Causes of Variation in Conflict-Induced Displacement: An Agent-Based Simulation Model

Political Science 418 Final Project

Allen Sirolly*

May 8, 2015

Abstract

This paper uses an agent-based modeling framework to investigate the effects of structural variants on displacement in conflict situations. In particular, it investigates the effect of the size and spatial concentration of household networks on the relative numbers of internally-displaced persons and refugees. The main result is that, within the model, the size of the network is strongly correlated with outcomes, and the spatial concentration weakly correlated. From a policy standpoint, this suggests that governments, humanitarian agencies, and other organizations can prioritize communities based on their network properties to avoid large-scale refugee crises. Given the simplicity of the model, however, the application of the results are limited, and several extensions are proposed.

1 Introduction

One often-overlooked aspect of civil conflict is the plight of displaced persons. Around the world, people who are forced to leave their homes face challenges in relocating to safe havens, adjusting to new living conditions, and, when conditions improve, returning to their former communities. Many are able to find refuge within their country of origin, while others are forced to leave their country entirely. Thus, displacement is an issue that affects not only national governments that are embedded in civil conflict, but also the governments of countries outside zones of conflict.

One of the main distinctions among displaced persons made in both research and policy work is the place of relocation. By the definitions used by the United Nations, an internally displaced person (IDP) is someone who "has been forced to leave his home... and has not crossed an internationally recognized state border"; a refugee, by contrast, is "any person

*Email: sirallen [at] sas.upenn.edu

who, owing to a well-founded fear of being persecuted... is outside the country of his nationality" (United Nations). At the end of 2014, there were 38 million IDPs worldwide, a 15 percent increase over 2013, according to estimates published by the Norwegian Refugee Council. While the number of refugees has not increased quite as substantially as the number of IDPs, there were still nearly 19 million at the end of 2014[1].

Thus, national governments, humanitarian agencies, and other organizations face mounting challenges in controlling the incidence of displacement in conflict-inflicted regions. In particular, these actors need to be able to leverage limited resources in order to best serve the interests of displaced communities and to protect communities from future displacement. By investigating some of the factors that cause variation in displacement outcomes, this paper aims to provide a theoretical basis for allocating those limited resources to communities that face the highest risk of displacement.

Of the sources of variation that could be explored, not all will provide meaningful insights. For example, it goes without saying that, all else held equal, an increase in violence will cause an increase in displacement. Therefore, it is not particularly interesting to focus on violence as an independent variable; instead, the research question that this paper addresses is, *Keeping the pattern of violence fixed, what other factors could cause variation in the severity and nature of displacement?* In this paper, I use an agent-based simulation model to offer an answer. In particular, I find that, within the model, the structure of networks among households at risk of displacement strongly influences the relative numbers of refugees and IDPs. Networks with few connections per household or very localized connections are more likely to produce large numbers of refugees, while those with many connections per household or far-reaching connections will see few refugees compared to IDPs.

Agent-based modeling techniques have been used to study issues pertinent to refugees, such as the spread of disease within refugee camps.[2] However, there are no such studies exploring the topic presented here. Yet, the agent-based approach is appropriate for two

[1] "Global Overview 2015" (7).

[2] For example, see Hailegiorgis and Crooks (2012).

reasons. First, the lack of precise micro-level data in conflict-affected zones makes an empirical analysis of this problem difficult, if not impossible. Second, even if reliable data were available, the complexity of displacement scenarios may obfuscate the effects of the variables of interest. Agent-based simulation, on the other hand, allows the researcher to investigate the question from a theoretical point of view. By introducing variation in the independent variables of interest—in this case, network properties—while excluding other nonessential variables, I can assess the effects of this variation on the dependent variable. If the underlying model is reasonable, I can draw the conclusion that these effects are likely to also manifest themselves, to a greater or lesser extent, in the real world.

The rest of the paper is organized as follows: Section 2 explains the operationalizations of several facts which motivate the model, and outlines the model's mathematical formulation and implementation. Section 3 describes the simulation experiment used to generate data. Section 4 presents the results of the analysis of these data and summarizes key findings. Section 5 discusses the conclusions in relation to the initial problem, as well as the associated policy implications and possible extensions to the model. The Appendix contains the model code and technical details which may have been glossed over in the other sections.

