A Coupled Food Security and Refugee Movement Model for the South Sudan Conflict

Christian Vanhille Campos[1], Diana Suleimenova[2], and Derek Groen[2]

[1] Universidad Complutense de Madrid and Université Paris Diderot.
[2] Dept. of Computer Science, Brunel University London, London, United Kingdom
Derek.Groen@brunel.ac.uk

Abstract. We investigate, through data sets correlation analysis, how relevant to the simulation of refugee dynamics the food situation is. Armed conflicts often imply difficult food access conditions for the population, which can have a great impact on the behaviour of the refugees, as is the case in South Sudan. To test our approach, we adopt the Flee agent-based simulation code, combining it with a data-driven food security model to enhance the rule set for determining refugee movements. We test two different approaches for the South Sudan civil war and find promising yet negative results. While our first approach to modelling refugees response to food insecurity do not improve the error of the simulation development approach, we show that this behaviour is highly non-trivial and that properly understanding it could prove determinant for the development of reliable models of refugee dynamics.

Keywords: multiscale modelling, agent-based modelling, forced displacement, data-driven simulation

1 Introduction

Forced displacement has reached record levels in 2018, and so has the number of people facing severe food insecurity. As of June 2018, 68.5 million people are forcibly displaced worldwide, 25.4 million of which are refugees[1]. According to World Food Programme (WFP)[2], 124 million people are currently facing *crisis* (IPC phase 3) food insecurity conditions or worse, which corresponds to a 15% increase with respect to last year's figures. Despite the current efforts by the global community, not enough is being done to improve the situation. Simulations could prove to be key in trying to understand what drives such crises and predict how they might evolve. This precisely is what a generalized simulation development approach (SDA) and FLEE code proposed by Suleimenova et al. [1] intends to do, attempting to establish accurate and reliable models for refugee destination predictions for a variety of conflicts over different periods of time. Indeed, achieving such a feat could be determinant in helping NGOs and governments alike make better-informed decisions regarding humanitarian support for refugee crises.

In this paper, we focus on the situation in South Sudan. It is a new country that was born in 2011, amidst important tensions and violence, leading to the outburst of civil war in December 2013 due to a conflict between the president and the vice-president. The conflict is characterized first by its ethnic component[3], which has led to claims of mass executions and ethnic cleansing[4]. Second, the conflict is characterized by the use of starvation tactics (such as banning humanitarian aid access to certain regions) to prevent the population from supporting the opposing forces[5]. The famine is exacerbated by the local climate with alternating dry

---

[1] https://www.unhcr.org/figures-at-a-glance.html
[2] https://sway.office.com
[3] https://www.aljazeera.com/indepth/features/2013/12/south-sudan
[4] https://unmiss.unmissions.org/media-briefing
[5] https://www.ipsnews.net/2018/05/south-sudan

and rainy (flood causing) seasons. Overall, the South Sudan civil war has had tremendous consequences for the civilian population, leading to a situation where there are now 7 million people starving (malnutrition is now the first cause of death) and 4.3 million people forcibly displaced (2.5 million of which are refugees)[6].

Given the different aspects of the South Sudan civil war and the famine its population is enduring, it seems fair to wonder what role the food security plays here in the movement of people. Several reports stress that many people are fleeing their homes due to starvation[7]. However, how the food conditions precisely relate to the conflicts remains unclear. Both NGOs and news companies are reporting that the famine, while being strongly linked to the climate, is also heavily influenced by the war [8]. Therefore, to assess how relevant to the modelling task food security is in South Sudan, we have decided to see how it relates to conflict occurrence by conducting a correlation analysis on these representative data sets. We conclude that while there exists a positive correlation between the two phenomena, food security conditions may still need to be accounted to construct a faithful model.

To test this, we explore the consequences of incorporating such results in the aforementioned SDA and on the overall validation outcome of the simulation. To do so, we have coupled FLEE with a data-driven food security model to create a multiscale model, and enhanced the rule set that guides agents' behaviour. We compare our results to those using FLEE alone and to observations by international agencies.

