range				
Path distance only	0.627	0.630	0.530	0.522
1 link away	0.613	0.621	0.510	0.500
2 link away	0.611	0.614	0.517	0.507
max. move speed (km/day)				
25	0.667	0.673	0.575	0.570
50	0.634	0.643	0.535	0.527
100	0.621	0.629	0.519	0.503
150	0.616	0.625	0.514	0.501
200	0.611	0.622	0.511	0.502
250	0.616	0.624	0.509	0.502

Table 7: Averaged relative difference values, averaged over time and all four base type of simulations using different agent awareness ranges, and different speed limits for agents. Note that we present results from 3 separate executions of the default type run: in the first data row, the third data row (labelled "1 link away") and the ninth data row (labelled "200").

3.3 We present *ssudan_ccamp* simulation results for all 10 camps validated against the UNHCR refugee registration data in Figure 7. The most populous camp in our simulation is Adjumani with more than 140,000 refugees over the simulation period and slightly overpredicted after 200 days into simulation compare to the data. The reason being that it is the closest camp for refugees fleeing from the South Sudan conflict. The forecast refugee counts in Kiryandongo and Kakume (at the start prior to 100 days) are in close agreement with the UNHCR data, while our simulations underpredict for Kule, Jewi and Khartoum camps. South Sudan has a record of being in conflict prior to our simulation start date. Kakuma (45239), Pugnido (42044), Rhino (5313) and Kiryanongo (15) camps had registered number of refugees fled prior to the simulation start; these, therefore, do not count towards the refugee arrival numbers.
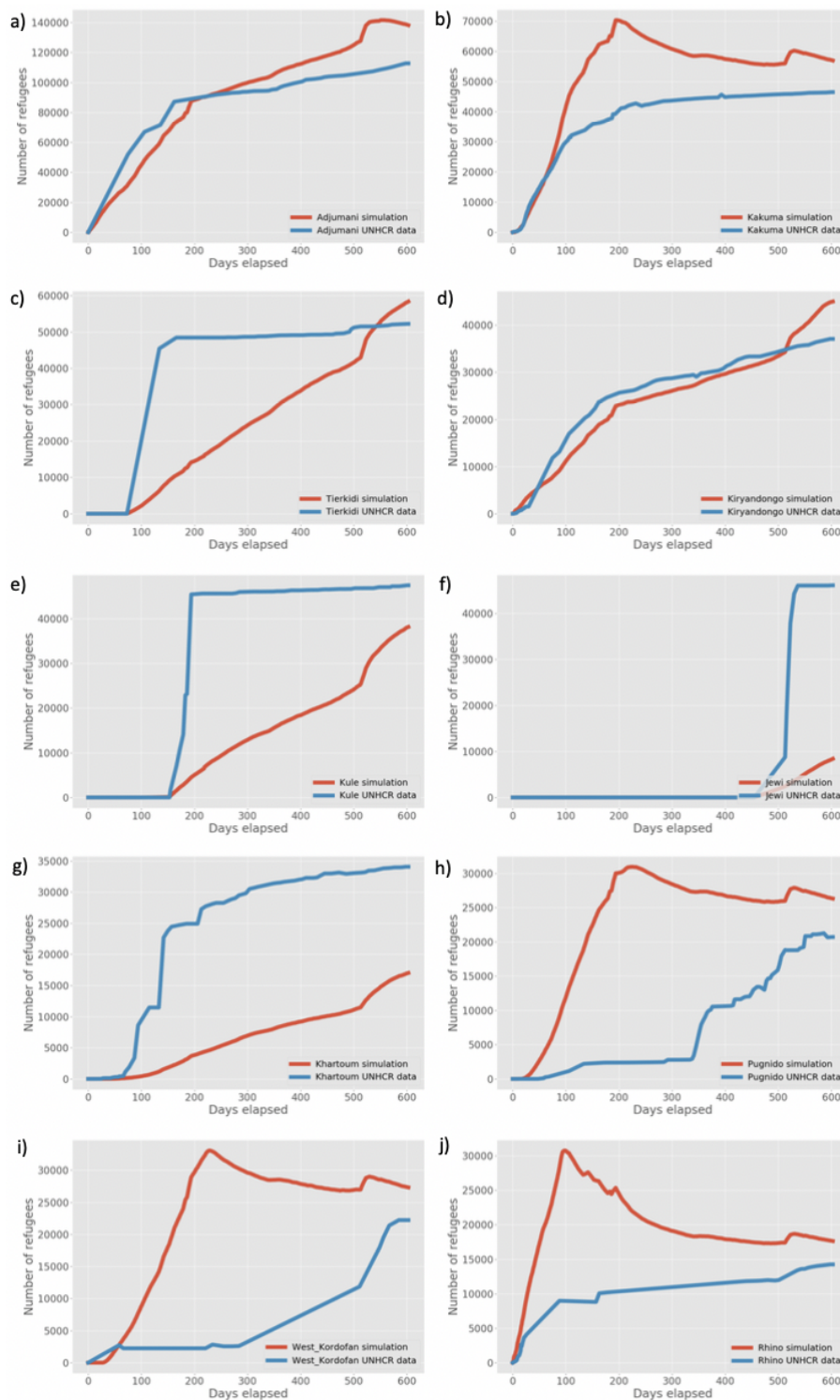
Figure 7: Number of refugees as forecast by our ssudan_ccamp simulation and validated against the UNHCR data for the South Sudan conflict. (**a-j**) Graphs are ordered by camp population size, with the most populous camp on the top to the smallest one on the bottom.