2 Methodology

My model contains many parts but is not particularly complicated. The agents are households, each of which contains a number of members (residents) m_i, as well as a 'capacity' greater than or equal to m_i.[3] The members are represented by an m_i-bit array (i.e., an array of 0's and 1's of length m_i), with the purpose of aiding the counting of IDPs and refugees. Links connect pairs of households and collectively form a network. For simplicity, and to maintain focus on the variables of interest, attacks are modeled exogenously.

Before describing the specifics of the model, it is important to elaborate on the operationalizations of relevant concepts. These are based on a set of stylized facts identified in Schmeidl, Mundt, and Miszak (2010) and UNHCR Action Sheet 13:

[3] As discussed in Schmeidl, Mundt, and Miszak (2010), there is frequently overcrowding in areas of refuge, justifying this excess capacity.

- Given sufficient risk, people will flee their homes preemptively [Brookings];
- IDPs prefer to live with host families, rather than seek refuge in camps [UNHCR];
- IDPs consider several factors when making a relocation choice [Brookings].

The first bullet suggests that risk perception may be an important source of variation in displacement, and should be modeled accordingly. The second suggests that, from a utility perspective, internal displacement is a "better" outcome than extranational refuge. The third suggests that preferences should be modeled to endow households with an appropriate level of agency. As stylized facts, the preceding bullets are generalizations of IDP behavior supported by the literature, although they are not expected to hold in every case. (The Brookings study focuses on displaced persons in southern Afghanistan, and the UNHCR fact sheet is a general guide to conflict-induced displacement situations.) Moreover, there are additional facts which have been observed empirically, but that I have not included in this model.

Subsections 2.1–2.4 describe these operationalizations and the components of the model formally, and 2.5 describes the software used in the implementation. A link to access the model code may be found in the Appendix.

2.1 Attacks

Violent attacks are modeled exogenously by a stochastic process with respect to magnitude and location. The magnitudes M_t are sampled independently from a power-law distribution f, and the locations S_t follow a Gaussian random walk in two dimensions.[4]

$$M_t \sim_{\text{iid}} f(x; \alpha, \underline{x})$$

$$f(x; \alpha, \underline{x}) = \tfrac{\alpha - 1}{\underline{x}} \left(\tfrac{x}{\underline{x}} \right)^{-\alpha}$$

$$S_t = S_{t-1} + \sigma \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t \sim_{\text{iid}} \mathcal{N}_2(\mathbf{0}, \mathbf{I})$$

[4]Following conventional notation in probability, the symbols following a semicolon represent parameters which may be adjusted by the modeler. For example, α and $\underline{x}$ are global variables in the NetLogo code.

The consequence of this specification is that subsequent attacks are expected to be "close" to one another, with the closeness parameterized by σ, the standard deviation of the random walk steps. It is important to note that the purpose of making the attack process stochastic is to add an additional safeguard against the possibility that the simulation results depend on a particular attack path. By running multiple trials and averaging, one may obtain a less biased estimate of the effect of the independent variable.

2.2 Household Risk

Each household maintains a risk level that can increase or decrease depending on the path of the attack process. The risk $R_{i,t}$ follows an exponential moving average

$$R_{i,t} = \alpha \cdot R_{i,t-1} + (1-\alpha) \cdot g(M_t, d(H_i, S_t); \beta)$$
$$g(M_t, z; \beta) = M_t e^{-\beta z}$$

where $d(H_i, S_t)$ is the distance between household H_i and the site of attack at time t, and g is a function proportional to the attack magnitude M_t and decreasing in z (a placeholder for distance), here an exponential decay. Intuitively, a given household's risk level will decay if the attack process drifts farther away, but will increase if attacks continue to occur in the household's vicinity. In the latter case, once the household's risk level exceeds its tolerance T_i, its members will become displaced and search over the household's network for a new residence.

At this point, a brief note is in order. Since the attacks follow a stochastic process, it is possible for households to compute the probability of future displacement, e.g., in the next time step. While this would improve the mathematical elegance of the model, it would assume perfectly perceptive, rational agents. Considering the task at hand, I didn't believe this would add much to (or substantially change) the results. Instead, I chose to use the above formulation as a heuristic which gives the risk model some desirable properties as described above.

