Localized Conflict Modeling for the DARPA World Modelers Project

THIS IS CURRENTLY UNDER CONSTRUCTION

Introduction

Motivation

Why do we want to create a conflict model?

Types of Conflict

There are many ways to categorize conflict. One important distinction is between intra-national conflict (subnational?) and international conflict (e.g. with border countries). Distance to the country's border may be an important indicator for the latter, but not for the former type. Another key distinction is whether a conflict is armed or unarmed, and the ACLED data set (ACLED, 2021a) focuses on armed conflict. The adjectives: state-based, non-state, and one-sided are also commonly used to categorize conflicts. Events such as protests (peaceful or non-peaceful) and riots may also be considered as types of conflict. Depending on which type of conflict one wishes to model and understand, different indicators may be more or less important.

Roots of Conflict

Funsten (2016) identifies the following three "Roots of Conflict":
(1) Limited resources, (2) Unmet needs and (3) Different values.
(Adapted from Shrumpf et al. (1991).)

Related Work

Here we briefly describe a few other conflict modeling projects and other types of models that have provided inspiration for the current model.

Kimetrica's Machine-Learning Conflict Model

Kimetrica has another type of conflict model based on machine-learning algorithms. That model is described here:
https://docs.kimetrica.com/world_models/domain_models/models_conflict_model_README.html
That model found that population count, mean rainfall amount, and land cover change index were the most important contributors to the onset of conflict. There are many differences between that model and the one proposed here. The proposed model is a simulation model that

associates different groups of indicators with different roles, such as potential for conflict, triggering of conflict and spreading of conflict. In that sense the current model is ***mechanistic***, or explainable, but it is also ***stochastic*** (i.e. it uses random variables and probability theory). In addition, it has a few adjustable control parameters that lead to different model behavior.

Violence Early Warning System (ViEWS) -- Uppsala University

See ViEWS (2021).

Conflict Pulse Model -- Armed Conflict Location and Event Data (ACLED)

See ACLED (2021b).

Extended Forest-Fire Model

See Drossel and Schwabl (1992) and Forest-fire Model (2021).

Infectious Disease Models

See Kim et al. (2019), Infectious Disease (2021), and Compartmental Models (2021)

Global Conflict Risk Index (GCRI) - European Commission

See Halkia et al (2019, 2020).

Modeling Methodology

Overview

We are in the process of converting a conceptual model into a mathematical/stochastic model. The conceptual model is loosely based on a cellular automaton model known as the "Forest-fire model". See Drossel and Schwabl (1992) and Forest-fire Model (2021). The conceptual model consists of 4 key components.

1. Unrest - The Potential for Conflict to Emerge. For each geospatial grid cell, we intend to compute a variable we will call ***unrest***, which is intended to quantify the potential for conflict within that grid cell. This will be a composite variable or index that is a function of several ***indicator*** variables. Here, we define an ***indicator*** to be any variable that may help to quantify a *qualitative property* of an object or system, but that by itself provides only an indirect or insufficient measure of that property. A familiar example of a qualitative property of a person is creditworthiness. The FICO score or index is a function of 6 different indicators that is used to quantify this somewhat abstract property. Another familiar example is obesity, and the Body Mass Index is computed as a function of two indicator variables (i.e. indicators for obesity), weight and height. Note that these are both ***quantitative properties*** (or quantities) associated with a person. When used to quantify obesity they play the role of an indicator, but otherwise they are just variables.

There are many indicator variables that can be used (and have been used) in an effort to quantify the potential for conflict and to predict the emergence of conflict. These can be roughly grouped into the following categories.

(1) **climatic conditions** (e.g. rainfall amount, drought, flooding, heat-wave, sea-level rise)
(2) **economic conditions** (e.g. food shortage, food prices, unemployment, income disparity, resource richness, inflation rate, GDP/capita, crop failure),
(3) **social conditions** (e.g. population count, population density, population growth, quality of life, standard of living, infant mortality rate, "identity group" mixture, education, access to health care, change in ethnic mix, maybe due to migration, land cover change, deforestation, "youth bulge", disasters, urbanization)
(4) **geographic conditions** (e.g. distance to border, distance to capital, remoteness)
(5) **political conditions** (e.g. time to next election, polarization, regime change, assasination of a leader, ethnic cleansing)
(6) **historical conditions** (e.g. time since the last conflict)

Note: We do not expect for all of these variables to be equally important in determining the degree of unrest. Previous modeling work in Ethiopia found that the amount of rainfall, population density and land cover change were among the most important indicators. Note that these 3 indicators may be less dominant in other countries.

Note: In the forest-fire analogy, unrest might correspond to the degree of "flammability", "combustibility", "dryness" or "fuel availability" (e.g. biomass). In the forest fire model, however, "flammability" is a boolean property -- it is 1 if there is a tree in the grid cell and 0 if there isn't.

2. Connectivity - The Potential for Conflict to Spread. For each geospatial grid cell, we also intend to compute a measure of connectivity that will be used to quantify the potential for a conflict in one cell to spread to other grid cells. This will also be a composite variable or index that is a function of several indicator variables. The indicator variables will characterize two types of connectivity:

(1) **Local / Geographic Connectivity** (geospatial or physical; e.g. transportation network accessibility, road density, topographic roughness, travel times)

(2) **Nonlocal / Electronic Connectivity** (access to nonlocal information; e.g. access to devices: cell phone, TV, radio, internet/broadband. Also social media usage, newspapers.