2 Related works

Understanding and forecasting forced displacement movements has become over the past years the main interest for international agencies and governments[9], thus motivating further scientific research in this field [2, 3]. Indeed, we have witnessed a recent increase in interest in the application of computational methods to forced population displacement forecasts, motivated by the complexity of these [4].

While many different techniques are gaining importance, such as machine learning [5], simulation-based approaches and agent-based modelling (ABM) [6]. In ABM, a complex system is modelled as a set of autonomous decision-making agents that behave accordingly with their environment based on a set of rules [7]. This has proven to be especially useful when modelling how individuals and their environment interact and evolve over time, as is the case in migration and forced displacement movements. This technique is thus being used extensively to model armed conflicts response [8], to predict humanitarian assistance need [9], to predict refugee routes [10] or to understand climate effects on migration [11]. However, these current models do not incorporate the effects of famine and starvation on refugee movements.

With regards to refugee arrival predictions, we rely on a generalized SDA, which is based on an ABM and is used to forecast refugee distributions across camps (see Figure 1). A full explanation of the SDA, which is beyond the scope of this short paper, is provided by Suleimenova et al. [1]. We run our simulations using a modified version of the FLEE code[10], obtaining results for the refugee distribution in camps, both without and with a model to reflect the food security aspects. We validate simulations against official UNHCR refugee registration counts.

---

[6]https://www.wfp.org/Situation-Reports/South-Sudan
[7]http://www.unhcr.org/news/briefing
[8]http://reporting.unhcr.org/sites/default/files/UNHCR
[9]https://ec.europa.eu/commission/priorities/migration_en
[10]Available at: `https://github.com/djgroen/flee-release`

The SDA is optimized [12] to allow for easy adaptation to different conflicts, policy choices and assumptions, using the FabFlee automation toolkit [13]. This adaptability, however, should not become a limitation in terms of accuracy, which is why special attention is paid to the relevant parameters the model should take to offer the most reliable results, while being generalized to any given conflict.
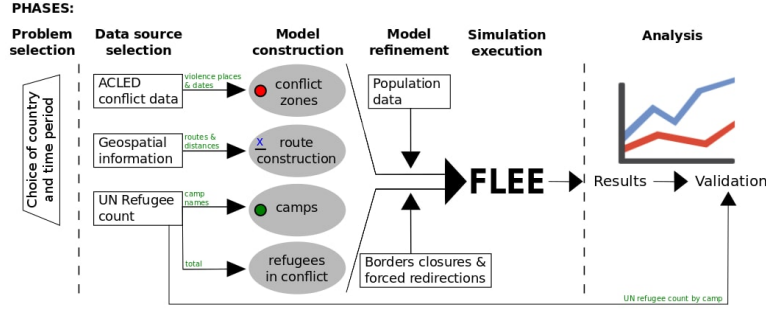

Figure 1: Schematic representation of the generalized simulation development approach and the FLEE code use (source: Suleimenova et al. [1])

3   Input data analysis

To understand the effects of the food security situation on the refugee dynamics in South Sudan, we need to assess how it relates to the conflict. To do this, we need data sources for conflict (or battle) occurrence and famine-like stress classifications. Conflict occurrence information is provided by ACLED[11], while food insecurity indexes are provided by the Integrated Food Security Phase Classification (IPC) project[12]. IPC offers a comprehensive classification of the population by country region and month according to its food security conditions. It distinguishes five different food security phases ranging from *Minimal stress* to *Famine* and classifies the population of different regions according to these.

We look for linear correlations in time and space to examine how these data sets relate to one another, as only a strong linear correlation would justify leaving food security conditions out of the model. We work with two different sets of variables, namely the temporal and spatial analyses, which differ in the normalization. On one hand we normalize the fraction of the population classified as IPC phase 3 or worse[13] and number of conflicts by their maximum over time, while on the other hand we normalize by their maximum for each state[14] (see Equations 1 and 2). We then create scatter plots of these variables with linear least-square fits, first for each different region ($F_{space}(X,t)$ vs $C_{space}(X,t)$ with X a fixed state of South Sudan) and second for each month of the conflict ($F_{time}(x,T)$ vs $C_{time}(x,T)$ with T a fixed month).