2.3 Household Networks

The household network is the main independent variable of interest in this experiment. To initialize the network, I specified two parameters: N, the number of connections per household, and D, the maximum distance at which households may form connections. Different values of N and D generate variation in two key network properties: density and geographic concentration, respectively.

With this formulation, I make several simplifying assumptions. The first is that households' links are independent,[5] which is unlikely to hold in reality. For example, two households that share a common family connection are also likely to be connected. The second assumption is that N and D are global variables; that is, there is no heterogeneity among households with respect to these parameters. While network properties vary spatially in reality, keeping these fixed in the model serves to strengthen the connection between variation in the parameters and variation in the dependent variable (the relative numbers of refugees and IDPs).

2.4 Relocation Preferences

Given that they are displaced, households must choose where to relocate from among the set of options corresponding to their network connections; indeed, according to Schmeidl, Mundt, and Miszak (2010), IDPs "make rational choices in selecting safe havens" (61). In addition, the authors identify five factors that IDPs in Southern Afghanistan considered when relocating: economics (cost), security, livelihoods, community networks, and cultural similarities. For simplicity, I include only those factors which are already present in the model, i.e., distance (cost) and security. In addition, I include separation of family, since this is likely to affect the livelihoods of displaced persons.[6] For each household H_j in its

[5]Although links are formed independently, displaced persons inherit the connections of their new residences, so there is actually a version of dependency in the model.

[6]However, as noted in Schmeidl, Mundt, and Miszak (2010), displaced men in Afghanistan often left their families temporarily to find work.

network, household H_i calculates a cost

$$\text{cost}_{ij} = \beta_{i1} d(H_i, H_j) + \beta_{i2} R_{j,t} + \beta_{i3} I(\text{split})$$

where d is the Euclidean distance function, $R_{j,t}$ is household H_j's risk at time t, and $I(\text{split})$ indicates whether relocating to household H_j would require household i's members to split up ('1') or not ('0').[7] Household H_i then forms preferences which correspond to these costs, i.e., the least costly option is most preferred, and so on.

Starting with the most preferred option, household H_i's members will relocate, given that the target has excess capacity. For each member that relocates, the IDP count is incremented by one. If all of the options have been exhausted and H_i still has remaining members, these members will become refugees, and the refugee count will be incremented accordingly.[8]

2.5 Software

Lastly, I used the NetLogo agent-based modeling platform to implement this experiment. I generated simulation data with the BehaviorSpace tool (described in more detail in Section 3 below) and analyzed it using the statistical software R. More details about the NetLogo language may be found here: `https://ccl.northwestern.edu/netlogo/`.

3 Experiment

While the model formulation described in Section 2 allows for testing multiple sources of variation, I chose to focus on the effect of networks on the total number of IDPs and refugees.

[7] In accordance with intuition, all β_i are restricted to be positive; that is, greater distance, higher risk of the target household, and a family split all correspond to higher relocation cost. Moreover, to keep the relocation decisions 'reasonable', each β is drawn uniformly from a small interval $[0.5 - c, 0.5 + c]$, and the variables are standardized to have mean 0 and range 1. Together, these steps guarantee that households assign comparable weight to each variable.

[8] As a technical note, recall from Section 2 that H_i's m_i members are represented by an m_i-bit array. In the model, individual relocations are ordered so that, in the event that the target household is also displaced in the future, its original members will be the first to relocate. However, H_i's members will still inherit the network connections of the target household. Whether this is a reasonable assumption may be checked empirically.

In particular, the ratio of refugees to IDPs after some number of time steps is the dependent variable of interest. Formally, I hypothesize that the values of the dependent variable will be smaller for networks with larger values of N and D (that is, more connections per household and more geographic spread). Given more relocation options and greater ability to move away from the site of violent attacks, household members should be better able to avoid becoming refugees.

