Title

Modeling pastoralist movement in response to environmental variables and conflict in Somaliland: Combining agent-based modeling and geospatial data

Authors

Erica L Nelson[1,3], Saira A Khan[1], Swapna Thorve[2], Gregg Greenough[1,3]

[1] Harvard T.H. Chan School of Public Health at Harvard University, Cambridge USA

[2] Department of Computer Science at University of Virginia, Charlottesville USA

[3] Harvard Medical School, Department of Emergency Medicine, Brigham and Women's Hospital, Boston USA

Key Words

Agent-based modeling, geo-spatial, environment, conflict, pastoralists, migration, Somaliland

Abstract

Pastoralism is widely practiced in arid and semi-arid lands and is the primary means of livelihood for approximately 268 million people across Africa. Critical environmental, interpersonal, and transactional variables such as vegetation and water availability, conflict and ethnic tensions, and private/public land delineation influence the movements of these populations across space and time. The challenges of climate change and conflict are widely observed by nomadic pastoralists in Somalia, particularly in the regions of Somaliland and Puntland, where resources are scarce, natural disasters are increasingly common, a protracted conflict has plagued communities for decades, and over 65% of the population rely on pastoralism as a primary livelihood. Bereft of necessary, real-time data, researchers and programmatic personnel often turn to post hoc analysis to create an understanding of the interaction between climate, conflict, and migration, and design programs to address the needs of nomadic pastoralists and those that drop out of pastoralism in search for alternate livelihoods. By designing an Agent-Based Model (ABM) that simulates the movement of nomadic pastoralists based on aggregated, typologically-diverse, historical data of environmental, interpersonal, and transactional variables in Somaliland and Puntland between 2008 and 2018, this study intends to identify how pastoralists respond to complex, changing environments over time. The subsequent application of spatial analysis, through Choropleth maps, Kernel Density Mapping and Standard Deviational Ellipses, characterizes the resultant pastoralist population densities in response to these spatio-temporal variables. Outcomes of these analyses demonstrate a large scale spatio-temporal trend of pastoralists migrating to the southeast of the study area with high density areas manifesting in the south of Nugaal, the northwest corner of Sool, and along the Ethiopian border. While minimal inter-seasonal variability is seen, multiple analyses support the consolidation of pastoralists to specifically favorable regions. While this ABM does produce compelling associations between pastoralist movements and terrestrial and conflict variables, it relies heavily on assumptions and incomplete data and is not necessarily representative of on-the-ground realities. Given the paucity of data regarding pastoralist decision-making and migration, validation remains challenging with current methods based on heuristics and descriptions in literature.

Background

Pastoralism is widely practiced in arid and semi-arid lands (ASALs) and is the primary means of livelihood for approximately 268 million people across Africa (FAO, 2018). Pastoral mobility is largely driven by the availability and quality of fodder and water to maintain livestock herds, influencing the spatial and temporal variability of pastoral migration patterns across landscapes (Pas, 2018; Sakamoto, 2016). While the non-sedentary lifestyle of pastoralism can be considered an effective adaptation technique to environmental changes, dependence on natural resources contributes to the risk-averse nature of pastoralism (FAO, 2018; Pas, 2016). Unpredictable climatic changes contribute to the increase in the severity and frequency of natural hazards, irregular rainfall patterns, extreme temperatures, and changing land cover, affecting the availability of natural resources required to support livestock herds (Avis & Herbert, 2016; Onyango, 2016; Sakamoto, 2016). These environmental effects are compounded by ongoing land degradation, land privatization, conflict, and numerous other factors, weakening pastoral systems (FAO, 2018; Onyango, 2016).

The challenges experienced by pastoralist communities are unevenly felt across the continent but are particularly pertinent in Somalia, where approximately 65% of the population relies on pastoralism as a primary source of livelihood (Carr-Hill, R. A., & Ondijo, D. 2012). In the last decade alone, Somalia has experienced numerous devastating environmental shocks over a short period of time. Between 2010 and 2012, a catastrophic drought with subsequent food insecurity and famine that affected a large part of the Horn of Africa affected 13 million people, many of whom were pastoralists (Slim, 2012). Of all the countries affected by the drought, Somalia was arguably the hardest hit in the region, which was aggravated by the inability to provide timely assistance due to instability, conflict, and lack of humanitarian coordination (ibid). Shortly after this period, Somalia experienced another drought between 2016 and 2017 that was followed by lower-than-normal rainfall in the following years, resulting in growing numbers of internally displaced populations (IDPs). The Food and Agriculture Organization (FAO) (2018) noted that it takes approximately five years for a livestock-dependent household to fully recover from the effects of severe drought. The severe droughts in Somalia in 2010 - 2012 and 2016 - 2017 then theoretically did not not give pastoral households time to fully recover. A variety of adaptation techniques are used to protect livestock-related livelihoods in times of drought or conflict, including forming agreements and alliances with members of the community, sharing of necessary resources, and diversification of livelihoods (Shaughnessy, 2018). At times, however, the expansion of private land and environmental degradation results in resource conflicts and occasional livestock death (Onyango, 2016).

The impact of environmental hazards is compounded by ongoing conflict and political instability that has troubled the country for decades. Violence has manifest itself in Somalia in numerous ways. Ongoing armed conflict between the Somali militia, the African Union Mission in Somalia, and non-state actors has led to widespread insecurity and displacement, particularly in Southern Somalia (IDMC, 2018). In the northern part of the country, communal violence prompted by clan/ethnic differences and the quest for political autonomy has been an ongoing problem in the contested regions of Puntland and Somaliland. In this region, Somaliland is working to gain international recognition for being an independent country while quarreling with neighboring Puntland over disputed land (Avis and Herbert, 2016). In 2018 alone, an estimated 578,000 people were displaced due to conflict, while an additional 548,000 people were displaced by regional disasters in Somalia (IDMC, 2018). To date, the total number of IDPs in Somalia is thought to be more than two million people, and this does not account for the many refugees who have fled the country altogether (ibid). Numerous organizations have documented conflict and disaster as the primary drivers of displacement for Somalis, including pastoralists (IDMC, 2018; UNHCR PRMN, 2019).

Pastoral migration, whether voluntary or forced, is a complex phenomenon that scientists continue to grapple with. While the drivers of migrations are broadly understood, less is known about *how* environmental and conflict variables affect patterns of movement across space and time in the past and how they may continue to evolve in the future. To the knowledge of the authors, there is very limited public information available about pastoral populations in Somalia or the self-declared autonomous region of Somaliland. Without information that captures the dynamic nature of pastoral populations, it is extremely challenging to understand the evolving patterns of movement in response to environmental changes and ongoing conflict on a regional scale. To date, several studies use computer models as a means to explore pastoral migration through virtual environments, which will be referred to in the following paragraph. Computer models enable the simulation of scenarios in which the relationships of variables are tested and analyzed over time and/or space providing insight into the observed phenomenon.

Agent-based models (ABM) are a subset of computer modeling techniques that are increasingly being utilized to examine the stochastic realities of human migration in response to environmental and conflict variables (Hailegiorgis et al., 2010). ABMs can simulate the actions of agents, in this case pastoralists, based on the interactions among agents and their surrounding environment. The bottom-up nature of ABMs enables the capture of granular patterns of movement that can be summarized to a systems-wide level. While some models aim to gain a holistic understanding of the adaptive behaviors of pastoralists in response to their environment (Sakamoto, 2016), others consider the possible effects of climate and/or conflict on pastoral movement (Ginetti et al., 2015, Hailegiorgis et al., 2010, Smith, Kniveton, Wood, & Black, 2011), as well as the interplay between pastoral movement, land privatization (Kennedy et al., 2010, Lesorogol & Boone, 2016), and disease transmission (Xiao, Cai, Moritz, Garabed, & Pomeroy, 2015). Computer models rely on a combination of data sources, both qualitative and quantitative evidence, predictive datasets, and/or informed assumptions to generate simulation scenarios. The extensive spatial and temporal variability of pastoral movement and the factors that trigger it warrants the implementation of a simulation model that captures its nuances and the wide-ranging variability.

The integration of geospatial data and agent-based modeling to study pastoral migration is less explored in literature. The existing models and findings, however, have contributed substantially to an ever-growing body of literature on the subject. Sakamoto (2016) integrates low-resolution multi-temporal satellite imagery analysis and agent-based modeling to study pastoral access to resources in dryland vegetation in northeastern Nigeria while the Center for Social Complexity and Department of Computational Social Science at George Mason University has examined the intersection of GIS and agent-based modeling through the development of a range of models on the ABM platform MASON, including the HerderLand, AfriLand, RiftLand, and RebeLand models. The models consider a range of scenarios in Eastern Africa that examine resource contention, the effect of environmental changes, availability of watering holes, and the effect of private land on pastoral movement on multiple scales (Kennedy et al, 2010). To the knowledge of the authors however, no ABM has been designed for or applied to Somaliland while southern Somalia is well studied.