3.4 There are no arriving refugees at the start of simulation period for several camps, namely Tierkidi, Kule and Jewi, illustrated in Table 8, as they opened after the conflict has commenced according to the UNHCR data. For instance, the Tierkidi camp has no arrivals prior to 73 days of simulation, but refugee counts increase over the simulation period and overpredict UNHCR data by the end of simulation period. In addition, the Jewi, Kule and Khartoum camps show slowly increasing and underpredicted refugee counts. Whereas, the Pugnido, West

Kordofan and Rhino camps are considerably overpredicted according to simulation results by almost 25000 refugees decreasing to 5000 refugees for each camp by the end of simulation period.

Countries	Camp names	Camp opened on
Ethiopia	Tierkidi	26th of February 2014
	Kule	17th of May 2014
	Jewi	15th of March 2015

Table 8: A list of camps that opened after the South Sudan conflict has commenced in neighbouring countries.

Examining policy decisions

3.5 There are various real-world policy implication instances, which have changed the course of refugee movements. To demonstrate, the Dadaab camp in Kenya, which was opened in 1991, currently hosts more than 260000 Somali refugees (Cannon & Fujibayashi 2018). Despite its populated occupancy, in 2016, the Kenyan authorities attempt to shut this camp, but a high court judge ruled out the authority's decision and allowed refugees to remain in the Dadaab camp. However, if the authority closed the camp, it would have forced refugees to return to their violent home country, hide illegally in Kenya or flee to other neighbouring countries. Another instance is refugee camps in the Gambia, which were planned to reopen in 2006 since there was an increase in refugee numbers. Nevertheless, the authority was hesitant to site camps near borders due to armed opposition groups who opposed a danger on refugee lives. Hence, the authority has decided to place Casamance refugees in the old camp at Bambali in the central Gambia. Since this camp was farther from borders, refugees refused to travel and settled in nearby villages of Gambia (Grant 2016). In this case, it is uncertain how refugees have impacted the Gambian villages and their residents, as well as how refugees managed themselves in the neighbouring country. Moreover, in May 2013, Jordan closed its borders with Syria to stop the influx of refugees, which instead increased illegal crossings into the country (Hargrave et al. 2016). It is an instance of a single country, while there are many other countries that have closed their borders to refugees, such as European countries, and increased illegal crossings and human trafficking.

3.6 These instances illustrate that policy decisions have influenced refugee movements towards destination camps. However, we do not know how these decisions affect refugees, their movement and overall refugee counts. Hence, we aim to investigate and understand the implications of these policy decisions on refugee arrivals, as well as inform other authorities and policymakers on the consequences of their decisions. We model policy decisions using four different scenarios of the South Sudan conflict.