To generate data using NetLogo's Behavior Space, I varied N, D, and σ over a coarse grid of values, as shown below. (The effect of varying σ is to make attacks more scattered, and the purpose is to check whether the key results are robust to this change. As shown in the next section, they are.) For each parameter setting, I ran 50 simulations, and for each simulation, I recorded the total number of IDPs and refugees after 250 time steps, from which I was able to calculate the ratio of refugees of IDPs. Figure 1 shows one such simulation at the final time step.

$$N : \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} \qquad \text{(number of connections per household)}$$

$$D : \{2,\ 6,\ 10,\ 14\} \qquad \text{(maximum connection distance)}$$

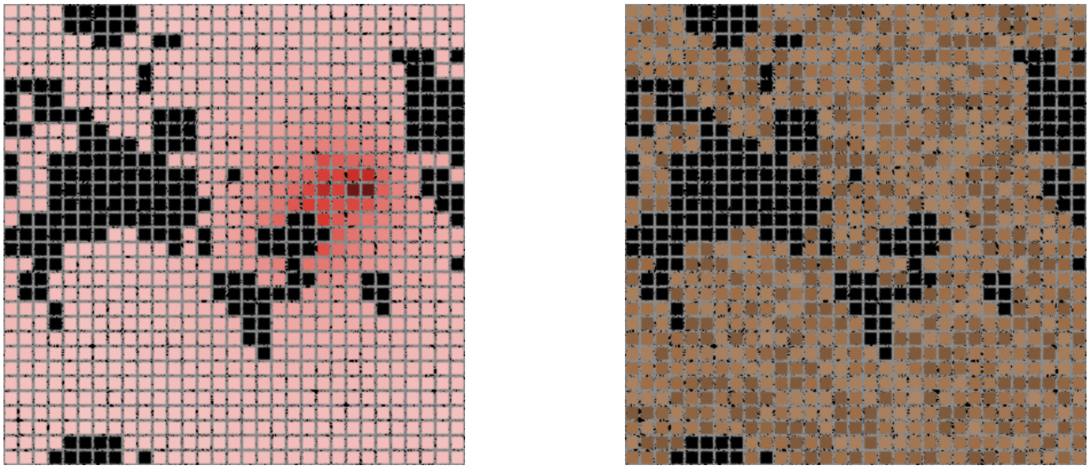
$$\sigma : \{1,\ 4\} \qquad \text{(standard deviation of attack steps)}$$

In total, the experiment consisted of 2400 simulations and took about 90 minutes to run.

4 Results and Analysis

The results are summarized in Figure 2. Each box plot depicts the distribution of the ratio of refugees to IDPs after 250 time steps for different values of N, the number of connections per household, and for a given (D, σ) pair. As mentioned in Section 3, comparing results for different values of σ ensures that they are robust to variations in the attack process.

As seen in Figure 2, for all parameters settings the ratio of refugees to IDPs is decreasing in N. The drop is most significant when N increases from 1 to 2; again for all parameters settings, the mean ratio of refugees to IDPs was greater than one when households only



(a) Household risk view.　　(b) Household capacity view.

Figure 1: Two views of the model. Black squares denote households whose members have been displaced, and are uninhabitable.

had one connection, but less than one when households had two or more connections. The ratio of refugees to IDPs also tends to decrease as D increases, although this relationship is less strong. As seen by comparing the bar plots in both columns, these results hold for both values of σ, indicating that they are robust to changes in the attack process. (On average, the refugee-IDP ratio for a given (N, D) pair was higher for $\sigma = 4.0$ than for $\sigma = 1.0$, indicating that a more 'unpredictable' attack process increased the relative number of refugees.)

5 Discussion

5.1 Conclusions

As established by the second stylized fact in Section 2, households tend to prefer internal displacement—in particular, residence with host families—to extranational refuge, indicating that outcomes with a smaller refugee-to-IDP ratio are "better" from a utilitarian perspective. (As discussed in 5.3, this is limiting.) Given this, the experiment mostly corroborated the hypothesis. However, as discussed in Section 2, the effect was less strong for an increase in D, indicating that, at least within the model, having geographically disperse connections is less important than having more relocation options in absolute terms. This