The purpose of this study is to examine the relationship between environmental change, conflict, and pastoral movement in Somaliland between 2008 to 2018 through agent-based modeling. The research generates synthetic movement patterns for nomadic pastoralists in the region, which are influenced by a series of environmental, interpersonal, and transactional variables. Evidently, computer modeling presents high levels of uncertainty but also has immense potential to improve humanitarian preparedness and response. To improve modeling capacities, it is necessary to continue developing the input data and methodology to identify the successes and limitations of the model. This paper adds to a growing body of literature that considers the use of predictive modeling and geospatial analysis to address issues in the humanitarian sector where climate and conflict variables have significant impact on population movement.

Methods

Setting

This agent-based model simulates the movement of nomadic pastoralists in response to conflict and environmental variability in northern Somalia, specifically in the regions of Somaliland and Puntland, between January 2008 and December 2018. Seven administrative regions (Figure 1) comprise Somaliland and Puntland, and the spatial extent of this analysis.

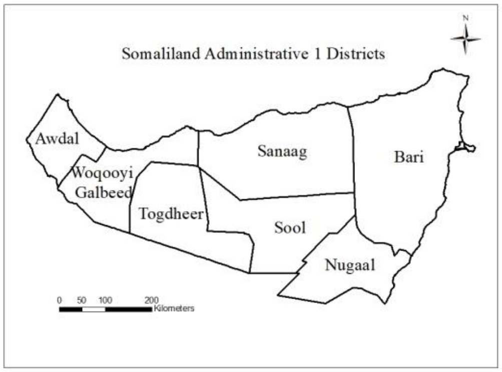


Figure 1. Map of the boundaries of administrative 1 districts in Somaliland and Puntland.

Data collection

Data utilized for this study (Table 1) were obtained from diverse sources, are of disparate typologies, and were quantified, normalized, and utilized in various ways to generate agents, assign attributes to those agents, and develop an environmental model. The following is a description of those data sources, manipulations, and the mechanism for normalization.

Table 1. Data utilized for the agent-based model, including ontological category, type of variable, source and date of collection.

Ontological Category	**Computational Variable**	**Variable Format**	**Data Source**	**Time when data was collected**
Agent demographics				
	Population distribution	Tabular population estimates by	UNFPA Population	2014

		administrative boundaries	Survey Data	
Environment data				
	Base map	Remotely sensed satellite imagery with superimposed administrative boundaries	Google Earth	2019
	Slope	Remotely-sensed raster layer extrapolated from elevation data	DIVA GIS	2008
	Surface water	Raster layer of remotely-sensed water sources utilized to create a NDWI Layer	NDWI Layer obtained from Google Earth	Aggregation from 1948 and 2018
	Artificial water sources	Geo-tagged data on artificial water sources	Somalia Water and Land Information Management (SWALIM)	Data from January 2008 until May 2018
	Vegetation data	Raster layer of remotely-sensed data utilized to	MODIS Terra Vegetation	Data from 2008 until

		create a SAVI layer	Indices 16-Day Global	2018
	Ethnic boundaries	Polygons that delineate a general understanding of geographic ethnic boundaries	Kenya Somalia Consortium, Clan map was digitized and geocoded	Data from 2015, as recorded in 1999
	Conflict data	Geo-tagged conflict data	Armed Conflict Location & Event Data Project (ACLED)	January 2008 through December 2018
	Settlements (Private/public land delineation)	Polygons created utilizing buffers around townships and cities	Metropolitan data provided by the Humanitarian Data Exchange (HDX), uploaded by UNOCHA Somalia	May 2011
	Land Cover	Polygon maps of land cover	Food and Agriculture Organization	May 2007

Manipulation and Normalization

The model environment is comprised of terrestrial variables including slope, surface water, points of artificial water sources, and vegetation rigor; interpersonal components including ethnic boundaries and locations of conflict; and a transactional variable that delineates private and public land. All data values were rasterized and aggregated to a 1 km² grid. Variables were then normalized to range values between 0-1, utilizing the equation, below.

$$Vnormalized = (Vcurrent - Vminimum) \div (Vmaximum - Vminimum)$$

Wherein V is the variable in consideration.

Slope was calculated from remotely-sensed elevation data obtained from DIVA GIS using the Slope Spatial Analyst in ArcMap 10.6.1. The presence or absence of surface water was ascribed a binary score of 1/0 (presence=1) based upon data downloaded through Google Earth Engine. Artificial water sources, such as wells and boreholes, were gleaned from a survey-based, geocoded dataset provided by SWALIM (W. Stephen, personal communication, April 18 2019), with cells containing more than 1 water source having higher values than pixels that contained only one or none.

The vegetation score was calculated on a raster grid containing seasonal median pixel values. To calculate the median pixel values, all available MODIS satellite data was seasonally aggregated, and the median pixel value was calculated. Following this calculation, the Soil-Adjusted Vegetation Index (SAVI), equation below, was applied to those median pixel values for each season in the study period to create a map of proxied vegetation availability. This index has proven more appropriate for the presence of vegetation in arid regions (Vani and Mandla, 2017). In this equation, *NIR* stands for near-infrared values and *L* equals the canopy background adjustment factor, which was set at 0.5 to minimize soil brightness variations and eliminate the need for further calibrations around soil-type (cite).

$$\mathrm{SAVI} = \frac{(1 + \mathrm{L})(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red} + \mathrm{L})}$$

Ethnic boundaries obtained from a static map provided by the Kenya Somalia Consortium were digitized, and rasterized, with buffers applied to the borders at 10 and 20 meters, with descending weighted values of 0.5 and 0.25, respectively. Conflict point data was gleaned from the ACLED database and rasterized. Grid cells that had higher incidence of conflict were attributed higher values, and a temporal lag was assigned within the model, with conflict occurring the previous season having half the effect on environmental favorability the following season.

In the absence of reliable data, the delineation of private and public lands was determined through expert opinion (HHI-UNICEF workshop, Nairobi, Kenya, June 3-4, 2019). Eighty percent of land in Somaliland was estimated to be publicly-owned, and the distribution of private land was assumed to be land which extends approximately 15 kilometers from the centroid of large metropolitan centers and five kilometers from smaller settlement centers. To adhere to the 80% threshold reported, the private land boundaries around cities and towns were modified to 14 and 4 kilometers, respectively.

Overview of the Computational Model

This agent-based model is developed in RePast (Recursive Porous Agent Simulation Toolkit) Simphony 2.6 using a Java-based simulation environment. Repast is a leading, open-source ABM development toolkit specifically designed for social science applications and has been well regarded in comparison to other ABM platforms (Railsback 2006). The Repast development framework provides all the basic functionality required to support the execution of an ABM, including scheduling mechanisms and diverse modeling functions. The established framework enables researchers to add components to customize the model to fit their needs, allowing the environment to be modified and incorporate a range of dynamic variables.

This agent-based model includes two entities: 1) agents, each representing a single nomadic pastoralist household unit, and 2) the physical environment, which is a geospatial landscape composed of both dynamic and static attributes. At the start of the simulation each generated agent is assigned attributes, including their geographic position at the start of the simulation, the name of the administrative unit they fall within, their ethnicity, and clan association. The number of agents generated per administrative unit was informed by a Population Estimation Survey conducted in 2014 (UNFPA, 2014). Within any given administrative unit, the agent start position was randomly generated with two constraints: 1) the agent must not be in unsuitable landscapes including water bodies and areas of bare soil (i.e. sand), and 2) the agent must be located 14 kilometers or more outside a major city and four kilometers or more outside of smaller settlements, i.e. on 'public land'. For every simulation run, the agent start position remains identical.

The gridded physical environment is composed of 1 km² grid cells and has a spatial extent of approximately 490,000 km², which covers the administrative regions in Somaliland and Puntland. The environment was designed to include eight variables, grouped into three thematic components: terrestrial variables, interpersonal variables, and transactional variables (Table 2). The environment variables are categorized to be either pull (attractors) or push (detractors) factors for nomadic pastoralists whose patterns of movement are influenced by the availability of water and suitable grazing land to support their herd. The factors were identified and included based on literature, workshops, and/or discussions with regional experts. Intuitively, the presence of vegetation (as proxied by SAVI) and availability of water are considered to be attractors, while steeper terrestrial slope, proximity to conflict or potential ethnic tension (as proxied by proximity to ethnic borders) are designated detractors.

Variables, such as surface water availability, conflicts, and vegetation cover, are subject to seasonal changes. In this model, four seasons exist per annum. The dry season from December to March is locally referred to as *Jilaal*, which is followed by the long rainy season, *Gu*, from April to June. The dry season that follows, *Hagaa*, spans from July to September while the short rainy season that takes place between October and November is known as *Deyr.* Regarding water availability, during the dry seasons, the surface water layer is disabled, as nomadic pastoralists tend to rely on man-made water sources such as wells and boreholes during these times and when surface water is extremely sparse (HHI-UNICEF workshop, Nairobi, Kenya, June 3-4, 2019). The conflict environmental surface changes seasonally during the model's timeline, with conflict point data being aggregated to each season and changing at the end of the ascribed three-month period. Vegetation availability scores also change seasonally, based upon the the medial pixel value of the imagery available for each defined season, as described above, with applied SAVI scores. The inclusion of these real, longitudinal data creates a reflection of not only the seasonal changes associated with wet and dry seasons, but also the changing environment due to climate and human variables over the course of the simulation.