The widespread use of cell phones and social media has led to a dramatic increase in "electronic connectivity". It is often viewed as the key factor in the "Arab Spring" phenomenon. See Arab Spring (2021). Note that electronic connectivity may have either a positive or negative impact on the spreading (e.g. when the family of someone who was killed calls for peaceful demonstrations, conflict may be reduced).

Note: It appears that the ***communication*** between people in different grid cells (near or far) is the primary mechanism by which conflict spreads. So connectivity is really getting at the ability to communicate between cells.

Note: This differs from the standard forest-fire model. In that model, only local (nearest neighbor) connectivity is present.

3. Trigger Events - The Initiation of Conflict. For each grid cell, there is a probability that some event will occur that initiates or sparks a conflict. (And that initial conflict may escalate or de-escalate, depending largely on how it is handled and the power discrepancy between the parties.) Conflicts are initiated as a result of negative interactions between people. This simple fact helps to explain the strong correlation between the occurrence of conflicts and population count or density. Let p(i) denote the probability that a conflict is initiated in a grid cell with index i. We then expect that p(i) will be an increasing function of unrest(i), the degree of unrest in the ith grid cell.

Note: This also differs from the forest-fire model of Drossel and Schwabl (1992). In that model, every grid cell has the same probability of a trigger event such as a lightning strike. (So this can be modeled with a 2D Poisson point process.) In this model, we expect this probability to vary spatially (and also in time) and to depend on the degree of unrest. (Of course, unquenched campfires and discarded cigarettes can also trigger a forest fire, but these are not naturally-occurring events. Since they are caused by people, the probability of these trigger events would increase with the population count of a grid cell. This is reminiscent of how the spatial distribution of meteorite findings are strongly correlated with population density.)

4. Conflict Duration - The Persistence of Conflict. Conflicts generally have a finite duration. We could model this in various ways. One of the simplest is to treat the duration of each conflict event as an independent, random variable drawn from some distribution that takes only nonnegative values. This could be an exponential distribution (Exponential Distribution, 2021) with parameter, lambda, for example. For a discrete time version of the model, an integer-valued random variable from a geometric distribution (Geometric Distribution, 2021) could be used to determine duration as a number of discrete time intervals. Using a Bernoulli random variable (parameter p) to determine whether conflict ends in a given time interval will result in conflict durations having a Geometric distribution with the same parameter, p. Lee et al. (2020b) discuss conflict duration in terms of “virulence”.

The Proposed Model (Prototype)

Here we describe the model, as currently envisioned, but this is still a work in progress. Referring to the last section, let U denote a geospatial grid of values that quantify unrest.
Let C1 denote a geospatial grid of values that quantify the average degree to which people in a given grid cell ***communicate with*** people in nearest neighbor (adjacent) grid cells. Let C2 denote a geospatial grid of values that quantify the average degree to which people in a grid cell communicate with (i.e. send or receive information to/from) people in any other grid cell. **Note:**

C2 depends on the degree to which a given cell is connected electronically to the rest of the country (or region or world). In general, U, C1, and C2 values would be expected to vary both spatially and in time. (These may be estimated by a machine learning algorithm or some other algorithm as functions of selected indicator variables.)

Note: While TVs, radios and newspapers allow people to ***receive*** information, they are much less likely to be used by people to ***send*** information to others. This is in sharp contrast with telephones (cell or land-line), and internet access which allow two-way communication. In order for information to spread from one cell to another, it has to first be received by some means and then sent or forwarded via two-way communication. Perhaps we should therefore distinguish between the degree to which people in a grid cell ***receive*** information electronically (by TV, radio, newspaper, phone, internet), and the degree to which they ***send*** it electronically (only by phone & internet). The time it takes for a person to act on information received (e.g. forward it, organize a protest, write about it) should probably also be taken into account.

Note: We could instead define C1 and C2 as follows. Let B1 and B2 denote the average degree to which ***any one person*** can communicate with (1) people in a nearest neighbor grid cell, or (2) people in any other grid cell in the country, respectively. We could then compute C1 and C2 by multiplying B1 and B2 by the population count of the grid cell. This would underscore the fact that -- especially for 2-way communication -- the total amount of communication is proportional to population size, and not simply related to the percentage of the population that can communicate electronically.

Conflict Emergence

Let E(i,j,k) denote the event that a conflict emerges during the kth discrete time interval, in the grid cell that has column and row indices i and j. Let p(i,j,k) be the probability of the event E(i,j,k):

$$p(i,j,k) = p\_emerge(i,j,k) = Prob[\, E(i, j, k)]$$

We assume that p(i,j,k) is an increasing function of U(i,j,k). Initially, we will simply consider the special case where p(i,j,k) is proportional to U(i,j,k), so that:

$$\begin{array}{ll} \text{if } (S(i,j,k) = 0), & p(i,j,k) = c\_emerge * U(i,j,k) / max\_i,j\ [\ U(i,j,k)\ ] \\ \text{if } (S(i,j,k) = 1), & p(i,j,k) = 0, \end{array}$$

where c_emerge is in (0,1]. Here, we have normalized the grid U(i,j,k) and then multiplied the result by a constant, c. Dividing by the array max and multiplying by a model parameter c in (0,1] guarantees that 0 <= p(i,j,k) < 1 for every grid cell --- as it must since it is a probability. Note that the parameter c can be adjusted to make conflicts more or less likely to occur. In the special case where all values in the U grid are equal, all cells will have the same probability of conflict emergence: p_emerge = c_emerge. (Note: Instead of dividing by the array max, we could have divided by the array sum, but then the upper bound on c is more complicated.)