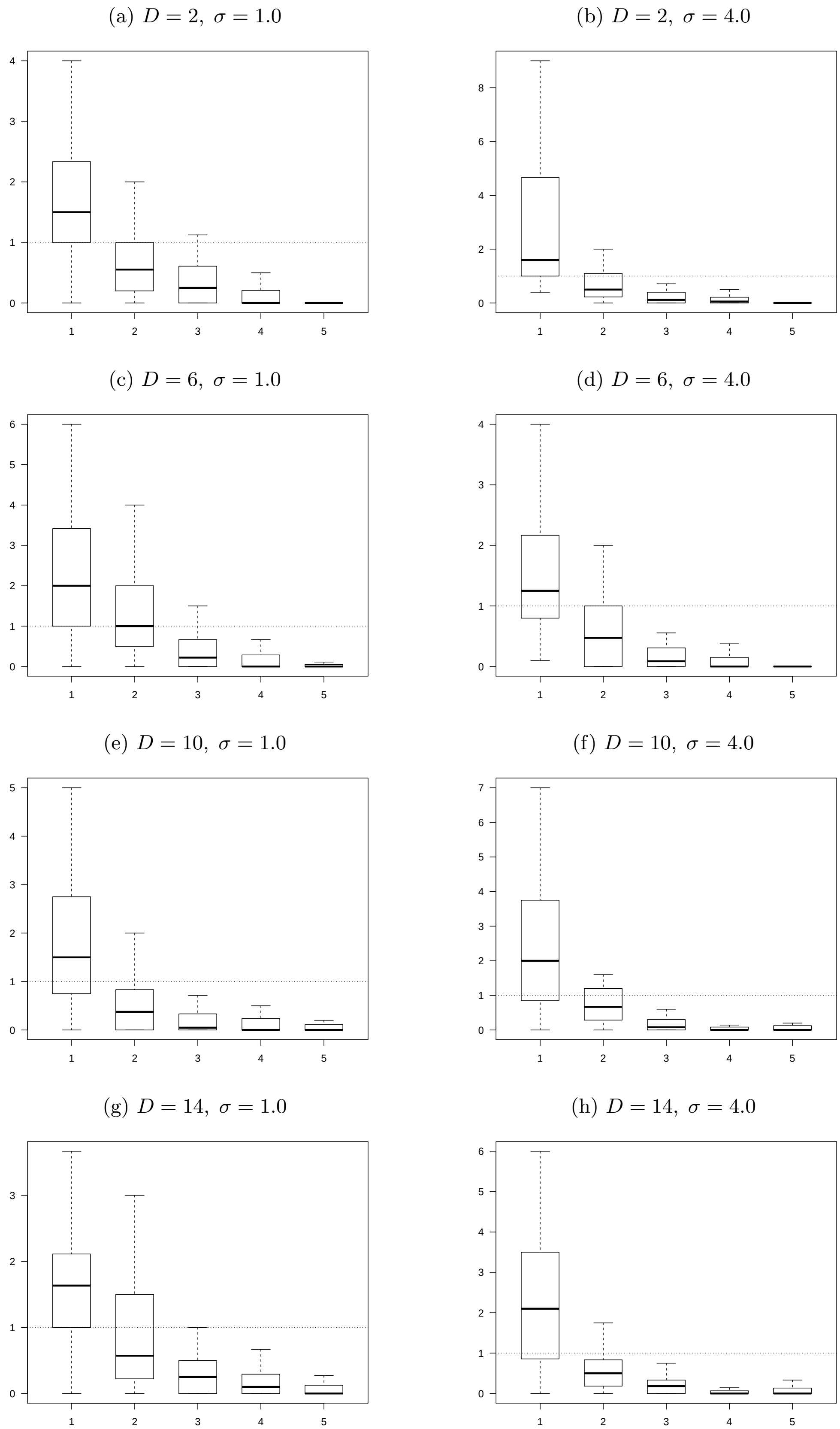


Figure 2: Distribution of the ratio of refugees to IDPs (y-axis) by D (maximum connection distance), and σ (standard deviation of attack steps). x-axis is N, the number of connections per household.

result may have been a consequence of measuring the dependent variable after 250 time steps, at which point the attack process may have affected most of the grid, so that any advantage gained from larger D would have been diminished. Also, the model of household preferences may have fuzzied the result, since the β_i's were randomly chosen, and are unlikely to have reflected the "true" relative weight that households give to distance compared to other factors.[9] Therefore, while the results do point to the importance of networks in determining displacement outcomes, more sophisticated experimental design and analysis are necessary to justify these conclusions and associated policy recommendations.