The favorability score for each land parcel is artificially created through the additive equation:

$$Score_{pixel} = v1 + v2 + v3 - (0.25 * v4) - v5 - (0.25 * v6)$$

Wherein,

v1 = normalized vegetative cover, as calculated with SAVI, where L=0.5

v2 = normalized surface water index (enabled in wet seasons)

v3 = artificial water point sources

v4 = normalized terrain gradient

v5 = normalized conflict frequency

v6 = ethnic boundary

Table 2. Attributable variables that are considered in environmental favorability score, relative impact and change status.

Thematic Variables	**Attributable Variable**	**Impact on Favorability**	**Change Over Time**
NA	Base map	NA	Static
	Gradient of the land	The greater the gradient, the less	Static

Terrestrial	(slope)	favorable	
	Surface water	Proximity to water source is favorable	Changes seasonally, only enabled in wet seasons
	Artificial water sources	Proximity to water source is favorable	Static
	Vegetation data	Higher SAVI score correlates with more vegetation, which is favorable	Changes seasonally
Interpersonal	Ethnic boundaries	Proximity to ethnic boundaries is less favorable (with a gradient buffer of 20km)	Static
	Conflict data	Proximity to conflict is less favorable (no gradient buffer)	Changes seasonally
Transactional	Private/public land ownership	Private land requires a transaction between the land-holder and the pastoralist to establish land-sharing. If no agreement is made, the pastoralist must find another land parcel	Static

Simulating Movement

Agents move throughout the simulated geographic environment based on the fundamental premise that nomadic pastoralists rely on livestock to support their livelihoods and are therefore heavily dependent upon access to water and vegetation (Figure 2). At every time tick, i.e. a month, the agent searches for a cell in the environmental grid with the highest favorability score within a search radius (referred to as scouting range) of its surrounding environment. This scouting range was a random distance generated between a 15- and 30-kilometer radius, based on expert consensus regarding the monthly mobility capacity of nomadic pastoralists. Once the most favorable cell has been identified, the agent moves to this location and determines whether it is located on private or public land. If the cell is public land, they are free to move to that location without any further delay and have the option to stay there for the duration of the season (between one and three months), provided this location continues to have the best favorability score. Once the season changes, the agent is required to seek out further land. This latter constraint is an effort to model resource depletion of that grid given the grazing requirements of a pastoralist herd.

However, if the cell happens to be on private land, the agent must negotiate a land-sharing deal with the local landowners. If the agent successfully makes a deal with the landowner, they are able to remain at that location for the duration of the season. However, if the agent is unable to obtain access, they are required to move to another grid cell within the same time tick, consider whether the new grid location is private or public land, and repeat the process described above. This secondary selection of a cell is determined by the next best score in the gridded environment.

If the agent is unable to make a deal on three separate occasions in any given season, the agent state switches from pastoralist to IDP, at which point the agent exits the simulation or 'drops out'. This is done with the assumption that the lack of access to grazing land leads to the death of livestock, and the pastoralist agent is required to seek alternative livelihoods and will potentially drop out of a purely pastoralist lifestyle.

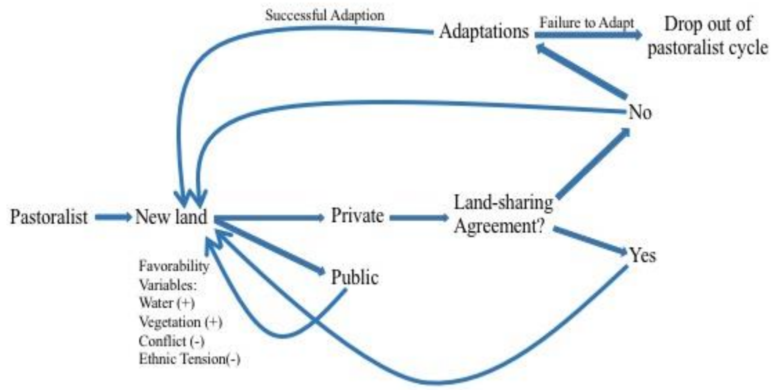


Figure 2. Conceptual model for the ABM of nomadic pastoralist agents seeking new grazing land

The model was run for eleven years, between January 2008 and December 2018, with one month change and decisional intervals. An explanation of the model described in a structured Overview, Design Concepts, Details protocol, which has been amended to include details regarding human decision making (ODD+D) (Müller et al, 2013), can be found in Appendix 1. Given the constraints of computational power, completing the simulation for all 225,767 pastoralist agents was not possible. Thus, a simple random sampling of 10% of the population in each administrative region was conducted, and the agent-based model was executed with this subset of the generated agents (Table 3).

Table 3. Number of synthetic agents generated per administrative unit for the simulation.

Administrative Unit	**Number of Agents (10% of admin pop)**
Awdal	2,851
Woqooyi Galbeed	4,374
Togdheer	2,428

Sool	2,898
Sanaag	4,776
Bari	1,911
Nugaal	3,337
Total	22,567

Analysis

Spatial analysis of this agent-based model is preliminarily aimed at identifying temporal trends of pastoralist density both across the seasons of a year and between seasons separated by a decade of evolving terrestrial and conflict variables. Somalia experiences four distinct seasons each year, as described above. Subsequently, the months of January, May, August, and October were identified as representative midpoints of these seasons and utilized in this spatial analysis.

By analysing the population distribution of pastoralist agents resultant of the ABM simulation during each of these seasonal time points through Choropleth Mapping of population counts, Kernel Density Mapping, and Standard Deviation Ellipses Analysis (SDE), a spatio-temporal understanding of movement is created based upon what can be considered ‘natural’ seasonal dynamics. The comparison of population distribution during specific seasonal periods between the years 2008 and 2018 heightens the analysis to appreciate how seasonal migration patterns have changed over the decade. The juxtaposition of these outcomes with environmental variables, such as artificial water sources, vegetation cover and conflict allow for associations regarding the interplay between changes in resource availability, conflict dynamics and pastoralist movements.

All spatial data was projected in WGS 1894 UTM Zone 38N and all analysis was performed utilizing the ArcGIS 10.7 platform by ESRI.

Population Counts and Differences

Population counts were created by identifying the location of each pastoralist agent at the aforementioned time periods and summed to generate aggregate population counts across the study region. These data were then spatially joined with administrative districts within Somaliland and Puntland. Choropleth maps delineating population counts per administrative district were created excluding those agents that lay outside of boundary lines. Classification was done using Jenks Natural Breaks without justification between the two time periods, as justification would obscure the significant difference in population counts in each district. Differences between the January 2008 population counts and January 2018 and October 2018 counts were calculated to create a change layer. These calculations were undertaken to explore the change in population counts between 2008 and 2018, both taking into account season (January 2008 verses January 2018) and the beginning and end of the simulation. Choropleth maps were subsequently created, and justified Jenks Natural Breaks were applied to create population change classifications.

Kernel Density Maps

Only a 10% random sampling of each of the regional populations were utilized in the ABM simulation, therefore there are a great number of agents (i.e. pastoralists) that go unrepresented when population distribution is mapped. Through the production of a kernel density surface, this underrepresentation is addressed by creating a predictive distribution surface map of pastoralist populations given known spatial inputs. Population density is calculated using a quadratic formula, seen below, with the highest weight ascribed to the known point location and tapering to zero at the edges of the search radius with the predicted population at any given cell in the output raster map being an accumulation of the values for each of the calculated surfaces.

$$Density = \frac{1}{(radius)^2} \sum_{i=1}^{n} \left[\frac{3}{\pi} \cdot pop_i \left(1 - \left(\frac{dist_i}{radius} \right)^2 \right)^2 \right]$$

For $dist_i < radius$

Wherein: i=1,...,n are the input points; pop_i is the population field value of point I, and $dist_i$ is the distance between point i and the (x,y) location (ESRI, n.d.). The cell size (x,y) is 0.0089831528, 0.0089831528 decimal degrees, which is equivalent to a 1 km². Given the UTM projection of the environmental basemap, the analysis was parameterized to a planar method of measurement. By creating these kernel density maps of the pastoralist population over the four seasons of 2008 and 2018, a spatio-visual interpretation of how pastoralist population has evolved over the course of ten years and its seasonal permutations may be developed.