Let S be a geospatial grid with the same dimensions as U, C1 and C2, such that the value of a grid cell in S is 1 if the cell is experiencing (or engaged in) a conflict, and 0 otherwise. For each model timestep, and for each grid cell, generate an ***independent***, random variable from a Bernoulli distribution (see Bernoulli Distribution, 2021) such that:

$$S(i,j,k) = 1 \text{ with probability } p(i,j,k) \text{ and}$$
$$S(i,j,k) = 0 \text{ with probability } (1 - p(i,j,k)).$$

Recall that the E(S(i,j,k) = p(i,j,k) for a Bernoulli random variable. Also recall that the expected value of a sum of random variables is equal to the sum of their expected values, *even if they are not independent*. Therefore, the expected number of grid cells in which a conflict is ***initiated*** in the 1st time interval (k=1, before any spreading) is given by:

E(Sum_i,j [S(i,j,1)] = (total conflicts in 1st interval) = Sum_i,j [p(i,j,1)]

The expected number of grid cells in which a conflict is initiated in the kth time interval (k >1, before spreading), will be less than this since cells already experiencing conflict (i.e. with S=1) are excluded (i.e. for them, p(i,j,k) = 0).

Conflict Spreading

For every grid cell that has S(i,j,k) = 1, the next step is to determine if conflict spreads to other grid cells (which could be nearest neighbors or far away due to electronic connectivity). If it does, the spreading process is repeated for those other grid cells until there is no further spreading. (That is, no cells have their S value changed to 1.) Since there is a lag time between cell-to-cell communication and a resulting action (to spread conflict into a new cell), we introduce a time lag parameter, tau, which can be taken to be a multiple of the model timestep. (In the current version of the model, tau = 1.)

There are 4 special cases of spreading that can be considered:
Case 1: no spreading,
Case 2: spreading only by local connectivity,
Case 3: spreading by only nonlocal connectivity, and
Case 4: spreading by both local and nonlocal connectivity.
For the last three of these cases, various spreading rules are also possible.

Case 1. No spreading. While there may be conflict in many cells, there is no spreading of conflict to other grid cells.

Case 2. Spreading only by local connectivity. In this case, conflict can only spread from a grid cell with conflict to other grid cells that have sufficiently high C1 and U values. Cells with a high C1 value should include nearest neighbor grid cells and possibly other cells, such as those separated by a very short travel time. In the current version of the model, the probability of spreading to any ***nearest neighbor*** is weighted by the U value of that neighbor, i.e.

$$p(n) = c\_spread * U(n) / max\_n(U(n)),$$

where c_spread is in (0,1] and n is a nearest neighbor index. If U is zero for all neighbors, then p=0. Notice that conflict cannot spread to a neighbor that has U=0, and this should include grid cells in a lake or in the ocean, for example. In the current version, conflict cannot "spread to" or "re-emerge" in a cell that already has conflict, but this could be changed. Various other spreading rules are also possible, such as:

(a) Conflict will spread to any neighbor, n, such that U(n) > U0, where U0 is a threshold value.
(b) Conflict will spread to any neighbor, n, such that U(n) >= a * U(i), with a in (0,1].
(c) The probability of spreading to any neighbor is weighted by the U and C1 values of that neighbor, for example using:

$$p(n) = c\_spread * [U(n) * C1(n)] / max\_n [U(n) * C1(n)],$$

where c_spread is in (0,1]. Here, the probability of spreading to a neighbor is low if either U(n) or C1(n) is low.

Case 3. Spreading only by nonlocal connectivity. In this case conflict can spread from a grid cell with conflict to any other grid cells that have sufficiently high C2 and U values.

Case 4. Spreading by both local and nonlocal connectivity. In this case, conflict can spread from a cell with conflict to:

(1a) any of its nearest neighbor grid cells if its own C1 value is sufficiently high, and if their value of U is sufficiently high, OR
(1b) any of its nearest neighbor grid cells with sufficiently high C1 and U values.

(2) any other grid cell with sufficiently high C2 and U values.

Conflict Duration

For every grid cell that had its S value set to 1 in the kth time interval (by emergence/triggering or spreading), generate an independent random variable from a Bernoulli distribution (Bernoulli Distribution, 2021) with parameter, p_resolve. If B=1, the conflict is ***resolved*** in time interval k and if B=0 the conflict continues. If we take X to be the random number of time intervals before the conflict is resolved, which is the duration of the conflict, then X will have a Geometric distribution (Geometric Distribution, 2021) with parameter p_resolve. The expected value of this random duration is then equal to 1/p_resolve. (So a smaller value of p_resolve results in a larger mean duration.) As long as conflict is active within a cell (some number of timesteps) it will attempt to spread to other cells.