5.2 Implications

To a limited extent, the results suggest that more attention should be given to networks in war and other conflict-related settings. Given that national governments and humanitarian agencies are often resource-constrained, efficiently prioritizing operations is crucial to the goal of minimizing casualties and displacement. When the networks of at-risk households are large and dispersed, humanitarian interventions may be less efficient at reducing the number of refugees than when networks are small and concentrated. As one example, communities of migrant workers may be less vulnerable in areas under imminent threat, by virtue of having long-distance network connections to their countries of origin. Whether this is indeed the case is an issue of empirical analysis.

5.3 Future Work

There are several improvements which may strengthen the conclusions above and make this model more useful in application (not necessarily listed in order of importance). First, the current specification does not distinguish between risk and displacement; in other words, whether a household becomes displaced depends only on the perceived risk, and not directly on the severity of an attack. In order to test the effect of risk perception (for example, a

[9] These weights are probably tricky to determine empirically. Unlike economic activity like consumption, a relocation decision in the event of conflict-induced displacement is rarely made, so revealed preference cannot be usefully applied. On the other hand, if one assumes that different individuals have similar preferences, then this may be possible.

systematic underestimation of risk) as a source of variation in displacement, it is necessary to model these separately. Second, as mentioned in 2.1, it is not clear whether the networks generated here are 'realistic'. While N and D do generate variation in the network structure, more work should be done to incorporate other properties which match those of real-world household networks. Third, it may be worthwhile to explicitly model humanitarian interventions, to more directly assess the effects of different policies. Fourth, while internal displacement may preferable to refuge from the perspective of household utility, this may not be true from the perspective of other actors such as national governments. This should be taken into consideration—for example, by experimenting with more sophisticated dependent variables and evaluation metrics—before making explicit policy recommendations. Finally, better data collection efforts by governments and humanitarian agencies may allow some amount of empirical analysis as a complement to this agent-based simulation approach. Considering the scale and impact of refugee crises around the world today, this would certainly be a worthwhile investment.

6 Appendix

6.1 Model and Analysis Code

Available here: `https://github.com/sirallen/refugee-model-2015`.

6.2 Note on Power Laws

To draw from a power-law distribution[10], one can use the quantile function, the inverse of the distribution function F:

$$Q(z;\alpha,\underline{x}) = F^{-1}(x;\alpha,\underline{x}) = \underline{x}(1-z)^{\frac{1}{1-\alpha}}$$

where z is drawn from the uniform distribution on $[0,1]$. However, in the attack model, large values of z (which correspond to large-magnitude attacks in the tail of f) resulted in unreasonably destructive attacks. To compensate for this, I drew z uniformly from $[0,1-\delta]$, which cuts out attacks in the $100(1-\delta)$th percentile and above. This fix resulted in a more well-behaved attack process, i.e., without 'extreme' events that unduly influenced the IDP and refugee counts.

[10] As a technical note, it must be that $\underline{x} > 0$ for the power law to be well-defined. Here, $\underline{x}$ is the minimum attack magnitude.

References

"Action Sheet 13: Shelter." Handbook for the Protection of Internally Displaced Persons, United Nations High Commissioner for Refugees, June 2006. [Link]

"Global Overview 2015: People internally displaced by conflict and violence." Internal Displacement Monitoring Centre, Norwegian Refugee Council, May 2015. [Link]

Hailegiorgis, Atesmachew, and Andrew T. Crooks. "Agent-based modeling for humanitarian issues: disease and refugee camps." In *The Computational Social Science Society of America Conference, Santa Fe, NM.* 2012. [Link]

Schmeidl, Susanne, Alexander D. Mundt, Nick Miszak (2010). "Beyond the Blanket: Towards More Effective Protection for Internally Displaced Persons in Southern Afghanistan." Brookings–Bern Project on Internal Displacement, May 2010. [Link]