Standard deviational ellipse

An unweighted, standard deviational ellipse analysis was undertaken to characterize the spatial distribution - specifically the location, dispersion and orientation (Wang et al., 2015) of pastoralists over time- utilizing classical statistical methods with the following equations:

$$C = \begin{pmatrix} var(x) & cov(x,y) \\ cov(y,x) & var(y) \end{pmatrix} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} \tilde{x}_i^2 & \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i \\ \sum_{i=1}^{n} \tilde{x}_i \tilde{y}_i & \sum_{i=1}^{n} \tilde{y}_i^2 \end{pmatrix} \quad \text{where}$$

$$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2$$

$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i$$

$$var(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i^2$$

Wherein x and y are the coordinates for each pastoralist at each defined time-point i, {xbar, ybar} represents the mean center for the features and n represents the total number of features (ESRI, n.d.). Given Rayleigh distribution, the two standard deviations applied includes 98 percent of the features. Comparing the size, shape and orientation of SDEs over four different seasons in 2008 and 2018 allows for the detection of relationships between environmental variables and pastoralist distribution not necessarily captured by density mapping alone. While SDEs have been utilized in the past to explore the relationships between environment and criminal activity (Kent and Leitner, 2007 and Chainey et al., 2008), to characterize racial segregation (Wong, 1998), and to assist in outbreak surveillance (Eryando et al., 2012), the application of SDEs to ABM outputs, specifically in the context of pastoralist migration, is novel.

Results

The results of the ABM yielded the location of each pastoralist agent at every time step, along with its scouting range, and the favorability score of the grid in which it inhabited. The following table (Table 4) quantifies the counts of pastoralist agents within each administrative boundary at each seasonal time-point during the periods chosen for analysis. In January 2008, Sanaag and Woqooyi Galbeed districts have the highest number of pastoralists, with 4601 and 3547, respectively, but all districts had populations greater than 2000. In October 2018, after nearly eleven years of simulated environmental change, Togdheer and Nugaal (with 2576 and 2502 pastoralist agents), in the south of the country (Figure 3), have the highest populations, and there is a compelling lack of pastoralists in Woqooyi Galbeed with only 73 inhabitants. In aggregate, there was a 52% decrease in the pastoralist population during the simulation period, equating to a 'drop out' of over 1,000 pastoralist agents per year.

Table 4. Pastoralist Agent Counts per Administrative District over time, a 10% sampling.

Administrative District	Pastoralist Population Counts								
	2008				2018				Total Difference
	Jan	May	Aug	Oct	Jan	May	Aug	Oct	
Bari	2405	2507	2447	2440	1574	1560	1528	1522	-883
Nugaal	2627	2555	2511	2519	2520	2522	25009	2502	-125

Sanaag	4601	4025	3692	3523	1784	1790	1777	1772	-2829
Sool	3405	4207	4299	4368	2591	2500	2345	2210	-1195
Togdheer	3179	3788	3818	3770	2391	2417	2465	2576	-603
Woqooyi Galbeed	3547	2839	1997	2041	70	78	68	73	-3474
Awdal	2786	2458	2022	1855	171	166	157	153	-2633
Total (Includes those that fall outside of specific districts)	22,575	22,461	20,802	20,586	11,114	11,051	10,853	10,819	-11756

All districts portray a decline in pastoralist population counts, resultant of pastoralist agents 'dropping out' of purely pastoralist cycles (Figure 3). However, certain districts have significantly greater declines in population than others, with Nugaal demonstrating only a decrease of 125 pastoralists, and Woqooyi Galbeed having a decline of almost 3,500. Also notable is the difference in population counts between administrative districts over time. In January 2008, the difference in population count between the most (Sanaag) and least (Bari) populated districts was 2,196, as opposed to October 2018 in which the difference between the most (Nugaal) and least (Woqooyi Galbeed) populated districts was 2,429. In general, the greatest decline in population per district was noted in the north-east of the study region with very little change when accounting for seasonal variability between January and October of 2018.

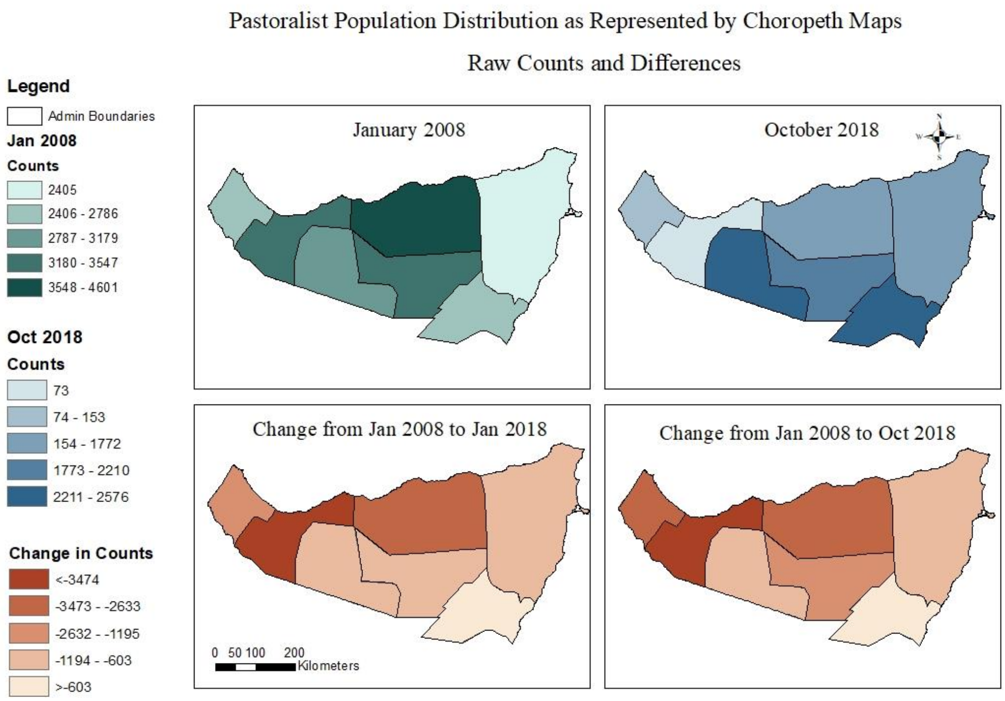


Figure 3. Choropeth maps of pastoralist agent population count and count difference at each time point defined.

Kernel Density

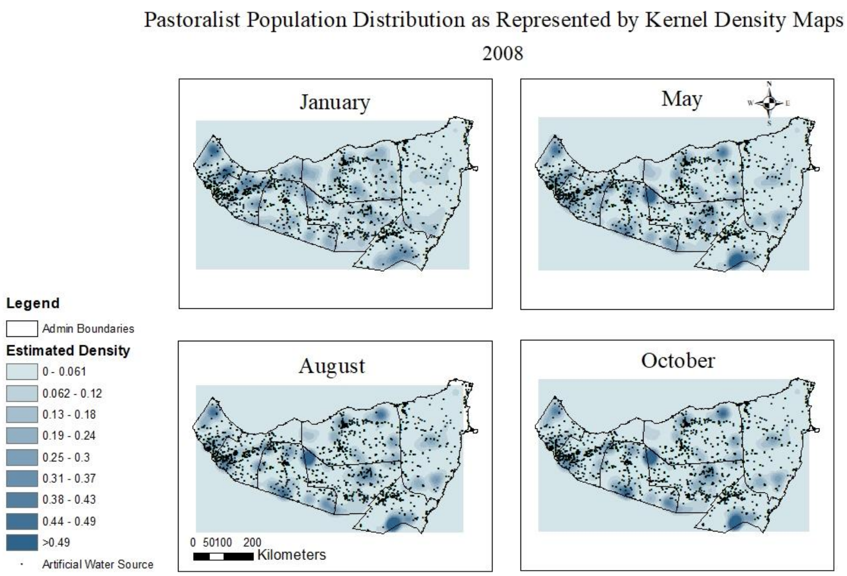
Kernel density (KD) analysis for the four seasonal time points in both 2008 and 2018 demonstrates pastoralist population density throughout Somaliland and Puntland as simulated by this ABM (Figure 4) with statistical characteristics identified in Table 5. In January 2008, the density across the study area ranged from 0 to 0.55 agent per square-kilometer with a standard deviation of 0.71. The map in Figure 4A shows that the agent density is fairly dispersed, but the highest densities are found in Awdal, the westernmost administrative region, while, Bari, the easternmost administrative region has the lowest estimate density. It is important to consider that January 2008 is the first month of the simulation, directly before which the agents were randomly generated in each administrative region, so the spatial dispersion is likely a remnant of the agent generation.

Table 5. Statistics of pastoralist population Kernel Density Maps disaggregated by time point

Month/Year	Minimum	Maximum	Mean	SD
January 2008	0	0.55	0.049	0.073
May	0	0.96	0.047	0.085
August	0	1.1	0.046	0.093
October	0	1.1	0.043	0.09
January 2018	0	1.1	0.024	0.078
May	0	1.2	0.023	0.079
August	0	1.5	0.023	0.086
October	0	1.4	0.023	0.083

Four simulated months later, in the following May, the maximum population density increased to 0.96 agents per square kilometer, without a large change in standard deviation. Certain high-density areas noted in January 2008, specifically in the south of Nugaal and the northwest corner of Sool, persist as relatively dense areas throughout the remainder of the simulation (Figure 4A). By August 2008, maximum density continues to increase, and additional clusters begin to form in Togdheer along the Ethiopian border as well as near the coast in Sanaag. Towards the end of 2008, high- and low-density locations appear to have reached homeostasis, with both large scale spatial patterns and standard deviations of population density remaining similar.