Expected Results

Depending on the values of the model parameters U, C1, C2, c_emerge, c_spread and p_resolve, this model can exhibit a wide range of behaviors, from rare conflicts that remain isolated to conflicts that spread to fill an entire region. Preliminary mathematical analysis of this model as well as early simulations suggests that ***if the model parameters are assumed not to vary with time***, then this model may evolve toward a “statistical steady-state” condition where:

(1) the mean rate at which new conflicts appear (by either emergence or spreading) eventually becomes equal to the mean rate at which they are resolved, and therefore

(2) the mean number of grid cells that are experiencing conflict in the kth time interval approaches a constant.

For details, see Appendix 1: Preliminary Mathematical Results.

Data Sources

Population Data

Global Data Lab, GDL Area Database 4.0,
https://globaldatalab.org/areadata/indicators/, (e.g. regpopm)
Global Rural-Urban Mapping Project (GRUMP) - v1,
https://sedac.ciesin.columbia.edu/data/collection/grump-v1/
Gridded Population of the World (GPW), v4, SEDAC,
https://sedac.ciesin.columbia.edu/data/collection/gpw-v4 (30 arcsec ~ 1 km)
LandScan, Oak Ridge National Laboratory, https://landscan.ornl.gov/
POPGRID Data Collaborative, https://www.popgrid.org/
POPGRID Viewer, https://sedac.ciesin.columbia.edu/mapping/popgrid/, SEDAC
PRIO-GRID 2.0, https://grid.prio.org/#/, (30-min grid cells, other products)
WorldPop, https://www.worldpop.org/project/list, (100 m grid cells)
WorldPop Viewer, https://apps.worldpop.org/woprVision/

Note: A table that compares the attributes of various population data sets is given here:
https://www.popgrid.org/data-docs-table1

Conflict Data

ACLED Project Data, https://acleddata.com/
GDELT Project, https://www.gdeltproject.org/, (using Google BigQuery)
ICEWS (Integrated Conflict Early Warning System),
https://en.wikipedia.org/wiki/Integrated_Conflict_Early_Warning_System
https://dataverse.harvard.edu/dataverse/icews (data in Harvard Dataverse)
https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html
Justice Data, http://www.justice-data.com/ (incl. post and during conflict justice)
Uppsala Conflict Data Program (UCDP), https://ucdp.uu.se/encyclopedia
UCDP Data Download Center, https://ucdp.uu.se/downloads/

Other Indicator Data

Global Data Lab, GDL Area Database 4.0,
    https://globaldatalab.org/areadata/indicators/, (e.g. internet, cellphone)
    (Not gridded; spreadsheets down to subnational admin regions)
Ki-Data (Kimetrica Data Publishing) https://data.kimetrica.com/
KDW (Kimetrica Data Warehouse) https://kdw.kimetrica.com/en/

Note: Some of the data sets listed under Population Data also provide data for other indicators.

Data Concerns

ACLED. In the ACLED event spreadsheets, spanning 1997 to 2019, for Horn of Africa countries (Ethiopia, Eritrea, Djibouti, Somalia, as well as Kenyra, Sudan, & South Sudan), the number reported in the Fatalities column is often repeated over many rows. For example, the highest value is 1369 and occurs in 73 rows, but many other exact values like this are repeated in many rows. This same number can be associated with multiple dates, multiple regions and multiple sources. As a result, we need some way to avoid overcounting.

Statistical analysis / feature selection

Which of these variables seemed to be significant? Any other significant statistical takeaways? Put a correlation matrix here.

For models where the output variable of interest is binary (e.g. conflict or no conflict, S=0 or S=1) that depend on several input variables, various ***binary regression*** models can be used instead of ***linear regression.*** Two such models are ***logistic regression*** (logit model) and ***probit regression*** (probit model). See Binary Regression (2021), Logistic Regression (2021), and Nailufar et al. (2015).

Model selection and validity

We chose to build a machine learning model because? We split into test and training, blah blah blah

Coming soon...

Results

The goal of this model is to predict the likelihood that a conflict will emerge in a given grid cell and to simulate the dynamics of how it is likely to spread and grow.

Next Steps

A prototype, stochastic, mathematical conflict evolution model has been described in the previous sections. The next step will be to implement this as a spatial simulation model in Python with the Numpy package, and to visualize the model predictions using the Matplotlib and Cartopy packages. It will also be necessary to prepare the unrest and connectivity grids (U, C1 and C2) as functions of multiple conflict indicators. A simple population count grid may be used initially for U, for development purposes. Existing Kimetrica models that use machine learning algorithms may be used to prepare the U, C1 and C2 grids. After the model has been implemented and tested with different indicators and parameters (like c_emerge, c_spread, p_resolve, and possible thresholds), it will be adapted to obtain the necessary indicator data using the Luigi framework, similar to other Kimetrica models. At some point, one or more Jupyter notebooks will also be created to make it easier to learn about and use the model.

Appendix 1. Preliminary Mathematical Results

Case 1. No Spreading

Consider a grid cell at location (i,j). Let p(k) = p(i,j,k), which depends on U(k) = U(i,j,k), and let B1(k) be a sequence of independent Bernoulli random variables with parameter p(k). Let B2(k) be a sequence of independent Bernoulli random variables with parameter p_resolve. In order to have S(k+1)=0, there are only two (mutually exclusive) possibilities:

(1) [S(k) = 0] and [no conflict was triggered at time k+1], or
(2) [S(k) = 1] and [the conflict was resolved at time k+1].