Considering that the variables of interest take values ranging from 0 to 1, in most cases the correlation is actually quite weak, with average values of the mean square error (MSE) above 0.05 both for the temporal and geographical analyses, and very few cases where a lower

---

[11]Available at: https://www.acleddata.com

[12]Available at: https://www.ipcinfo.org/

[13]Note that this percentage is often considered to represent the population at risk, and is widely referenced as a good estimator of the food security of a certain region

[14]Note that by defining variables in this way we guarantee that they all take values ranging from 0 to 1 and that they are coherently normalized, making it easier to compare the data sets

MSE than that is observed (see Table 1). Moreover, we observe an important variability in the parameters of the fits (i.e. the slope of correlations), with standard deviations of the same order as or even higher than the average values. Overall, conflict emergence cannot simply portray food security conditions in a model as not only the quality but also the precise form of the correlations between the two data sets is highly variable, making it very difficult to express the effects of one in terms of the other.

$$F_{time}(x,t) = \frac{N^P_{IPC \geq 3}(x,t)}{\max_x[N^P_{IPC \geq 3}(x,t)]} \qquad\qquad C_{time}(x,t) = \frac{N^C(x,t)}{\max_x[N^C(x,t)]} \qquad (1)$$

$$F_{space}(x,t) = \frac{N^P_{IPC \geq 3}(x,t)}{\max_t[N^P_{IPC \geq 3}(x,t)]} \qquad\qquad C_{space}(x,t) = \frac{N^C(x,t)}{\max_t[N^C(x,t)]} \qquad (2)$$

The temporal correlations are largely consistent regarding the sign of the slope, with more than 96% of the cases having a positive relationship as compared to only 27% for spatial correlations. This result shows that famine is not local to conflict zones but is still frequently higher when more conflict events occur. This suggests that starving tactics may indeed have been implemented in this struggle, as indicated in the literature.

Table 1: Comparison between the temporal and spatial agreement of the analyzed data. The percentage of data sets with a *MSE* below 0.05 provides a comparative measure of the quality of the spatial and temporal correlations respectively. Such a value of the *MSE* corresponds to an error of the order of 20% for this set of variables, which is already considerably high. The average slope, its standard deviation and the percentage of data sets with same-sign (positive) correlations provide a measure of the consistency of these. Full results are available at DOI: https://10.17633/rd.brunel.8053340.

	States	Months
$MSE \leq 0.05$	36.4%	11.5%
Positive slope	26.9%	96.2%
Average slope	0.130	0.524
Std. dev. of the slope	0.259	0.267

4 Implementation

As we have presented so far, the food conditions appear relevant to the model, at least in the South Sudan civil war. We now present a simple tentative way to include these features in the model that will also allow for it to be extrapolated to any other conflict to which the model might be applicable. We confront the results obtained by implementing these modifications with the official UNHCR refugee counts and compare the error with the results of the original model offers [1].

Modification 1: We make movechances in other locations dependent on the IPC index of each region at every time-step using the following function: $IPC_{x,t} + 0.3(1 - IPC_{x,t})$, where $IPC_{x,t}$ is the fraction of the population estimated by IPC to be in a *stress* situation in the region of the location at a given time. With this definition we guarantee that, on average, the fraction of the population estimated to be affected by food insecurity conditions leaves such locations, in an attempt to recreate real refugee behaviour.

Modification 2: We assume that food insecurity leads to refugees departing exclusively for that reason. We do this by expanding the list of spawn locations to include both conflict locations and locations with IPC indexes greater than 0. The weighted probability of agents spawning in each location is set as follows: (a) for conflict zones, equal to the full population according to the last census. (b) for IPC locations that are not conflict zones, equal to the population times $IPC_{x,t}/100$.

