SCAMP User Manual
Chapter 2: SCAMP Model Package Files

Laura Sappelsa, H. Van Dyke Parunak, and Jonathan A. Morell

This chapter defines the files and variables that make up a SCAMP model, contained in the `SCAMP/data/<modelName>` directory, and describes how you can set many of these variables to default values to allow incremental model development and exploration.

We deliver SCAMP with the model of civil conflict that we developed in the DARPA Ground Truth program, and often show examples from this model. SCAMP refers to the modeling platform, while Ground Truth refers to the DARPA program and the model we built there.

Section 1 is an index to the various configuration files and where this chapter discusses them.

Section 2 describes the high-level concept of a scenario and how to define it.

Sections 3-6 describe the Excel workbook (**model.xlsx**) and associated files that specify the various variables in the model. These tabs and files fall into four categories, corresponding to SCAMP's four perspectives.

1. The Causal Event Graph (CEG) defines the various *events* in which agents can participate, and how they are causally related to one another. This graph, with its parameters, is the only required component of the model. The next three are optional.
2. Modelers can define different groups with which agents can be affiliated. Each of these groups can have a hierarchical goal network (HGN) that allows agents to reason about how their choices impact the goals of their own and of other groups.
3. Agents can move through geospace (GEO), as defined in a set of tabs and maps.
4. A social perspective (SN) allows agents to become acquainted with one another as they interact. These interactions can allow them to change their group membership.

In addition to defining the variables, we discuss the open-source software tools we use to construct some of the supporting files. In addition to discussing the contents of each group of parameters, we outline the methods we use to allow a group of analysts to collaborate in constructing the model.

Section 7 describes the default mechanism that allows you to run a model without having to define all of the variables, so that you can get an idea of its behavior and refine it iteratively.

1 Index of Model Package Files

Table 1 provides a list of the model package files in the order they should be created. It lists the perspective, file (without the <version> tag), tab in **model.xlsx** (if applicable), column(s) within the tab (if applicable), component of the perspective (if applicable) and where they are described. The suffix '*' (e.g., "Pref*") indicates that a column name includes other model-specific information, as discussed in detail below.

Table 1. List of model package files, tabs, and variables, organized by perspective.

Perspective	File	Tab	Columns	Component	Where Described
CEG	model.cmap	NA	NA	NA	Section 3.3
CEG	model.xml	NA	NA	NA	Section 3.3

Perspective	File	Tab	Columns	Component	Where Described
CEG	model.xlsx	symbols	Symbol, Value	symbols	Section 3.4
CEG	model.xlsx	fixedFeatures	Id, Feature, Abbr	NA	Section 3.4
CEG	model.xlsx	groups	Pref*, PrefGoal*, PrefPresence*	groups	Section 3.5
CEG	model.xlsx	groups	IndependentAgency, ReceivesAffiliation, MakesAffiliation, GhostsPerShift, ShiftsPerGen, MaxGhostSteps, GroupAvatars, IndividualAvatars, AffiliationThreshold, PrefVariation, UseHGN	agents	Section 3.5.3
CEG	model.xlsx	events	Ag*	agency	Section 3.5.3
CEG	model.xlsx	events	Feat*	features	Section 3.7
CEG	model.xlsx	events	Sticky*	stickiness	Section 0
CEG	model.xlsx	events	TransitTime, EffectTime	delays	Section 3.9
CEG	model.xlsx	TLHR calculation	(all columns)	delays	Section 3.9.2
CEG	model.xlsx	events	Probability	NA	Section 3.10
HGN	hgn.cmap	NA	NA	NA	Section 4.2
HGN	hgn.xml	NA	NA	NA	Section 4.2
HGN	model.xlsx	zips	(all columns)	zips	Section 4.3
GEO	model.ora	NA	NA	NA	Section 5.3
GEO	codebook.docx	NA	NA	NA	Section 0
GEO	model.xlsx	events	Region, Dest*	destinations	Section 5.5
GEO	model.xlsx	regions	id, Name, Layer, Color, Falloff	Regions: Falloff	Section 5.6
GEO	model.xlsx	regions	DelayDiv*	Regions: DelayDiv	Section 5.7
GEO	model.xlsx	groups	StartLocation, Feature#Name, Pct	startLocations	Section 5.8
GEO	model.xlsx	groups	MvmtDelay	mvmtDelays	Section 5.9
Social	model.xlsx	groupChanges	(all columns)	groupChanges	Section 6.2
Social	model.xlsx	namedAgents	Name, Leader?, Influencer?, NumBackups, Notes	namedAgents	Section 6.3
(all)	model.xlsx	defaults	type, ColumnType, Default	defaults	Section 7