Recall that for two ***independent*** random events X and Y, the probability of their intersection is given by: P[X and Y] = P[X] P[Y]. (A random event can be that some random variable takes a particular value.) Notice that B1(k+1) and B2(k+1) have no dependence on S(k). Note also that for two ***mutually exclusive*** events X and Y, P(X and Y) = 0. (e.g. S=0 and S=1). Also recall that the probability of the union of events X and Y is given by: P[X or Y] = P(X) + P(Y) - P(X and Y). (This one does not require independence.) The probability that S(k+1)=0 can therefore be computed as follows:

P[S(k+1) = 0] = P[{(S(k) = 0) and (B1(k+1) = 0)} OR {(S(k) = 1) and (B2(k+1) = 1}]
= P[S(k) = 0] P[B1(k+1) = 0] + P[S(k) = 1] P[B2(k+1) = 1]

P[S(k+1) = 1] = 1 - P[S(k+1) = 0]

Let R(k) = P[S(k) = 0], let a(k) = pe(k) = p_emerge(k) = p(i,j,k), and let b = pr = p_resolve. We then have the recurrence equation:

R(k+1) = R(k)*[1 - pe(k)] + [1 - R(k)]*pr

$$R(k+1) = pr + R(k) * [1 - pe(k) - pr]$$
$$R(k+1) = R(k) * a(k) + b.$$

where $b = pr$ and $a(k) = [1 - pe(k) - pr]$. Note that $R(0) = 1$ since $S(0)=0$ (i.e. S is initialized to 0). This is a ***recurrence equation*** for R(k). In order to solve it, we must know a(k), which requires knowing both pe(k) and U(k). However, in the special case where the unrest, U, is the same for every time interval, k --- so that $U(k) = U(0)$ --- we have $a(k) = a(0) = a = f(U(0))$ (i.e. a function of U(0)). In this special case (still with no spreading) we can solve the recurrence equation for R(k) (with Mathematica's RSolve function), to get

$$R(k) = \{\{a^k * (a + b - 1) - b \} / (a - 1)\}.$$

Since $a = (1 - pe - pr)$ and $b = pr$, $(a + b - 1) = -pe$, and $(a - 1) = -(pe + pr)$. After these substitutions we can simplify to get

$$P[S(k) = 0] = R(k) = [pr + pe *(1 - pe - pr)^k] / (pr + pe)$$

Notice that the term $(1 - pe - pr)$ will be negative if $(pe + pr) > 1$ (e.g. $pe = pr = 0.6$). While the equation is still valid in this case, $(1 - pe - pr)^k$ will be negative for odd values of k and positive otherwise; this will cause R(k) to oscillate rather than decreasing smoothly to its limiting value. In the limit as k goes to infinity, this sequence rapidly converges to the constant:

$$c = pr / (pe + pr) < 1.$$

If $pe = pr$, this constant equals 0.5. Keep in mind that pr is a constant (i.e. it does not depend on i, j or k). For this result we assumed that U(i,j,k) is the same for all k, but this means that pe can still be different for every grid cell, that is, $pe = pe(i,j)$. Note that if $pe \ll pr$, then c approaches 1 and the given cell will be very ***unlikely*** to have conflict. Similarly, if $pe \gg pr$, then c approaches 0 and the given cell is very ***likely*** to have conflict. If U does not depend on i, j or k, then pe won't either.

Case 1 as a Two-State Markov Process

For the grid cell at (i,j), $S(i,j,k) = S(k)$ is a sequence of zeros and ones (the two possible states -- conflict or no conflict).

$P[S(k+1) = 1 \mid S(k) = 0] = P[B1(k+1) = 1] = pe$ (new conflict emerges)
$P[S(k+1) = 0 \mid S(k) = 0] = P[B1(k+1) = 0] = (1 - pe)$ (current non-conflict continues)
$P[S(k+1) = 1 \mid S(k) = 1] = P[B2(k+1) = 0] = (1 - pr)$ (current conflict continues)
$P[S(k+1) = 0 \mid S(k) = 1] = P[B2(k+1) = 1] = pr$ (current conflict is resolved)

These probabilities give the transition matrix for a two-state Markov process:

M = [[(1 - pe), pe], = [[0 to 0, 0 to 1]
[pr, (1 - pr)]] [1 to 0, 1 to 1]]

The probability of going from any state to another state in k steps is given by M^k, the kth power of the transition matrix, M.

P[S(k+1) = 0] = P[S(k) = 0] P[B1(k+1) = 0] + P[S(k) = 1] P[B2(k+1) = 1]
P[S(k+1) = 1] = P[S(k) = 0] P[B1(k+1) = 1] + P[S(k) = 1] P[B2(k+1) = 0]

P[S(k+1) = a | S(k) = b] = P[{S(k+1) = a} and {S(k) = b}] / P[S(k) = b]

Case 2 as a Two-State Markov Process

For the grid cell at (i,j), S(i,j,k) = S(k) is a sequence of zeros and ones (the two possible states -- conflict or no conflict).