By January 2018, the mean density of pastoralist agents has approximately halved (0.024 from 0.043) resultant of pastoralist 'drop out'. Areas of highest population density are now centralized predominantly within four districts: Sanaag, Togdheer, Sool and Nugaal (Figure 4B). Certain early areas of pastoralist accumulation, such as that in the south of Sool and along the Ethiopian border persist as the most densely populated. In general, as the simulation progresses, pastoralists appear to tend towards clustering, with maximum densities at 1.5 and 1.4 pastoralists per $1km^2$ in the final two seasons. These areas of high density seem to shift eastwards over the eleven-year period, or inversely, the density of pastoralists in the western districts of Awdal and Woqooyi Galbeed declined significantly.



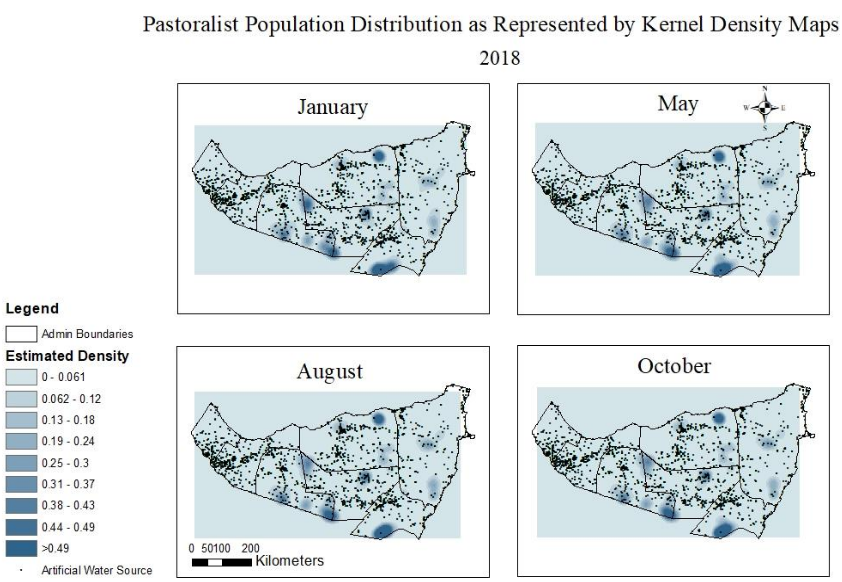


Figure 4.A and B: Kernel Density Maps of pastoralist population agent position at ascribed seasonal timepoints as determined by the ABM simulation for 2008 (A) and 2018 (B). Density unit classifications have been justified across all analyses.

These density maps have been overlaid with artificial water sources, which consists of wells, boreholes, dams, and other artificial facilities. The artificial water sources were included in the ABM and were one of the pull factors in the simulation. In Sanaag, Sool, Woqooyi Galbeed, the areas in which high numbers of water points are situated, understandably evidenced higher densities of agent concentration during 2008. However, this association was weakened when appreciating population density in 2018. Some relationship could be intimated between high density areas and artificial water sites in northern parts of Sool and Sanaag, but as of January 2018, there are no high-density clusters in the westernmost part of the country despite the presence of artificial water sources.

Standard Deviational Ellipse

Standard Deviational Ellipses (SDE) were produced for four time points in 2008 and 2018, resulting in the creation of eight SDEs (Figure 5 with statistical descriptions in Table 6 and 7). In 2008, there was minimal variation in the SDE area for all four seasonal time points, which ranged between 578,912 km² to 602,925 km² with a mean area of 595,865.25 km² and a difference of 24,013 km². The centroid of the ellipses experienced a minor shift in a Southeastern direction, with the rotation remaining consistent across seasons. This analysis indicates that pastoral grazing in 2008 was spatially dispersed across the region and portrayed only minor seasonal variation throughout the year.

Table 6. Characteristics of standard deviational ellipses in 2008

Month	Area km²	Centroid (X,Y) in meters	Rotation angle
January 2008	578,912	694,477.91, 1,057,479.56	101.32
May 2008	601,005	696,424.92, 1,047,957.08	100.05
August 2008	600,619	711,125.21, 1,041,804.01	99.69
October 2008	602,925	712,926.92, 1,039,990.71	99.29

Table 7. Characteristics of standard deviational ellipses in 2018

Month	Area km²	Centroid (X,Y) in meters	Rotation angle
January 2018	501,606	805,533.81, 996,938.63	89.25
May 2018	501,688	804,695.96, 996,794.97	88.24
August 2018	505,096	806,454.28, 995,321.20	88.39
October 2018	503,455	805,817.12, 995,725.31	87.77

In 2018, the area of the SDE ranged from 501,606 km² to 505,096 km² (Table 7) with a mean area of 502,961.25 km² and a difference of only 3,490 km². The centroid of the SDE moved slightly between the seasons and, similar to 2008, the rotation only varied by a few degrees. As demonstrated in Figure 5, the SDEs calculated for 2018 largely exclude the district of Awdal, in the west. When comparing the 2008 and 2018 standard deviational ellipses, these results support previous analysis that the overall grazing patterns of pastoralists have become more compact, with a mean SDE areal decrease of 92,904 km². Similar to the outcomes of the Kernel Density Maps, the SDEs demonstrate a shift in pastoralist density towards the southeast. And while there is minimal seasonal variability within years, it is notable that SDE areal differences between seasons were higher in 2008 at 24,013 km² in comparison to only 3,490 km² in 2018, evidencing a decrease in variability towards the end of the model. It is important to note, however, that this analysis does not account for the variability within the 11-year period and only considers the beginning and end years to capture the absolute change over time.

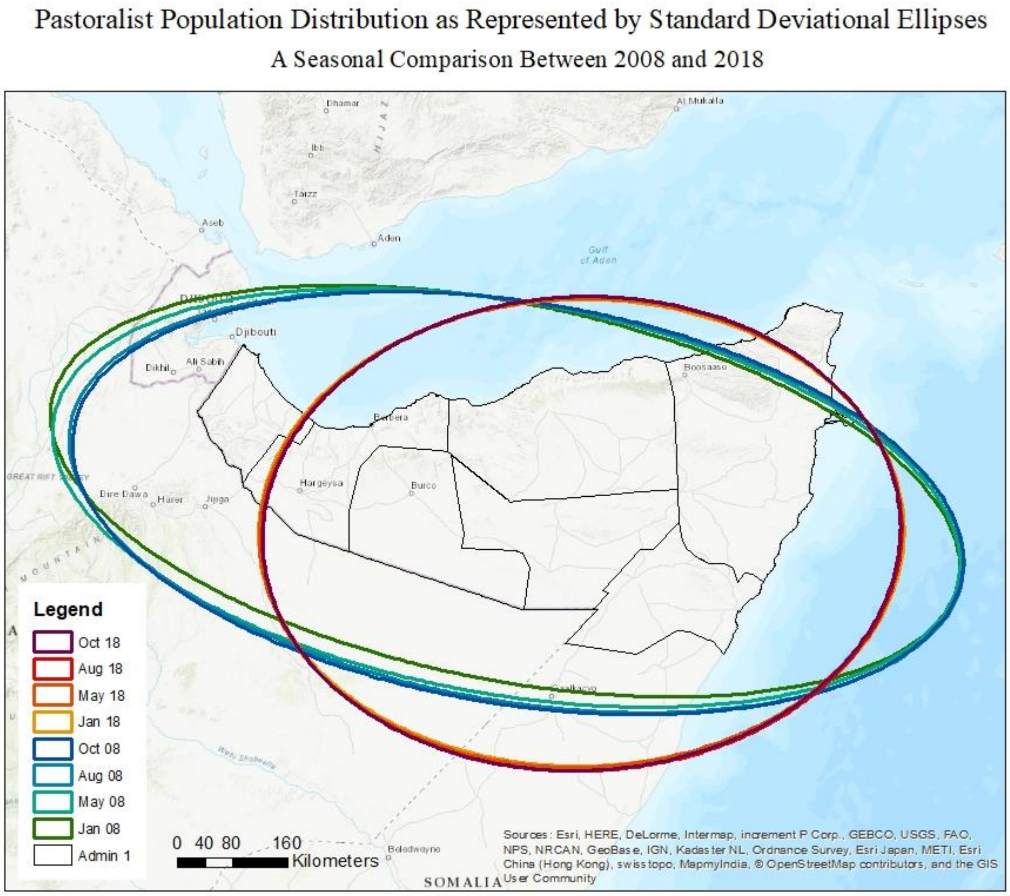


Figure 5. Pastoralist population distribution as demonstrated by Standard Deviational Ellipses, disaggregated seasonally for 2008 and 2018. Two standard deviations were applied to capture 98% of features.