The SCAMP program expects to find its input files stored according to a particular directory structure under SCAMP/data/<modelName>. Table 2 lists the files alphabetically, the subdirectory where they should be placed, the perspective that they address, and where this chapter discusses them. The `model` subdirectory contains files processed by SCAMP, and the `source` subdirectory contains files from which these are generated. SCAMP will generate a third subdirectory alongside these two, `geo`, containing geospatial information. The file names are all case sensitive, and <groupName> must be spelled and capitalized the same way that you record it in **model.xlsx**. (SCAMP itself doesn't read the `source` subdirectory, but you will find it very helpful to have your model source available along with the files that SCAMP does read.) We will refer to these three subdirectories as the source subdirectory, the model subdirectory, and the geo subdirectory, without repeating SCAMP/data/<modelName>.

Table 2. List of model package files, organized alphabetically.

File	Subdirectory	Perspective	Where Described
codebook.docx	source	GEO	Section 0
<groupName>.hgn.cmap	source	HGN	Section 4.2
<groupName>.hgn.xml	model	HGN	Section 4.2
CEG.cmap	source	CEG	Section 3.3
map.xcf	source	GEO	Section 5.2
map.ora	model	GEO	Section 5.2
model.xlsx	model	(all)	(all)
CEG.xml	model	CEG	Section 3.3

In the Ground Truth model included with the simulator, <modelName> is SCAMP.CEG0073.

The **model.xlsx** workbook includes numerous columns for the convenience of the modeler that are not used by the program. Table 3 lists these names.

Table 3: Column names in model.xlsx ignored by SCAMP.

Tab	Unused Columns
events	Region
symbols	(All used)
fixedFeatures	id
groupChanges	order, TriggerText, FromGroupText, FromLocText, ToGroupText, To LocText, ConditionalLocText, OLDProb, PromotersText, PromoterLocText, BlockersText, BlockerLocText
groups	Feature1 Name, Feature2 Name, Feature3 Name
defaults	(All used)
namedAgents	Notes
regions	(All used)
zips	Group, Goal Text, Event Text
TLHR calculation	(Tab unused, all columns unused)

In our experience, there are important underlying principles to consider and certain areas in the development of the model package that are likely to become sources of error. Throughout this document, the symbol ⚠ indicates these areas.

One common source of trouble is having information (even spaces) in cells in model.xlsx outside the rows and columns defined in this document. Such extraneous data will almost certainly crash SCAMP, and the error messages will show an array reference error. If this happens to you, carefully examine each tab for extraneous characters, and either delete them or copy the

legitimate cells from the tab into a new one, delete the faulty tab, and rename the new one appropriately.

2 Defining a Scenario

The files defined in this document are a computer-readable representation of a scenario that you, as the modeler, have in mind and wish to study. Before we delve into the technical details of these files, let's think about the scenario you want to study.

Defining a scenario is like writing a novel. You as the modeler are the author. It is up to you to define the characters and their personalities, the groups to which they belong and the goals that motivate those groups, the kinds of events in which they participate, and the geography of he world in which they live. In a novel, you capture these details in sentences, paragraphs, and chapters, to be processed by a human reader. In SCAMP, you capture them in the various model files, to be processed by a computer program. But you still have the freedom (and the responsibility) to define the world in its varied perspectives.

2.1 Groups

SCAMP's great strength is in modeling people, their psychological influences and social dynamics as they interact. So at the outset, you should decide what kind of people you are modeling. Are you modeling the interaction of patients, doctors, and insurance companies? Of governments and their citizens? You can take "people" very broadly—perhaps you want to study an ecological system consisting of fish, sportsmen, polluters, and environmental control officers. Each kind of agent is a group. Every agent must belong to a group, though they can have sympathies with other groups.

2.2 Events

The heart of SCAMP is the Causal Event Graph (CEG, Section 3), a directed graph whose nodes are events. An event is something that someone does and that unfolds over time (possibly very short). SCAMP allows you to define a special Environment group to do "acts of nature," such as droughts, economic collapse, hurricanes, or other events that could impact your scenario but that are not performed by the agents you are modeling. An edge from one event to another indicates that it would be a coherent story for an agent who has participated in the first event to go on and participate in the second. The graph starts with a special event START, and ends with a special node STOP, and with a few exceptions, there must be a path to every event from START, and a path from every event to STOP. Each path from START to STOP should be a credible narrative of what an individual agent might do. Thus the entire graph is a *narrative space*.