P[S(k+1) = 1 | S(k) = 0] = P[B1(k+1) = 1} or (spreading)]   > pe   (new conflict)
P[S(k+1) = 0 | S(k) = 0] = P[B1(k+1) = 0 and (no spreading)] < (1-pe) (cont'd peace)
P[S(k+1) = 1 | S(k) = 1] = P[B2(k+1) = 0] = (1 - pr)   (current conflict continues)
P[S(k+1) = 0 | S(k) = 1] = P[B2(k+1) = 1] = pr   (current conflict is resolved)

As before, let A(k+1) denote no spreading at time k+1.
P[A(k+1) and {B1(k+1) = 0}] = P[A(k+1) | B1(k+1) = 0] P[B1(k+1) = 0]
P[not(A(k+1)) or {B1(k+1)}=1] = P[not(A(k+1)] + P[B1(k+1)=1] - P[not(A(k+1)) and {B1[k+1]=1}]
= 1 - P[A(k+1)] + P[B1(k+1)=1] - 0??
= probability of new conflict

Let P[spreading] = P[not(A(k+1)] = ps.
P[S(k+1) = 1 | S(k) = 0] = ps + pe   (new conflict)
P[S(k+1) = 0 | S(k) = 0] = 1- (ps + pe)   (cont'd peace)
P[S(k+1) = 1 | S(k) = 1] = (1 - pr)   (current conflict continues)
P[S(k+1) = 0 | S(k) = 1] = pr   (current conflict is resolved)

M = [[1 - (ps+pe), (ps + pe)],
[pr, (1 - pr)]

But it looks like ps = ps(k) is an increasing function of k since neighbor cells become more likely to have conflict over time. If ps does not depend on k, we have:

Limit as k-> Infinity { M^k } = [[pr/(pe+pr+ps), (pe+ps)/(pe+pr+ps)],
pr/(pe+pr+ps), (pe+ps)/(pe+pr+ps)]

Maybe we can set up a recurrence equation for P[A(k)].
If there was conflict at time k, then there can be no spreading at time k+1 because conflict either continues or gets resolved.   If there wasn't conflict at time k, and it wasn't triggered….

$P[(\text{no spreading at time } k+1\ ] = P[$
$P[\text{not spreading}] = P[\ (\text{conflict at time } k) \text{ or } (\text{no conflict and}$
$P[\ (\text{conflict at time } k) \text{ and } (\text{wasn't resolved at time } k+1) \text{ OR}$

$P[A(k+1)] = P[S(k)=1 \text{ and } B2(k+1)=0] +$

Case 2. Case 1 Plus Spreading

Here we attempt to generalize the results of Case 1 to include spreading. Notice that the inclusion of spreading increases the probability that $S(k+1) = 1$, and therefore decreases the probability that $S(k+1) = 0$. So the results of Case 1 provide an upper bound on $P[S(k)=0]$. Now, in order to have $S(k+1)=0$, there are two possibilities:

(1) [$S(k) = 0$] and [no conflict was triggered at time k+1] and in addition,
[no conflict was spread from a neighbor at time k+1], or
(2) [$S(k) = 1$] and [the conflict was resolved at time k+1].

Let $A(k)$ be the event that "no conflict was spread from a neighbor at time k". We then have

$$
\begin{aligned}
P[S(k+1) = 0] = P[\ &\{(S(k) = 0) \text{ and } (B1(k+1) = 0) \text{ and } A(k+1)\} \quad \text{OR} \\
&\{(S(k) = 1) \text{ and } (B2(k+1) = 1\}\ ] \\
= P[&(S(k) = 0) \text{ and } (B1(k+1) = 0) \text{ and } A(k+1)] \ + \ P[S(k) = 1]\, P[B2(k+1) = 1]
\end{aligned}
$$

While it is clear that $A(k+1)$ depends on values in the S grid (but not on $S(k)$ for the cell in question), it is not entirely clear whether $A(k)$ and $B1(k+1)$ are independent. In the model, spreading is not attempted if a cell has already been triggered (i.e. $B1 = 1$). But perhaps we can assume it is always attempted, even if the result doesn't make a difference when the cell has already been triggered. We may be able to use conditional probabilities here:

$$P[\ A(k+1) \mid B1(k+1) = 0)\ ] = ? = P[A(k+1) \text{ and } \{B1(k+1) = 0)] \,/\, P[\ B1(k+1) = 0)]$$
$$P[\ A(k+1) \mid B1(k+1) = 1)\ ] = 1 = P[A(k+1) \text{ and } \{B1(k+1) = 1)] \,/\, P[\ B1(k+1) = 1)$$

If $A(k)$ and $B1(k+1)$ are independent, the right-hand side in both cases must reduce to $P[A(k+1)]$. The formula and notation $P(X \mid Y) = P(\ X \text{ and } Y) / P(Y)$ is used to compute the ***conditional probability*** that event X occurs, ***given*** that the event Y has occurred. Notice that if X and Y are independent, then $P(X \text{ and } Y) = P(X)\, P(Y)$ and then $P(X \mid Y) = P(X)$.