5 Simulation setup

To implement these changes in the code, we started using FabFlee. The FabFlee tool [12] is an automation tool that facilitates the study of the effects of policy decisions on the refugees' situation, and relies on the Flee code for underlying simulations. FabFlee allows for easy changes in simulation settings, such as border closures, forced redirection, speed changes, etc. The basic workflow of the tool consists of (i) loading the conflict we want to work on, including locations, routes and conflicts information, (ii) performing desired changes to the simulation settings and (iii) instantiating the modified version of the conflict and run it to obtain the results. The advantage of performing multiple different simulations for the same conflict without having to manually edit every part of the code each time makes FabFlee well suited for studying the effects of food security on refugee migrations: instead of having to construct a complete simulation from scratch, we can just add a few modules to FabFlee to account for these modifications. In this way we also ensure that both simulations (with the above-described modifications and without) run under the same general conditions.

To include the proposed modifications themselves we incorporated a separate submodel in the original FLEE code [1], that either updates the move-chances of neutral locations or modifies the spawning probabilities of refugees over time. Starting from an appropriate CSV file containing the IPC index of the different regions at different times, The data-driven food security submodel updates the affected locations' parameters when a change occurs in the IPC index of their region as the simulation evolves. Lastly, we implemented new commands in FabFlee: one to run the simulation including the food security aspects, and one to generate side-by-side comparison graphs of key metrics between both types of simulations.

6 Results

In Figure 2, we present the time evolution of the total error we obtain with the two proposed modifications together with the total error of the original FLEE implementation under the same conditions for comparison. To calculate this error we use the following equation:

$$E(t) = \frac{\sum_{x \in S}(|n_{sim,x,t} - n_{data,x,t}|)}{N_{data,all}} \tag{3}$$

where $n_{sim,x,t}$ gives the number of refugees in camp x at time t according to the simulation and $n_{data,x,t}$ according to the UNHCR data. S is the set of all camps and therefore $N_{data,all} = \sum_{x \in S} n_{data,x,t}$.

In Figure 2a, we compare the movechance modification of FLEE with the original implementation, correcting the original such that the average movechance across locations is identical in both simulations. We perform this correction, because the validation error is known to be sensitive to this parameter. The error behaviour of both runs are highly similar, with the error of the modified implementation being $\approx -1.79\%$ worse on average. These observations imply

that our first proposed implementation of the food considerations doesn't translate to a more accurate prediction of the refugee distribution across camps in South Sudan.

In Figure 2b, we compare the error of the simulation with modified spawning behaviour with the original FLEE run. Here, the movechance is identical in both runs, so no corrections are required. We observe a higher error in the modified run which lasts for 517 days, upon which it drops and reaches up to an $\approx 21.95\%$ improvement on the final day of simulation. The average improvement (fraction of decrease in the error) over time is negative ($\approx -9.49\%$), meaning that this implementation of the food considerations also does not lead to a better refugee distribution forecast in this conflict.

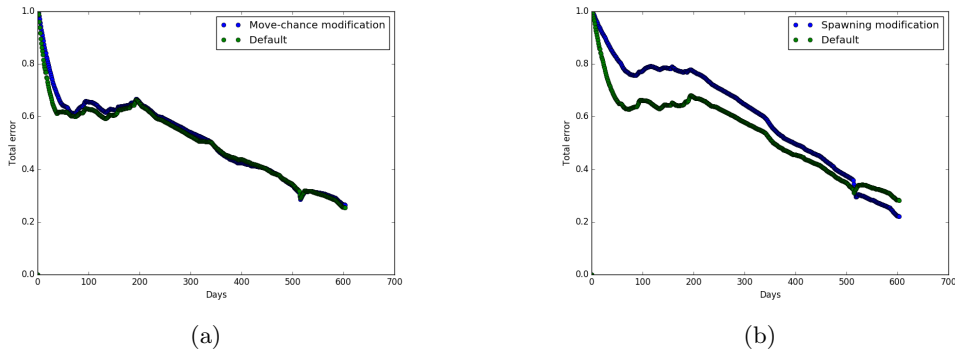(a)   (b)

Figure 2: Total error of the two modified implementations of FLEE compared to the original rule set when using (a) modified movechance calculations, and (b) modified spawn location algorithm. The consistency of these results has been tested by running multiple replicas of the same simulation, obtaining very low variability.