A natural way to construct a narrative space is to tell stories about a typical agent. For each story, write down each event in which the agent participates, in the order in which the agent would participate in them. Write several stories. You can reuse parts of a story: an agent's day might always start out the same way (eating breakfast, commuting to work, and working all day), but end in different ways (dinner with friends, a walk in the park, going to the gym). So your graph may well branch. As you tell individual stories, pay attention to places where you can join them together in this way. The longer your individual stories, and the more places you link them together, the more variety of behavior you can expect from your model.

The kinds of stories you might tell about a government agent will not be the same as those you tell about citizens, so in general you will have a separate graph for each group. If there are events

that make sense for two groups, you can share them between the individual group graphs, thus joining the graphs together, but be sure that there is a way for agents of both groups to get to such events, and also paths from the event that make sense for both groups. All these group graphs can share the same START event. They can share a common STOP event, or each can have its own <groupName>STOP event.

The edges in the story graph that form coherent narrative snippets are called *agency edges*. Agents move over these edges as their lives unfold, choosing between alternatives when they have more than one successor event available. But sometimes one event can *influence* another, even if agents cannot move directly between them. Note such pairs of events. We will show you how to define *influence edges* between them.

The resulting Causal Event Graph is the heart of SCAMP. You can enrich its behavior with the other three perspectives: goals, geospatial movement, and social interactions. Each of these is connected to the CEG.

2.3 Goals

Agents' decisions are based not only on the immediate options they face, but also by their longer-range *goals*. These goals can be arranged in a hierarchy, with subgoals indicating what must be done to achieve higher-level goals. Each group has its own hierarchy of goals and subgoals, called a Hierarchical Goal Network (HGN, Section 4). For example, the government's top-level goal might be, "Remain in power," and it might do so in one of two ways: "Exercise overt force" (e.g., a police state), or "Control economic incentives." Each of these in turn could have subgoals.

In principle, you could go on breaking goals into subgoals forever, but at some point you have to stop, leaving a bottom level of subgoals. You will connect these subgoals to individual events in the CEG that either support them or block them. (We call this connection process "zipping.") Each group has its own CEG subgraph and its own HGN, but it is perfectly OK to zip events in one group's subgraph to subgoals in another group's HGN.

2.4 Which comes First—the Events or the Goals?

Modelers need to start their work with an understanding of both the *events* in which agents participate, and the *goals* the agents pursue as they engage in events. In writing down your story, which should you define first—an agent's actions (events), or the reasons for those actions (goals)? Of course, you will find yourself going back and forth, but under different circumstances, you might start with either one. It may help to ask yourself three questions.

Does it make sense to think of the agents as having goals? Agents may exist in a world where their behavior is driven tactically, where only short-term decision-making matters. As an example, agents may exist in a world where their only motivation is a high-level goal such as "survival." In such cases one might be able to envision intermediate goals and the making of plans, but really, the best way to understand agent behavior is in almost exclusively tactical terms, and goals may not be needed at all, or can be added after the events are defined.

Do I want to model how agents' existing behavior fits with their goals? Sometimes your emphasis is on the empirical reality of the agents' actions, which you have observed in the world. You can imagine the goals that lie behind them, but your first concern is to capture empirical behavior. As an example, imagine agents as being "governments." In this scenario, the high-level

goal might be “remain in power,” with subsidiary goals such as “satisfied electorate” and “loyalty of the military.” You might begin by describing the events that are taking place. Next, think of the goals that hose events support or hinder. In a scenario such as this, the ordering of modeling goals and events may not matter. The events are as they are. Goal trees can be constructed. It does not matter what comes first as long as connections between goals and events are drawn correctly.

Am I constructing a theoretical world in order to explore possible relationships? The purpose of your model may be to test theory. As an example, imagine that you wish to test a hypothesis about how townships pass zoning ordinances. You may have reason to believe that the racial and economic mix of residents affects such ordinances. Here, you might define invent events such as “immigrants move into town,” “long-term residents retire and move away,” and “existing industries close because of foreign competition.” You may then decide that the actions of various types of agents matter, e.g. newcomers, existing residents, school system, and city council. You may have reason to believe that these agents and these events are salient, and may wish to build an artificial world in which you can test possible relationships. In a case like this, it would be advisable to construct a goal network first so that a rigorous process of theorizing and literature searching can be done to identify the relevant events.