If we assume that $A(k+1)$ and $B1(k+1)$ are independent, we then have

$$P[S(k+1) = 0] = P[(S(k) = 0) \text{ and } A(k+1)]\, P[B1(k+1) = 0] \ + \ P[S(k) = 1]\, P[B2(k+1) = 1]$$

$$P[(S(k) = 0) \text{ and } A(k+1)] \ = P[\ A(k+1) \mid (\ S(k) = 0\ )]\ P[\ (\ S(k) = 0\ )\ ]$$

Likely incorrect beyond this point, due to lack of independence.

If we assume that A(k+1) and S(k) are independent, then the probability that S(k+1)=0 can be computed as follows:

$$
\begin{aligned}
P[S(k+1) = 0] &= P[\ \{(S(k) = 0) \text{ and } (B1(k+1) = 0) \text{ and } A(k+1)\ \} \text{ or } \\
&\qquad \{(S(k) = 1) \text{ and } (B2(k+1) = 1\}\ ] \\
&= P[S(k) = 0]\ P[B1(k+1) = 0]\ P[\ A(k+1)\ ] \\
&\quad + P[S(k) = 1]\ P[B2(k+1) = 1]
\end{aligned}
$$

$$P[S(k+1) = 1] = 1 - P[S(k+1) = 0]$$

Again, for the given grid cell (i.e. fixed i,j), let R(k) = P[S(k) = 0], let a(k) = p_e(k) = p_emerge(k) = p(i,j,k), let b = p_r = p_resolve, and let q(k) = P[A(k+1)]. (Notice that q(k)=1 returns us to the case of no spreading.) We then have the recurrence equation:

$$
\begin{aligned}
R(k+1) &= R(k)*[\ 1 - p\_e(k)]*q(k) + [1 - R(k)]*p\_r \\
R(k+1) &= p\_r + R(k) * [\ q(k)*(1 - p\_e(k)) - p\_r\ ] \\
R(k+1) &= R(k) * a(k) + b.
\end{aligned}
$$

where b = p_r and a(k) = q(k)*[1 - p_e(k)] - p_r. As before, R(0) = 1. Now, in the special case where the unrest, U, is the same for every time interval, k --- so that U(k) = U(0) --- we have a(k) = a(0) = a = f(U(0)) (i.e. a function of U(0)), and q(k) = q(0) = q. We got the same recurrence equation in Case 1 and solved it to get

$$R\_k = \{\{a\text{^}k * (a + b - 1) - b\ \} / (a - 1)\}.$$

But now, since a = q*(1 - pe) - pr and b = pr, this simplifies to

$$R(k) = \{pr + [1 - q(1 - pe)] * [q(1 - pe) - pr]\text{^}\{k\}\ \} \ / [1 + pr - q(1 - pe)].$$

In the limit as k goes to infinity, this sequence rapidly converges to the constant:

$$c = pr / [1 + pr + q*(pe -1\ )$$

Work in progress:

The expected total number of grid cells in which conflict **emerges** in the kth time interval is given by:

The expected total number of grid cells to which conflict **spreads** in the kth time interval is given by:

The expected total number of grid cells in which a conflict **terminates** in the kth time interval is given by:

References

ACLED (2021a) Curated Data, https://acleddata.com/curated-data-files/, Accessed May 25, 2021

ACLED (2021b) Conflict Pulse Model, Armed Conflict Location and Event Data (ACLED) project, https://acleddata.com/conflict-pulse/.

Arab Spring (2021) Wikipedia, https://en.wikipedia.org/wiki/Arab_Spring, Accessed: May 12, 2021

Asynchronous Cellular Automaton (2021), Wikipedia, (see random independent update scheme) https://en.wikipedia.org/wiki/Asynchronous_cellular_automaton

Bernoulli Distribution (2021), Wikipedia, https://en.wikipedia.org/wiki/Bernoulli_distribution, Accessed: May 12, 2021

Binary Regression (2021), Wikipedia, https://en.wikipedia.org/wiki/Binary_regression, Accessed: May 31, 2021

Binomial Distribution (2021), Wikipedia, https://en.wikipedia.org/wiki/Binomial_distribution, Accessed: May 12, 2021

Boschee, E., J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz (2018) ICEWS Weekly Event Data, https://doi.org/10.7910/DVN/QI2T9A, Harvard Dataverse, V278

Boschee, E., J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, M. Ward (2015) ICEWS Coded Event Data, https://doi.org/10.7910/DVN/28075, Harvard Dataverse, V30, UNF:6:NOSHB7wyt0SQ8sMg7+w38w== [fileUNF]

Boschee, E., J. Lautenschlager, S. Shellman, A. Shilliday (2015) ICEWS Dictionaries, https://doi.org/10.7910/DVN/28118, Harvard Dataverse, V4, UNF:6:WBLzLxVC+rRZzmsx/wwtHA== [fileUNF]

Cellular Automaton (2021), Wikipedia, https://en.wikipedia.org/wiki/Cellular_automaton, Accessed: May 13, 2021

Clifford, P. and A. Sudbury (1973) A model for spatial conflict, Biometrika, Vol. 60, No. 3, pp. 581-588, Oxford University Press, https://www.jstor.org/stable/2335008

Combustibility and Flammability (2021) Wikipedia, https://en.wikipedia.org/wiki/Combustibility_and_flammability, Accessed: May 13, 2021

Compartmental Models (2021) Compartmental models in epidemiology, Wikipedia, https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology, Accessed: May 14, 2021.