3.7 First, we compare the refugee arrivals in camps between three scenarios, which are a model without camp and border closures (*ssudan_links*), a model with camp closures (*ssudan_ccamp*), and a model containing an additional border closure between South Sudan and Uganda, enforced until day 302 that is halfway into the simulation (*ssudan_cborder*). We present our comparison results in Figure 8, and find no significant differences between *ssudan_links* and *ssudan_ccamp*. However, we do find differences between these two scenarios and *ssudan_cborder*, which results in 40% fewer refugee arrivals on day 302. This implies an increasingly long travel time for refugees up to day 302, the day that the border is again reopened. In addition, the delaying effect of border closures lingers in our simulation results after borders have been reopened, with approximately 15% fewer arrivals on day 400, for instance. This emergent behaviour can by definition not be validated against reality (we're examining a counterfactual). However, explanations for such delays are possible. For instance, refugees may fear that recently opened borders are more likely to be closed again, or may not be immediately aware that a previously closed border has again reopened.
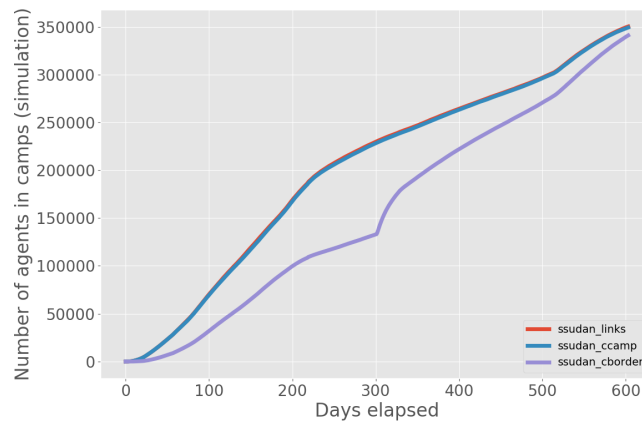
Figure 8: Comparison of the total number of refugees in camps between ssudan_links (red line), ssudan_ccamp (blue line) and ssudan_cborder (violet line) simulations.

3.8 Second, to explore how changes in camp capacities affect simulation results, we changed the capacity of the most populous camp, namely Adjumani. For the first instance, we decreased the original capacities of 112734 refugees by half. The second instance involved an increase in the original capacity by 50%. In Figure 9, we present the number of refugees for Adjumani camp *ssudan_adjumani1* (capacity: 56367 refugees) and *ssudan_adjumani2* (capacity: 169101 refugees). We find that a reduction of capacity in Adjumani results in up to 16% fewer refugee arrivals in camps, which implies considerably longer refugee travel times. However, increasing the capacity at Adjumani by allocating more resources appears to only result in a very limited increase in refugee arrivals (< 4%). Based purely on these results, we find that, in a setting where aid resources are heavily constrained, the default capacity of this camp is close to optimal.
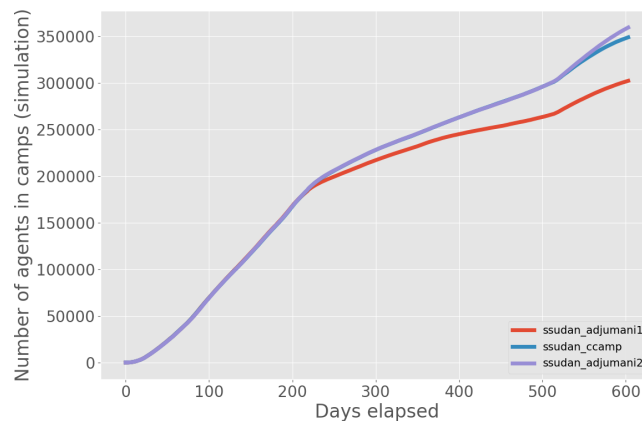

Figure 9: Comparison of number of refugees in camps between three simulations with capacity change for Adjumani camp in comparison to the base model of ssudan_ccamp.

3.9 Finally, we explore how the enforced redirection of arriving refugees from one camp to another can affect the distribution of refugees across all the camps. As an exemplar, we created a scenario (*ssudan_redirect*) where all refugees arriving in Kule, Jewi and Pugnido are redirected to the Tierkidi camp, which has its capacity increased accordingly, creating a counterfactual situation where Tierkidi is the single central camp in Ethiopia receiving refugees from South Sudan. This kind of centralised management of incoming refugees has been known to occur in some other conflict situations, such as Mauritania (Mbera camp) in the North Mali conflict in 2012 Suleimenova et al. (2017a).