These are not pure choices. It is not hard to envision scenarios that combine some groups who act only in tactical ways, an existing set of events that must be recognized, and there are hypotheses whose exploration requires constructing an idealized world. What matters is that modelers make a reasoned choice of “event first” or “goal first” based on reasoned answers to each of these questions.

2.5 Geospace

Another enhancement to the CEG is a geospatial model (Section 4.4), which is basically a map of the terrain involved in your scenario with distinctive locations and regions identified. If you use this perspective, each agent always has a current location somewhere in geospace. You will assign each group a region within which its agents originate, and SCAMP will distribute them randomly within that space. In addition, you can define a spatial destination with some events. When an agent participates in such an event, SCAMP will put it into geospace, and it will move toward the destination. Along the way, it may interact with other agents who are there at the same time, thus delaying or diverting its path. When it reaches its destination, it will return to the event and go on to pick the next one.

2.6 Social Dynamics

The third enhancement to the CEG is extended social dynamics (Section 6). If this is active, agents keep track of other agents whom they have encountered, either on events in the CEG or in geospace, and their preferences can change based on these encounters. Eventually, an agent may develop a desire to change its home group. The social network portion of SCAMP will allow you to define sophisticated rules, attached to selected events or locations, to make such changes happen. These rules can change not only an agent’s group, but also its current event, its current location, or even its existence (modeling the birth and death of agents).

The important insight to keep in mind is that the CEG is central, and that every trajectory through the CEG should be a coherent story about the behavior of one of your agents.

2.7 Overview of the Modeling Process

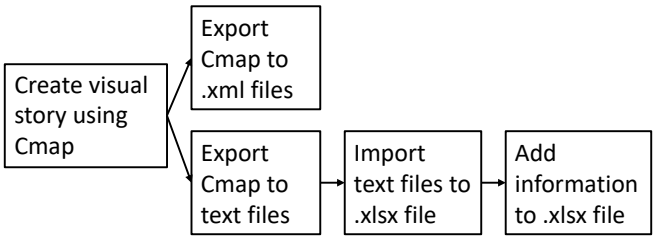

Figure 1: High-level view of SCAMP modeling process

Figure 1 shows the overall flow of the SCAMP modeling process.

The process begins with a graphics-based application called Cmap, which lets domain expert modelers construct various maps that represent a story. The SCAMP *program*, however, cannot read Cmap files. Also, an entire SCAMP model includes information about the story that cannot be put into Cmap files. Fortunately, Cmap has the capacity to output its maps as text files that have two characteristics. First, they allow modelers to add information. Second, they can be read by the SCAMP code, thus allowing the models to run.

Modelers load the text files into **model.xlsx**, the main SCAMP spreadsheet in the familiar Excel spreadsheet format. During the course of the SCAMP process, modelers will add their expert information to this file. Other Cmap information is converted into .xml files. The modelers will have to "push the buttons" to create the .xml files, but once modelers do this, they will have no need to work with the .xml files again. There are also a few intermediate steps between Cmap and Excel that involve text files, but don't worry about that. They are there to help with processing, and we will explain them carefully.

As you read these instructions you will find much detail about how the Cmap files should be constructed and what should be done with and to the Excel files. It will get detailed, and the detail matters because the SCAMP code is unforgiving in what it expects. But throughout, keep the overall process in mind. It will keep you from getting lost in the detail.

2.8 CmapTools

To construct the CEG and HGNs, you will use the CmapTools concept modeling program, a freeware utility developed by the Florida Institute for Human and Machine Cognition [1] and available online. Here are some useful hints.