7   Discussion

In this paper, we have explored the importance of food security for refugee movements in the current South Sudan conflict and proposed a computational model based on automated SDA. Although we find that food security conditions are important in refugee dynamics in South Sudan through our data analysis, it has not yet become clear to us in what way the two phenomena exactly interrelate. We investigated two hypotheses in a very tentative fashion as a demonstrator. First, food security influences the likelihood of travelling refugees to depart and second, food insecure locations may act as a cause for refugees choosing to depart in the first place. When implemented in our coupled simulation, neither scenario leads to a lower error, but one of the hypotheses does affect a sensitive parameter in the simulation (movechance).

To understand the relation clearly requires more in-depth investigations. These include bolstering the simulations against the sensitivity of movechance (e.g., by estimating refugee departure dates in the validation set, see [12] for some considerations on this), testing a wider range of hypotheses, and to validate across a much wider range of conflicts. We expect that extensive automation and more sophisticated coupling (as discussed in [14]) will be essential in making this possible. Lastly, we need to do extensive data analysis to improve the model's accuracy. Examples include studies on the origins of the refugees (ethnic, religious, etc.), group dynamics, other environmental conditions and average trajectories. A better understanding of these aspects should then lead to a significant improvement of the model's predictive capacities.

Acknowledgements

References

[1] D. Suleimenova, D. Bell, and D. Groen, "A generalized simulation development approach for predicting refugee destinations," *Scientific Reports*, vol. 7:13377, 2017.

[2] G. Ahmed, Mohammed N. andBarlacchi, S. Braghin, F. Calabrese, M. Ferretti, V. Lonij, R. Nair, R. Novack, J. Paraszczak, and A. S. Toor, "A multi-scale approach to data-driven mass migration analysis," *CEUR Workshop Proceedings*, vol. 1831, 2016.

[3] V. M. Garcia-Guerrero, "A probabilistic method to forecast the international migration of mexico by age and sex," *Papeles de Poblacin*, vol. 22, pp. 113–140, 2016.

[4] S. Edwards, "Computational tools in predicting and assessing forced migration," *Journal of Refugee Studies*, vol. 21, no. 3, pp. 347–359, 2008.

[5] A. Sfyridis, T. Cheng, and M. Vespe, "Detecting vessels carrying migrants using machine learning," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-4/W2, pp. 53–60, 10 2017.

[6] J. A. Sokolowski and C. M. Banks, "A methodology for environment and agent development to model population displacement," in *Proceedings of the 2014 Symposium on Agent Directed Simulation*, (Tampa, Florida), Society for Computer Simulation International, 2014.

[7] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 3, pp. 7280–7287, 2002.

[8] N. E. Williams, M. L. O'Brien, and X. Yao, "Using survey data for agent-based modeling: Design and challenges in a model of armed conflict and population change," in *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications* (A. Grow and J. Van Bavel, eds.), pp. 159–184, Cham: Springer International Publishing, 2017.

[9] A. T. Crooks and S. Wise, "GIS and agent-based models for humanitarian assistance," *Computers, Environment and Urban Systems*, vol. 41, pp. 100 – 111, 2013.

[10] G. A. Hébert, L. Perez, and S. Harati, "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* (L. Perez, E.-K. Kim, and R. Sengupta, eds.), (Cham), pp. 45–58, Springer International Publishing, 2018.

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

[12] D. Suleimenova and D. Groen, "How policy decisions affect refugee journeys in South Sudan: A study using automated ensemble simulations," *Journal of Artificial Societies and Social Simulation*, Submitted.

[13] D. Suleimenova, D. Bell, and D. Groen, "Towards an automated framework for agent-based simulation of refugee movements," in *Winter Simulation Conference (WSC), 2017*, pp. 1240–1251, IEEE, 2017.

[14] D. Groen, J. Knap, P. Neumann, D. Suleimenova, L. Veen, and K. Leiter, "Mastering the scales: A survey on the benefits of multiscale computing software," *Phil. Trans. R. Soc. A (in press)*, 2018.