3.10 We present a comparison of arrivals across seven camps in both scenarios in Figure 10. Here, Kule, Jewi and Pugnido are excluded from the comparison, as they do not host any refugees in the modified simulation. In comparison to the *ssudan_ccamp* simulation results for individual camps, we attain different distribution of refugees across camps in *ssudan_redirect*. By Day 180, Tierkidi has received twice as many arrivals in *ssudan_redirect*
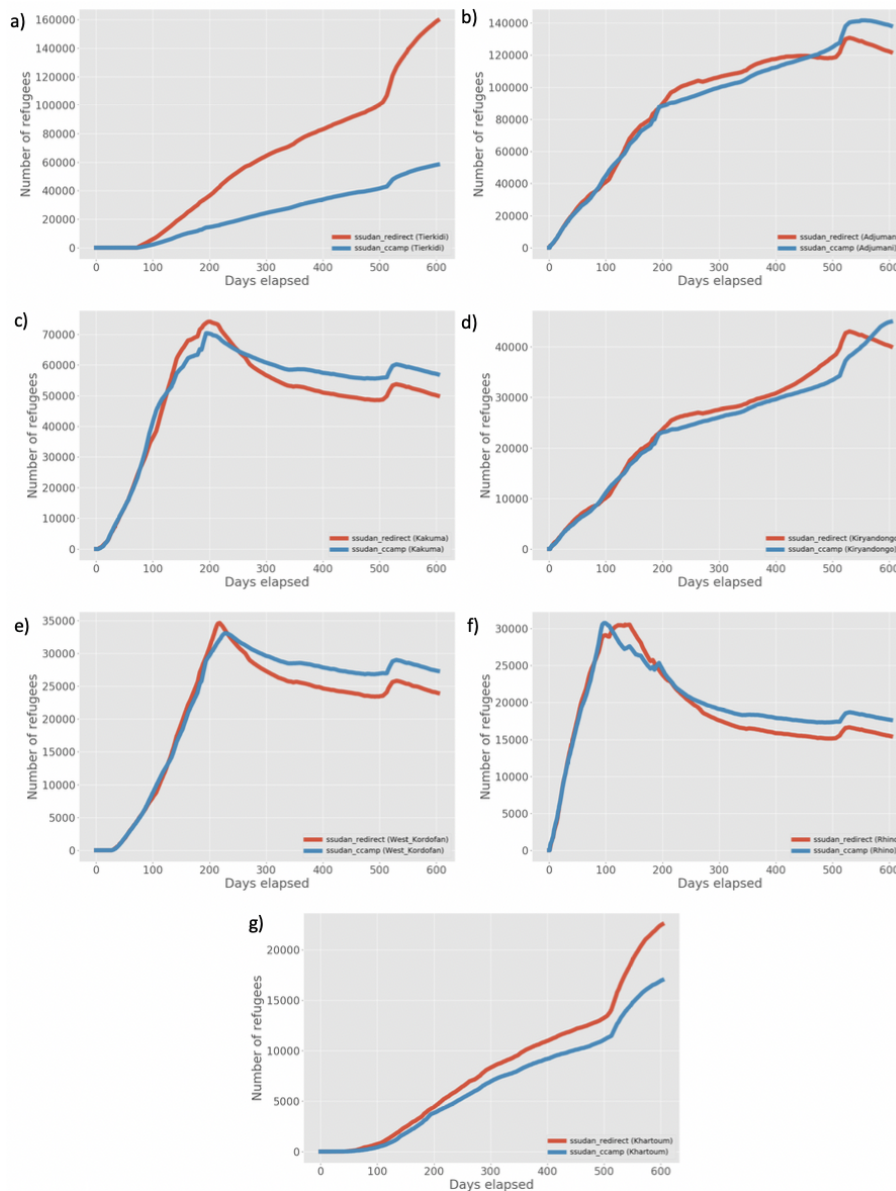
Figure 10: Comparison of the number of refugees in seven camps as forecast by our *ssudan_ccamp* and *ssudan_redirect* simulations for the South Sudan conflict. (**a-g**) Graphs are ordered by camp population size, with the most populous camp on the top to the smallest one on the bottom.