Discussion

This model's findings show minimal seasonal variations within years of population distribution. These outcomes are unanticipated, given the assumption that during dry periods when water and vegetation are more scarce, greater degrees of migration would occur. It is feasible that the inclusion of water and vegetation proxy indicators into a simple, non-weighted, favorability score masked any potential effects at this time scale. Further discussion and validation of the weight of these variables in the decision to migrate needs to be undertaken to better model any presence or absence of seasonal movement.

In general, these analyses do demonstrate compelling trends of consolidation, clustering and a large-scale spatial movement of agent density in the south-easterly direction. Between April and December of 2016, Somalia experienced a significant drought and tremendous food insecurity (FEWS NET, 2017). The districts of Sool, Sanaag, and Bari were relatively spared, which may account for the general spatial trend of simulated pastoralists migrating to the southeast as a whole and to Sool and Sanaag, specifically. While Awdal was also relatively less affected, this region showed a significant drop in pastoralist populations within our ABM simulation, intimating that further environmental, conflict or transactional factors were at play. Similarly confounding is the manifestation of high pastoralist density in the south of Nugaal, where rainfall during this season was only 30-40% of a thirty year average. The positioning of artificial water sources did appear to yield some influence over pastoralist density in 2008, but these associations largely resolved by the end of the model period and do not account for the persistence of this density location or the trends over time. A more rigorous interrogation of the environmental favorability surface could yield an explanation for which variables, such as conflict, topology, or private/public landholding swayed pastoralist dynamics.

Existing studies that use ABM to simulate pastoral mobility have typically used the availability of water and pasture as drivers of migration. Water sources can include surface water, artificial sources, or a range of other water sources. Some studies have even incorporated complex hydrological modeling and have incorporated information on rainfall. The studies that consider vegetation and pasture availability most commonly incorporate the Normalized Difference Vegetation Index (NDVI), which captures vegetation vigor in any given area. While NDVI is a widely used vegetation index in academic studies because it is highly correlated with leaf-area index and biomass, which are attributes of measuring greenness, it is acknowledged that soil reflectance interferes with NDVI calculations, not accurately capturing the vegetative ground conditions. Due to the known limitations of NDVI, this research instead uses SAVI to more accurately capture the vegetative cover of Somaliland and

Puntland, which is sparsely vegetated. Beyond water and vegetation, some have incorporated historical conflict while others actually identify areas of possible contention instead.

Typically, studies have considered the incorporation of variables one by one, where the agent makes a decision based on a decision tree. This study however, normalized all the variables to have values between 0 and 1, after which they were aggregated to a single favorability score. The favorability score is considered in relation to the score of the surrounding grid cells prior to the agents physically moving to another cell. While there is one favorability score, the data layers can be weighted in a variety of combinations to understand the effects of individual variables rather than the collective whole. Following the generation of ABM results, geospatial analysis techniques were used to analyze the data. The integration of geostatistics and ABM is for this particular application is not commonly documented, so the use of these methods and the generation of the results contribute valuable information to the academic sphere and unifies the multiple drivers of migration through a spatial and geographic lens.

In building the conceptual model, we invoked a number of disciplines for our assumptions. Social science, climatology, conflict studies and spatial epidemiology all contributed to its development. Such a multi-disciplinary approach is required to understand real-world complex ecosystems and for us, to ideally produce data that would allow us to create inferences between change in terrestrial, interpersonal and transactional variables and population movement and density.

Limitations and challenges

The model in theory includes a large number of agents making monthly decisions based on the influence of eight variables over a 10-years period resulting in hundreds of thousands of potential data points, requiring a vast amount of computing power. For us, this is limiting for two reasons: first, it required us to use a sample of agents, introducing some degree of statistical error that we could not determine after running the model; second, we could only look at beginning and end point outcomes (2008 and 2018) and not iterative trends (points in time) over the study time period that could be linked to specific climate or conflict events. Arguably then, to get the benefit of a full simulation, ABMs must have the computing power available and this fact could significantly limit its field applicability.

The model also makes a number of assumptions about pastoralist behavior that while duly gleaned from informed local non-governmental service providers, have not been scientifically validated or have an evidence-based understanding of real-time migration beyond anecdotal observations. As a result, the variables are weighted based on these assumptions. As such migration patterns cannot be fully validated. Without clear behavioral and decision-making benchmarks, this makes the model output challenging to validate; rather, validation is done visually using heuristics from the social science domain literature and expert opinion. That said, the benefit of the agent-based model presented here provides some insight into pastoralist behavior and further invites theory and generates hypotheses for deeper study.

Conclusion

The agent-based model, introduced here, is a useful tool to understand the behavior of individuals in a spatial and temporal context. Short of tracking individuals prospectively and manually querying their decision-making, the ABM, coupled with an ethnographic understanding of the factors and drivers of livelihood decisions, can provide a dynamic view of population ecosystems. In the case of nomadic pastoralists in the Horn of Africa where the history of seasonal migration lends itself to a simulated model, the ABM affords the ability to study the complexities of individual attributes in an aggregated and collective fashion with the added benefit of exploring layers of geographic and sociological variables.

In the humanitarian sphere, two critical independent and interdependent variables have the potential to trigger migration: climate-related environmental conditions, of which there are several, and conflict. Somaliland, with its long history of ethnic conflict over land use, its exposure to drought and water scarcity, and a basal level of livelihood migration amongst its nomadic pastoralists, offers a crucible in which to explore how these variables interplay in a model, assuming non-linearity and the need for non-parametric approaches to spatial analysis.

The use of ABM to understand pastoral migration in Somaliland and the ways it which it changes in response to environmental variability and conflict highlighted several important findings. This model was developed with limited data sources and capacity for large scale analysis, and still demonstrated the potential value added of the methodology in understanding pastoralist migration. In an attempt to evolve this model into one that is better representative of reality, the researchers intend to further refine the model, validating behavioural assumptions and including information and decisional pathways regarding economic markets, the effects of livestock disease, pastoral adaptation techniques, and resource depletion to develop a more holistic view of the pastoralist system in Somaliland and Puntland.

References

Avis, W., & Herbert, S. (2016). *Rapid fragility and migration assessment for Somalia*. 47.

Boone, R. B., & Lesorogol, C. K. (2016). Modeling Coupled Human–Natural Systems of Pastoralism in East Africa. In *Building Resilience of Human-Natural Systems of Pastoralism in the Developing World* (pp. 251-280). Springer, Cham.

Carr-Hill, R. A., & Ondijo, D. (2012). Assessment of the Education, Livelihoods, Living Conditions and Welfare Of Somali Pastoralists: A Representative, Multi-Sectoral Survey Conducted to Provide a Baseline for Programming.

Cioffi-Revilla, C., & Rouleau, M. (2009). MASON AfriLand: A regional multi-country agent-based model with cultural and environmental dynamics.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security journal*, *21*(1-2), 4-28.

Eryando, T., Susanna, D., Pratiwi, D., & Nugraha, F. (2012). Standard Deviational Ellipse (SDE) models for malaria surveillance, case study: Sukabumi district-Indonesia, in 2012. *Malaria Journal*, *11*(1), P130.

ESRI. (n.d.). How Directional Distribution (Standard Deviational Ellipse) works. Accessed through https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-directional-distribution-standard-deviationa.htm#targetText=The%20ellipse%20is%20referred%20to,the%20axes%20of%20the%20ellipse.&targetText=The%20latter%20is%20termed%20a%20weighted%20standard%20deviational%20Ellipse.

ESRI. (n.d.). How Kernel Density works. Accessed through https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm

FAO. (2018). Pastoralism in Africa's drylands. Rome. 52 pp. Licence: CC BY-NC-SA 3.0 IGO.

FEWS NET. (2017). Severe drought, rising prices, continued access limitations, and dry forecasts suggest Famine is possible in 2017. Accessed through http://fews.net/east-africa/somalia/alert/january-16-2017

Ginnetti, J., & Franck, T. (2014). Assessing Drought Displacement Risk for Kenyan, Ethiopian and Somali Pastoralists–Technical Paper. *Internal Displacement Monitoring Centre, Last modified April*, *26*.

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological modelling*, *221*(23), 2760-2768.

Hailegiorgis, A., & Kennedy, W., & Rouleau, M., & Bassett, J., & Coletti, M., & Balan, G., & Gulden, T. (2010). An Agent Based Model of Climate Change and Conflict among Pastoralists in East Africa.

IDMC. (2018). Somalia. Accessed through http://www.internal-displacement.org/countries/somalia

Kennedy, W. G., Gulden, T., Hailegiorgis, A. B., Bassett, J. K., Coletti, M., Balan, G. C., Gulden, T., & Cioffi-Revilla, C. (2010). An agent-based model of conflict in East Africa and the effect of the privatization of land. In *Proceedings of the 3rd World Congress in Social Simulation*.