Drossel, B. and F. Schwabl (1992) Self-organized critical forest-fire model, Physical Review Letters, Vol. 69, No. 11.

Exponential Distribution (2021) Wikipedia, https://en.wikipedia.org/wiki/Exponential_distribution, Accessed: May 14, 2021

Forest-fire Model (2021), Wikipedia, https://en.wikipedia.org/wiki/Forest-fire_model, Accessed: May 12, 2021

Funsten, R. (2016) How understanding conflict can help improve our lives, TEDxTryon, YouTube, https://www.youtube.com/watch?v=fdDQSHyyUic, 11 min, 58 sec.

GDELT Project (2021) A global database of events, https://www.gdeltproject.org/

Geometric Distribution (2021) Wikipedia, https://en.wikipedia.org/wiki/Geometric_distribution, Accessed: May 14, 2021

Goldstone, J.A. (2002) Population and security: How demographic change can lead to violent conflict, Journal of International Affairs, Vol. 56, No. 1, 20 pp.

Hadley, L, (2019) Borders and the feasibility of rebel conflict, Borders in Globalization Review, Vol. 1, No. 1, pp. 66-82, https://doi.org/10.18357/bigr11201919259

Halkia, M., S. Ferri, M. Papazoglou, M-S. van Damme, G. Jenkinson, K-M. Baumann, D. Thomakos (2019) Dynamic global conflict risk index, JRC Technical Reports, European Commission, Publications Office of the EU, https://doi.org/10.2760/846412.

Halkia, M., S. Ferri, M. Papazoglou, M-S. van Damme, D. Thomakos (2020) Conflict event modelling: Research experiment and event data limitations, Conference: LREC 2020 Workshop on Automated Event Extraction of Socio-political Events from News (AESPEN2020), Marseille, France.
https://www.researchgate.net/publication/341378782_Conflict_Event_Modelling_Research_Experiment_and_Event_Data_Limitations

Hamblin, R.L., M. Hout, J.L.L. Miller, B.L. Pitcher (1977) Arms races: A test of two models, American Sociological Review, Vol. 42, No. 2 (April 1977), pp. 338-354, https://doi.org/10.2307/2094609

Hatfield and McCoy Feud (2021), Wikipedia, https://en.wikipedia.org/wiki/Hatfield%E2%80%93McCoy_feud, Accessed: May 12, 2021, classic example of self-perpetuated conflict.

Infectious Disease Models (2021), Mathematical modeling of infectious disease, Wikipedia, https://en.wikipedia.org/wiki/Mathematical_modelling_of_infectious_disease, Accessed: May 14, 2021.

Kim, M., K., D. Paini, and R. Jurdak (2019) Modeling stochastic processes in disease spread across a heterogeneous social system, Proceedings of the National Academy of Sciences, Vol. 116, No. 2, 401-406, https://doi.org/10.1073/pnas.1801429116.

Lee, E.D., B.C. Daniels, C.R. Myers, D.C. Krakauer, and J.C. Flack (2020a) Scaling theory of armed-conflict avalanches, Physical Review E, Vol. 102, No. 042312, 12 pp., http://doi.org/10.1103/PhysRevE.102.042312

Lee, E.D., B.C. Daniels, C.R. Myers, D.C. Krakauer, and J.C. Flack (2020b) Emergent regularities and scaling in armed conflict data, arXiv:1903.07762v4 [physics.soc-ph].

Logistic Regression (2021), Wikipedia, https://en.wikipedia.org/wiki/Logistic_regression, Accessed: May 12, 2021

Morrow, J.D. (1986) A spatial model of international conflict, The American Political Science Review, Vol. 80, No. 4 (Dec. 1986), pp. 1131-1150, American Political Science Association, Stable URL: https://www.jstor.org/stable/1960860.

Nailufar, B., S. Wijaya, and D. Perwitasari (2015) Landscape modeling for human - Sulawesi crested black macaques conflict in North Sulawesi, Procedia Environmental Sciences, 24, pp. 104-110. http://doi.org/10.1016/j.proenv.2015.03.014. (example of logistic regression)

Rustad, S.A., H. Buhaug, Å. Falch, and S. Gates (2011) All conflict is local: Modeling subnational variation in civil conflict risk, Conflict Management and Peace Science, Vol. 28, No. 1, pp. 15-40.

UCDP (2021) Uppsala Conflict Data Program, https://www.pcr.uu.se/research/ucdp/.

USAID (2012) Conflict Assessment Framework: Application Guide, United States Agency for International Development. 48 pp., https://pdf.usaid.gov/pdf_docs/PNADY740.pdf

ViEWS (2021) A political Violence Early Warning System, Dept. of Peace and Conflict Research, Uppsala University, https://www.pcr.uu.se/research/views/.

Walther, O.J., S.M. Radil, D.G. Russell, and M. Trémolières (2020) Introducing the Spatial Conflict Dynamics indicator of political violence, arXiv:2003.01750v2 [physics.soc-ph].

Yurevich, P.A., M.A. Olegovich, S.V. Mikhailovich, and P.Y. Vasilievich (2018) Modeling conflict in a social system using diffusion equations, Simulation: Transactions of the Society for Modeling and Simulation International, Vol. 94, No. 12, pp. 1053-1061, http://doi.org/10.1177/0037549718761573.