than in *ssudan_ccamp*, while the other six camps retain similar arrival rates. However, after Day 180 the number of refugees in the other six camps becomes lower in *ssudan_redirect* than in *ssudan_ccamp*, while the number of refugees in Tierkidi remains considerably higher. This behaviour can primarily be attributed to the Pugnido camp, which reaches full capacity around Day 180 in *ssudan_ccamp* (see Figure 7), but which is redirected to Tierkidi in *ssudan_redirect*, a camp with a higher (combined) capacity.

Conclusion

4.1 Forecasting forced displacement, especially refugee movements, is both very important and very challenging. Forecasting the distribution of refugee arrivals to potential destinations, as governments and NGOs can efficiently allocate humanitarian resources and provide protection to vulnerable refugees. Through the use of computational modelling and the automation approach presented here, we are able to systematically explore the possible impact of specific policy decisions while accounting for the sensitivity to at least some of individual

parameters and assumptions in the model. To achieve this, we have extended the simulation development approach by Suleimenova et al. (2017a) and used it to forecast refugee arrivals in camps in the South Sudan crisis. Our approach, which relies on the FabSim3-based FabFlee toolkit, publicly available as part of the EU-funded VECMA project (`https://github.com/djgroen/FabFlee`). Though the runs in this paper were all performed on local resources, the FabSim3 toolkit has been applied extensively to execute simulations on supercomputers. We aim to enable this functionality for FabFlee and perform much larger parameter and policy explorations in the near future using it.

4.2 We demonstrated our automated ensemble simulation approach by analysing the effect of policy decisions on refugee journeys in the South Sudan conflict. This conflict is relatively difficult to simulate, primarily due to the lack of roads and difficult food circumstances. While investigating the latter aspect requires new model development and is beyond the scope of this paper, we did update the model to include several walking routes, and were able to achieve a much lower validation error (averaged relative difference) as a result. All policy decisions presented here are purely hypothetical, and largely derived from having observed similar decisions being made in the three conflicts we analysed previously in Suleimenova et al. (2017a).

4.3 In terms of policy decision examples, we incorporated camp and border closures, two camp capacity changes and a forced redirection. As expected, border closure and a reduction in camp capacity result in fewer refugee arrivals as more refugees end up travelling to other camps. Likewise, an increase in camp capacity results in a limited increase in refugee arrivals at the destination camps. However, we also found several unexpected behaviours, such as a lingering effect in prolonged refugee journey times once a border is again reopened, and a clear boost in refugee arrivals when refugees are redirected to a reduced number of camps with larger capacities. These findings help to understand the effects of policy decisions on refugee arrivals and inform other similar conflict situations. We believe these policy decisions in particular warrant more in-depth investigation, using simulation and data analysis approaches that take into more relevant factors and circumstances, and can also leverage the benefits from the automation approach we presented here.

Acknowledgements

Model Documentation

We use the FLEE agent-based modelling code for our migration simulations and FabFlee to automate our migration modelling workflows (Groen et al. 2019b). FabFlee is a plugin for the FabSim3 automation toolkit, and can be installed using a one-liner command once FabSim3 is set up. The Flee code is available at `https://github.com/djgroen/flee-release/releases/tag/1.0` while the FabFlee plugin can be found at `https://github.com/djgroen/FabFlee/releases/tag/v1.0`. Information on how to set up FabSim3 is available at `https://fabsim3.readthedocs.io`. Flee, FabFlee and FabSim3 are all written in Python3 and have been released under the BSD 3-clause license. All output data publicly available on Figshare under a CC-By 4.0 license with DOI `http://dx.doi.org/10.17633/rd.brunel.11395977.v1`.