Kennedy, W. G., Hailegiorgis, A. B., Rouleau, M., Bassett, J. K., Coletti, M., Balan, G. C., & Gulden, T. (2010). An agent-based model of conflict in East Africa and the effect of watering holes. Conference on Behavior Representation in Modeling and Simulation.

Kennedy, W. G., Hailegiorgis, A. B., Rouleau, M., Bassett, J. K., Coletti, M., Balan, G. C., Gulden, T., & Cioffi-Revilla, C. (2010). MASON HerderLand: Origins of Conflict in East Africa.

Kent, J., & Leitner, M. (2007). Efficacy of standard deviational ellipses in the application of criminal geographic profiling. *Journal of Investigative Psychology and Offender Profiling*, *4*(3), 147-165.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., ... & Schwarz, N. (2013). Describing human decisions in agent-based models–ODD+ D, an extension of the ODD protocol. *Environmental Modelling & Software*, *48*, 37-48.

Onyango, P. (2016). Nomadic Pastoralism and Natural Resource Use Conflicts in East Africa. Yanda P. Z., & Mung'ong'o, C. G. *Pastoralism and Climate Change in East Africa.* Dar es Salaam, Tanzania: Mkuki na Nyota Publishers Ltd.

Pas, A. (2018). Governing Grazing and Mobility in the Samburu Lowlands, Kenya. *Land*, *7*(2), 41. https://doi.org/10.3390/land7020041

UNHCR. (2019). Somalia Internal Displacement Monitored by UNHCR Protection and Return Monitoring Network (PRMN). Accessed through https://unhcr.github.io/dataviz-somalia-prmn/index.html#reason=&month=&need=&pregion=&pdistrictmap=&cregion=&cdistrictmap=&year=2019

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, *82*(9), 609-623.

Sakamoto, T. (2016). Computational Research on Mobile Pastoralism Using Agent-Based Modeling and Satellite Imagery. *PloS One*, *11*(3), e0151157. https://doi.org/10.1371/journal.pone.0151157

Shaughnessy, S. (2018). How do East Africa's Pastoralists deal with risk? (Unpublished Dissertation). University of Cambridge, Cambridge, United Kingdom.

Slim, H. (2012). IASC Real-Time Evaluation of the Humanitarian Response to the Horn of Africa Drought Crisis in Somalia, Ethiopia, and Kenya: IASC.

Smith, C., Kniveton, D., Black, R., & Wood, S. Climate Change, Migration and Agent-Based Modelling: Modelling the impact of climate change on forced migration in Burkina Faso.

Vani, Volite & Mandla, V.R.. (2017). Comparative study of NDVI and SAVI vegetation indices in Anantapur district semi-arid areas. International Journal of Civil Engineering and Technology. 8. 559-566.

Wong, D. W. (1998). Measuring multiethnic spatial segregation. *Urban Geography*, *19*(1), 77-87.

Xiao, N., Cai, S., Moritz, M., Garabed, R., & Pomeroy, L. W. (2015). Spatial and temporal characteristics of pastoral mobility in the far north region, Cameroon: data analysis and modeling. *PLoS One*, *10*(7), e0131697.

Appendix A: ODD+D Protocol

The ODD+D protocol documents information used during the modeling process. The ODD protocol (Overview, Design Concepts, and Details) was developed to standardize the documentation of agent-based modeling (Grimm et al, 2010). Müller et al. (2013) suggested the amendment of a section that outlines human decision making, which is relevant to the ABM described in this paper. Therefore, the ODD+D protocol was applied to this study.

1. Overview

1.1. Purpose

1.1.1. What is the purpose of this study?

The purpose of this study is to understand the impact of seasonal, environmental changes and conflict on the movement of nomadic pastoralists in Somaliland between 2008 and 2018.

1.1.2. For whom is the model designed?

This model is designed for humanitarian researchers studying the interaction between conflict, environment, and migration and those studying human migration more generally. This case study also adds to a growing body of literature addressing the value of computer models in the humanitarian sector.

1.2. Entities, state variables, and scales

1.2.1. What kind of entities are in the model?

Each modeled agent in the simulation represents a nomadic pastoralist household. The agent remains in a constant state unless the impact of exogenous variables is so severe that the agent state changes from nomadic pastoralist to internally displaced person (IDP).

The second entity in the model is the spatial environment. The environment is not only a modeled entity but also drives the behavior and movement of the agents.

1.2.2. By what attributes (i.e. state variables and parameters) are these entities characterized?

The nomadic pastoralists' clan and ethnicity are identified according to the region in which the agent is generated. The spatial coordinates of the agent's position are recorded at each timestep in addition to the distance traveled to that location. Lastly, the grid cell value corresponding to the agent's position is also recorded as an attribute. If the agent state changes from nomadic pastoralist to IDP, the agent effectively exits the simulation and no additional attributes are collected.

The entity representing the spatial environment is composed of aggregated variables in grid cells measuring 1 km^2. Each grid cell is characterized by a normalized value between 0 and 1. The pixel value reflects the state of the environment as characterized by the following variables: vegetation cover, water availability, terrain gradient, conflict events, and the presence of ethnic boundaries.

1.2.3. What are the exogenous factors/drivers of the model?

The drivers of the model are largely captured in the spatial environment as described above. The presence of water and vegetation are considered pull factors whereas conflict, proximity to ethnic boundaries, and steep gradient are incorporated as push factors in the model. Once the agent has identified the most favorable grid cell, it must determine whether they are positions on private or public land. If it is public, they can continue their actions, however if the land is private, the agent must establish a deal with the landowner. If the deal is made the agent can remain on that parcel for the duration of a season, if they cannot make a deal then they must continue to search for alternate land. If they are unable to do so within a given season, the agent state changes from pastoralist to IDP and they exit the simulation.

1.2.4. <u>If applicable, how is space included in the model?</u>

The modeled environment is a spatial grid, where each grid cell measures 1 km², collectively covering the full extent of Somaliland. The spatial environment is artificially created through the combination of GIS datasets and information derived from satellite imagery. The individual data layers are incorporated into the spatial environment through the following equation:

$$Pixel\ value = v1 + v2 + v3 - (0.25 * v4) - v5 - (0.25 * v6)$$

Where,

v1 =normalized vegetative cover, as calculated with SAVI, where L=0.5

v2 = normalized surface water index (enabled in wet seasons)

v3 = artificial water point sources (enabled in dry seasons)

v4 = normalized terrain gradient

v5 = normalized conflict frequency

v6 = ethnic boundary

The additive pixel value is therefore representative of the condition of the artificial environment, which is considered the primary driver of pastoral migration.

1.2.5. <u>What are the temporal and spatial resolutions and extents of the model?</u>

The variables are aggregated into a single environmental layer which has a resolution of 1 km² and whose full extent measures 490,000 km². The simulation considers a monthly time step and runs for 11 years between 2008 and 2018.

2. Design Concepts

2.1. Theoretical and empirical background

2.1.1. <u>What entity does what, and in what order?</u>

Every month, the agent considers a traversable distance in which it searches for a cell with the highest value relative to the cells in the surrounding environment. The cell value reflects the aggregation of environmental variables is composed of favorable variables (high water and vegetation availability) and unfavorable variables (high prevalence of conflict, proximity to ethnic boundaries, or steep terrain. Once the highest grid cell has been identified, the agent moves to this location and is able to stay there for the duration of the season (1 - 3 months).

Once an agent moves to a new grid cell, the agent determines whether they are located on private or public land. If the agent is on public land, they are free to remain at that location without restraints for the duration of a season. If the agent happens to be on private land, the agent must negotiate a land-sharing deal with the local landowner. If the agent successfully strikes a deal with the landowner, they are able to remain at that location for the duration of the season. However, if the agent is unable to obtain access, they are required to move to another grid cell within the same time tick and again consider whether the new grid location is private or public land and repeat the process described above. If the agent moves to grid cells in private land and is also unable to make a deal on three separate occasions in any given season, the agent state switches from pastoralist to IDP, at which point the agent exists the simulation.

2.1.2. <u>Which general concepts, theories or hypotheses are underlying the model's design at the system level or at the level(s) of the submodel(s) (apart from the decision model)? What is the link to complexity and the purpose of the model?</u>

Nomadic pastoralists primarily rely on livestock to support their livelihoods. Pastoralism is therefore deeply intertwined with environmental conditions and is heavily dependent on the availability of water and pasture. While nomadic pastoralists typically traverse familiar environments, unpredictable changes in the landscape have required many to travel longer distances to unfamiliar locations in order to find water and pasture. The changing patterns of movement have at times resulted in ethnic clashes and/or conflict over critical resources. It is therefore the purpose of the model to better understand how conflict and changing environmental variables impact pastoral migration. Through modeling these complex relationships, we hope to develop a better understanding of how agent-based modeling may be used in understanding complex humanitarian crises, and to identify what the gaps and limitations of this methodology are. Of course, this model does not account for all the complexities experienced in reality.