References

Anderson, J., Chaturvedi, A. & Cibulskis, M. (2007). Simulation tools for developing policies for complex systems: Modeling the health and safety of refugee communities. *Health Care Management Science*, *10*(4), 331–339

Anderson, J., Chaturvedi, A., Lengacher, D. & Cibulskis, M. (2006). Modeling the health of refugee camps: An agent-based computational approach. In D. Lee & B. Nutter (Eds.), *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems*, (pp. 641–645). Washington, D.C.: IEEE

Cannon, B. J. & Fujibayashi, H. (2018). Security, structural factors and sovereignty: Analysing reactions to Kenyaâ ̆AŹs decision to close the Dadaab refugee camp complex. *African Security Review*, *27*(1), 20–41

Collins, A. J. & Frydenlund, E. (2016). Agent-based modeling and strategic group formation: A refugee case study. In T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Z. T. Huschka & S. E. Chick (Eds.), *Proceedings of the 2016 Winter Simulation Conference*, (pp. 1289–1300). Washington, D.C.: IEEE

Disney, G., Wisniowski, A., Forster, J. J., Smith, P. W. F. & Bijak, J. (2015). Evaluation of existing migration forecasting methods and models. Tech. Rep. October, Centre for Population Change, Southampton, UK

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

Entwisle, B., Williams, N. E., Verdery, A. M., Rindfuss, R. R., Walsh, S. J., Malanson, G. P., Mucha, P. J., Frizzelle, B. G., McDaniel, P. M., Yao, X., Heumann, B. W., Prasartkul, P., Sawangdee, Y. & Jampaklay, A. (2016). Climate shocks and migration: An agent-based modeling approach. *Population and Environment*, *38*(1), 47–71

Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K. P. & Wilkinson, H. (2018). Computational modelling of public policy: Reflections on practice. *Journal of Artificial Societies and Social Simulation*, *21*(1)

Gleditsch, K., Salehyan, I. & Schultz, K. (2008). Fighting at home, fighting abroad: How civil wars lead to international disputes. *Journal of Conflict Resolution*, *52*(4), 479–506

Grant, C. (2016). Shifting policy on refugees from encampment to other models. Institute of Development Studies. K4D Helpdesk Report

Groen, D. (2016). Simulating refugee movements: Where would you go? *Procedia Computer Science*, *80*, 2251–2255

Groen, D., Bhati, A. P., Suter, J., Hetherington, J., Zasada, S. J. & Coveney, P. V. (2016). FabSim: Facilitating computational research through automation on large-scale and distributed e-Infrastructures. *Computer Physics Communications*, *207*, 375–385

Groen, D., Knap, J., Neumann, P., Suleimenova, D., Veen, L. & Leiter, K. (2019a). Mastering the scales: a survey on the benefits of multiscale computing software. *Philosophical Transactions of the Royal Society A*, *377*(20180147)

Groen, D., Richardson, R. A., Wright, D. W., 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. V. (2019b). Introducing VECMAtk — Verification, validation and uncertainty quantification for multiscale and HPC simulations. In J. M. F. Rodrigues, P. J. S. Cardoso, J. Monteiro, R. Lam, V. V. Krzhizhanovskaya, M. H. Lees, J. J. Dongarra & P. M. A. Sloot (Eds.), *Computational Science âĂŞ ICCS 2019 19th International Conference, Faro, Portugal, June 12âĂŞ14, 2019, Proceedings, Part IV*, (pp. 479–492). Berlin/Heidelberg: Springer

Hargrave, K., Pantuliano, S. & Idris, A. (2016). Closing borders. The ripple effects of Australian and European refugee policy: Case studies from Indonesia, Kenya and Jordan. Overseas Development Institute. Humanitarian Policy Group. Working Paper. Available at: `https://www.odi.org/sites/odi.org.uk/files/resource-documents/10862.pdf`

Hassani-Mahmooei, B. & Parris, B. W. (2012). Climate change and internal migration patterns in Bangladesh: An agent-based model. *Environment and Development Economics*, *17*(06), 763–780

Hebert, G. A., Perez, L. & Harati, S. (2018). *An agent-based model to identify migration pathways of refugees: The case of Syria*, (pp. 45–58). In L. Perez et al. (eds.) Agent-Based Models and Complexity Science in the Age of Geospatial Big Data, Advances in Geographic Information Science. Springer International Publishing AG

ICG (2014). South Sudan: A civil war by any other name. International Crisis Group. Africa Report N217. Available at: `https://www.crisisgroup.org/africa/horn-africa/south-sudan/south-sudan-civil-war-any-other-name`

Jacobsen, K. (1996). Factors influencing the policy responses of host governments to mass refugee influxes. *International Migration Review*, *30*(3), 655–678

Johnson, R. T., Lampe, T. A. & Seichter, S. (2009). Calibration of an agent-based simulation model depicting a refugee camp scenario. In *Proceedings of the 2009 Winter Simulation Conference*, (pp. 1778–1786). Austin, Texas: IEEE

Kniveton, D., Smith, C. & Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, *21*, S34–S40

Lin, L., Carley, K. M. & Cheng, S. F. (2016). An agent-based approach to human migration movement. In T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka & S. E. Chick (Eds.), *Proceedings of the 2016 Winter Simulation Conference*, (pp. 3510–3520). Arlington, VA: IEEE

Macal, C. & North, M. (2014). Introductory tutorial: Agent-based modeling and simulation. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference*, (pp. 6–20). Savannah, GA: IEEE

Martin, A. (2005). Environmental conflict between refugee and host communities. *Journal of Peace Research*, *42*(3), 329–346

Quinn, J. A., Nyhan, M. M., Navarro, C. C., Coluccia, D., Bromley, L. L. & Luengo-Oroz, M. A. (2018). Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *376:2128*

Reid, K. (2018). South Sudan conflict, hunger: Facts, FAQs, and how to help. World Division. Available at: `https://www.worldvision.org/refugees-news-stories/south-sudan-refugee-crisis-facts`

Salehyan, I. (2014). Forced migration as a cause and consequence of civil war. In E. Newman & K. DeRouen Jr. (Eds.), *Routledge Handbook of Civil Wars*, (pp. 267–278). London: Routledge

Schmeidl, S. (1997). Exploring the causes of forced migration: A pooled time-series analysis, 1971-1990. *Social Science Quarterly*, *78*(2), 284–308

Schmeidl, S. & Jenkins, J. C. (1998). Early warning of humanitarian disasters: Problems in building an early warning system. *International Migration Review*, *32*(2), 471–486

Sfyridis, A., Cheng, T. & Vespe, M. (2017). Detecting vessels carrying migrants using machine learning. In *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2nd International Symposium on Spatiotemporal Computing*. Cambridge, MA: ISPRS

Sokolowski, J. A. & Banks, C. M. (2014). A methodology for environment and agent development to model population displacement. In *Proceedings of the 2014 Symposium on Agent Directed Simulation*. Tampa, FL: Society for Computer Simulation International

Sokolowski, J. A., Banks, C. M. & Hayes, R. L. (2014). Modeling population displacement in the Syrian city of Aleppo. In A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference*, (pp. 252–263). Savannah, GA: IEEE

Suleimenova, D., Bell, D. & Groen, D. (2017a). A generalized simulation development approach for predicting refugee destinations. *Scientific Reports*, *7:13377*

Suleimenova, D., Bell, D. & Groen, D. (2017b). Towards an automated framework for agent-based simulation of refugee movements. In W. K. V. Chan, A. DAmbrogio, G. Zacharewicz, N. Mustafee, G. Wainer & E. Page (Eds.), *Proceedings of the 2017 Winter Simulation Conference*, (pp. 1240–1251). Las Vegas, NV: IEEE

UNHCR (2010). Convention and protocol relating to the status of refugees. United Nations High Commissioner for Refugees. Available at: `https://www.unhcr.org/uk/protection/basic/3b66c2aa10/convention-protocol-relating-status-refugees.html`

UNHCR (2014). Inter-agency appeal for the South Sudanese refugee emergency. United Nations High Commissioner for Refugees. Available at: `https://www.unhcr.org/partners/donors/531f13539/inter-agency-appeal-south-sudanese-refugee-emergency-january-december-2014.html`

UNHCR (2015). South Sudan. Humanitarian response plan. United Nations High Commissioner for Refugees. Available at: `https://data2.unhcr.org/en/documents/download/29790`

UNHCR (2018a). Figures at a glance. United Nations High Commissioner for Refugees. Available at: `https://www.unhcr.org/uk/figures-at-a-glance.html`

UNHCR (2018b). Situations: South Sudan. United Nations High Commissioner for Refugees. Available at: `https://www.unhcr.org/uk/south-sudan-emergency.html`