2.1.3. <u>On what assumptions is/are the agents' decision model(s) based?</u>

To our knowledge, there is very limited publicly available data about nomadic pastoralists in Somaliland. The majority of the data that does exist is often outdated. For this reason, this model heavily relies on informed assumptions.

The researchers of this project relied on academic publications, grey literature, and local knowledge to inform agent decision making. The following assumptions were made in this model:

1) SAVI is the indicator used to indicate the presence of vegetation, which is an assumed pull factor for pastoralists
2) Artificial water points and surface water are both used as a pull factor
3) We assumed that the presence of conflict, proximity to ethnic boundaries, and slope in part influence pastoral migration. The likelihood of pastoralists striking a deal with local landowners is assumed, as is the condition for dropping out of pastoralism.
4) Distance scouted and traveled are derived from delphi consensus after conversations with local experts who interact with Somali pastoralist

populations. The thresholds utilized in the model are assumptions derived from these consensuses.

2.1.4. Why is/are certain decision model(s) chosen?

The agent decision model is largely based on heuristics.

2.1.5. If the model/submodel (e.g. the decision model) is based on empirical data, where does the data come from?

Agent decisions do not come from empirical data and primarily rely on heuristics. Data that comprise the environment and come from empirical sources are cited in the Methods section of the article, Table 1.

2.1.6. At which level of aggregation were the data available?

The data used to inform agent generation was disaggregated by administrative regions. More granular data at smaller administrative units was largely unavailable.

The variables included in the spatial environment were available at several different spatial resolutions. For instance, the global surface water index was applied to data from Sentinel 2, which captures data at 10 m resolution. The vegetation index was calculated using data from MODIS Terra at 250 m resolution. However, to create one continuous spatial layer, the data layers were aggregated to 1 km resolution. The lower resolution was thought to adequately capture movement on a regional scale and also minimized the time and effort required by the machines running the model.

2.1.7. What are the subjects and objects of the decision-making? On which level of aggregation is decision-making modelled? Are multiple levels of decision making included?

Decision-making is modeled on a household level that can then be aggregated to an administrative level.

2.1.8. <u>What is the basic rationality behind agent decision-making in the model? Do agents pursue an explicit objective or have other success criteria? How do agents make their decisions?</u>

The basic rationale behind the agent decision-making process is that the agent must identify a grid cell that is suitable for grazing, given a combination of factors. This is done by seeking out the grid cell with the highest additive score in a random proximity to the agent's location. Once the grid cell has been successfully identified, the agent must confirm that they are able to graze here, since the land may be privately owned. If the agent is consistently able to occupy a grid cell without any problems, they are successful in their mission.

2.1.9. <u>Do the agents adapt their behavior to changing endogenous and exogenous state variables? And if yes, how?</u>

The agent's decision-making process remains constant, their pattern of movement changes in response to the spatial environment.

2.1.10. <u>Do social norms or cultural values play a role in the decision-making process?</u>

Social norms do play a role in the decision-making process. Ethnic conflict is not uncommon in Somaliland, which is why ethnic boundaries are considered in this model. The assumption is therefore, the closer the agent moves to an ethnic boundary, the higher the likelihood for conflict to occur. Additionally, pastoralists are known to strike deals with landowners to gain access to pasture and water as an adaptation technique.

2.1.11. <u>Do spatial aspects play a role in the decision process?</u>

Agent decision-making processes and movements are entirely informed by variables that are spatially represented. Distance, terrain gradient, and the general land cover conditions are all considered prior to agent movement. Agents move between 15 - 30 kilometers per timestep, which was determined through discussions captured at a workshop in Nairobi, Kenya on June 3 - 4, 2019.

2.1.12. Do temporal aspects play a role in the decision process?

While the agent is able to move at every time tick, the agent is also able to remain at any given grid cell for the duration of a full season, after which they will be required to move. This encourages the agent to follow typical seasonal movement patterns.

2.1.13. To which extent and how is uncertainty included in the agents' decision rules?

Uncertainty is included in the agents' decision rules through the incorporation of probabilities. Additionally, all the probabilities and even the equation of the additive model can be modified to understand the agents' behaviors under varied circumstances.

2.2. Learning

2.2.1. Is individual learning included in the decision process? How do individuals change their decision rules over time as a consequence of their experience?

Learning is not included in the decision process.

2.2.2. Is collective learning implemented in the model?

No

2.3. Individual sensing

2.3.1. What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions? Is the sensing process erroneous?

Individual agents are assumed to sense the surrounding spatial environment (exogenous variable) prior to moving in the simulation.

2.3.2. What state variables of which other individuals can an individual perceive? Is the sensing process erroneous?

None.

2.3.3. What is the spatial scale of sensing?

The spatial extent of the sensing is determined through a randomized distance between 15 - 30 km.

2.3.4. Are the mechanisms by which agents obtain information modelled explicitly, or are individuals simply assumed to know these variables?

The individuals are assumed to know the variables they are sensing.

2.3.5. Are the costs for cognition and the costs for gathering information explicitly included in the model?

No

2.4. Individual prediction

2.4.1. Which data do the agents use to predict future conditions?

The agent does not predict future conditions.

2.4.2. What internal models are agents assumed to use to estimate future conditions or consequences of their decisions?

N/A

2.4.3. Might agents be erroneous in the prediction process, and how is it implemented?

N/A

2.5. Interaction

2.5.1. Are interactions among agents and entities assumed as direct or indirect?

In this first model iteration, there is no agent interaction.

2.6. Collectives

2.6.1. Do the individuals form or belong to aggregation that affect and are affected by the individuals? Are these aggregations imposed by the modeler or do they emerge during the simulation?

A single agent is considered a singular household unit. The agents do not aggregate during the simulation and are affected individually.

2.6.2. How are collectives represented?

N/A

2.7. Heterogeneity

2.7.1. Are the agents heterogeneous? If yes, which state variables and/or processes differ between the agents?

The agents are minimally heterogeneous. The agent state variables include geographic position, clan, and ethnic affiliation. Due to limited, outdated, and often unreliable information, no further agent characteristics were included in an effort to minimize assumptions and uncertainty.

2.7.2. Are the agents heterogeneous in their decision-making? If yes, which decision models or decision objects differ between the agents?

Agents are homogeneous in their decision-making.

2.8. Stochasticity

2.8.1. What process (including initialization) are modelled by assuming they are random or partly random?

The agent start locations are randomly generated with two constraining rules as described in section 2.9.1

2.9. Observation

2.9.1. What data are collected from the ABM for testing, understanding and analyzing it, and how and when are they collected?

At every time tick, the following data are collected: 1) geographic position of the agent, 2) the score associated with the pixel within which the agent is situated, and 3) the clan and ethnic affiliation of the agent. The spatial analysis of these attributes will provide insight about where and how far agents move and what the relationship is between agent movement and individual variables.

2.9.2. <u>What key results, outputs or characteristics of the model are emerging from the individuals? (Emergence)</u>

We anticipate that varying spatial and temporal patterns of migrations will emerge from the model in response to changing environmental and conflict factors.

3. Details

3.1. Implementation details

3.1.1. <u>How has the model been implemented?</u>

The model was implemented using Java in Repast.

3.1.2. <u>Is the model accessible, and if so where?</u>

The source code is stored on GitHub.

3.2. Initialization

3.2.1. <u>What is the initial state of the model world, i.e. at time t=0 of a simulation run?</u>

The state of the model at t = 0 mimics the environment as it was in January 2008. The agents generated at t = 0 are generated randomly throughout Somaliland with two conditions: 1) Agents cannot be in areas that are labeled as water bodies or sand, and 2) agents cannot be within 14 km and 4 km radius from a major city or settlement, respectively.

3.2.2. <u>Is the initialization always the same or is it allowed to vary among simulations?</u>

The simulation initialization is always the same.

3.2.3. <u>Are the initial values chosen arbitrarily or based on data?</u>

The initial values of the simulation environment are based on aggregated GIS and remotely sensed data that were captured at the time of the simulation date.

The number of agents generated per administrative units are informed by data captured by UNPF in 2013.

3.3. Input data

3.3.1. Does the model use input from external sources such as data files or other models to represent processes that change over time?

The spatial environment changes on a seasonal basis, these files were preprocessed by analysts. The input variables and their sources are listed below:

- Agent characteristics | Population distribution is derived from a UNFPA Population Survey conducted in 2014.
- Vegetation | SAVI as calculated on imagery collected by the MODIS Terra satellite
- Conflict | Armed Conflict Location & Event Data Project
- Artificial water sources | Somalia Water and Land Information Management
- Natural water sources | NDWI layer obtained from Google Earth Engine
- Slope | Calculated from DEM obtained from DIVA GIS
- Ethnic boundaries | Obtained from the Kenya Somalia Consortium
- Public/Private land delineation | Point data obtained from HDX, from UNOCHA