1. CmapTools stores its Cmaps in a folder "My Cmaps" that it will create in your Documents directory when you install it, and does not let you directly store them somewhere else. So the first time you create a Cmap, you should move it from that directory into wherever you keep your other model files. We recommend you keep your Cmaps in the source subdirectory. If you open a Cmap from this directory, modify it, and save it, it will update in this directory, not in the My Cmaps folder.
2. Keep an eye on filenames that Cmap generates. Often it insists on adding ".cmap" to them. Other than the legitimate .cmap extension on the Cmap model file, you can rename the files to remove the extra characters.
3. If you construct large models, you may encounter "out of memory" error messages when you try to save cmap files. To avoid this, increase the memory of CmapTools. Refer to the instructions in CmapTools FAQs under the section "How do I increase the memory of CmapTools and CmapServer?" [2]

3 Causal Event Graph (CEG)

The causal event graph is a straightforward, intuitively clear representation that draws on the natural preference of human cognition for stories over other forms or reasoning [3]. It is the core of the model. By moving agents over the causal representation rather than embedding the causal model in the agents, SCAMP provides a central representation of the overall causal process. Movement over the graph is constrained by the relations between the nodes, and by features that characterize each node. In the CEG, nodes are events and edges are the relations between them.

3.1 CEG Model Package Files

The CEG portion of the model package includes two files and three tabs in the model workbook. Table 4 summarizes their contents.

Table 4. CEG model package components

Model Package Component	Location	Purpose
CEG Cmap file	Separate file	Allows modeler to construct and modify structure of the CEG
CEG xml file	Separate file	Version of the CEG that SCAMP consumes
groups	model.xlsx tab	Defines the different groups supported in the model
events	model.xlsx tab	Defines information specific to individual events in the CEG, including which groups have agency for them, their fixed features, event time information, as well as which events are "sticky."
TLHR calculation	model.xlsx tab	"The Land has Rest" calculation associated with the event time information.

3.2 CEG Construction

The model building process should make sense to the user. While we provide examples taken from the initial Ground Truth model, the user should tailor the instructions to their model. For example, the sim does not distinguish between colors or edge styles. These options can be used to facilitate visualization and understanding of the CEG.

The CEG has three components: events, agency edges and influence edges. The events define the alternative activities among which agents can choose. Agency edges indicate the choices available to an agent who is currently participating in an event. Influence edges indicate causal influences among events between which agents cannot move directly.

3.2.1 Defining Events

The events define the choices available to agents for their activities. Strictly speaking, these are event *types*; a specific event occurs when agents begin to participate in an event type.

SCAMP's perspective on events is behavioristic, not mentalistic. An event is something that an external observer could detect, not a psychological event. SCAMP's preference-feature and goal mechanisms do model internal processes, but the events themselves are external, and not appropriate for modeling internal psychological events accessible only to the participating agents.

1. Open CmapTools.
2. To create a new CEG, Select File, then New Cmap. A new window will open.
3. Select File, Save Cmap As
   a. The Save Cmap As window will open
   b. Under Name, enter <modelName>. It will save in the .cmap file format.

   c. Cmap files are saved on the hard drive under users/<yourname>/Documents/My Cmaps. It's a good idea to save a copy elsewhere on your hard drive. Ultimately, the file goes in the source subdirectory, and you may want to keep a copy there.

When editing a CEG, save the file prior to making any changes. The CEG will automatically save as you edit it, and if it is not saved with a new name before making edits, the previous version will be lost.

4. Double-click anywhere in the window to create a node.
5. A single start node serves as the origin of all nodes that otherwise have no incoming edges.
   a. Double-click ???? inside the node and type START. This label must be all upper-case.
   b. All nodes must be accessible from the START node.
6. Create another node and replace ???? with an event id number and label.
   a. All nodes must have unique names, with no commas in the name.
   b. Each node must have only one event.
   c. Each event must have a unique number associated with it.
   d. There must be no space before the number.
   e. There should be only a single space between the event id number and the event label.
   f. Event labels are comprised of nouns and verbs.
   g. An event must be something that happens, not just a situation that exists.
   h. Do not use passive voice.
   i. Events must be agented by at least one group.
   j. A group has agency in an event when it is contributing to making that event happen.
   k. Agents can only visit events for which they have agency.
   l. Each group will have its own subgraph.
   m. Examples (Figure 2 illustrates some event labels).
      i. "Government cuts public services" is agented by the Government.
      ii. "Government cuts public services" is agented by the Government.
   n. If geospace will be part of your model, some of your events must involve or imply movement to a destination.
7. Keep track of references that inform event labels and edges.

Figure 2. Example from Ground Truth: Labeling events in the CEG.

The Environment group (ENV) generates background events (nature or acts of God) that are not the direct result of decisions by actors. ENV also includes events by agents that are not otherwise included in the model, and limited in number to the extent that creating a new group is deemed unnecessary (for example, occasional interventions by other governments). Consider how a potential ENV event would affect the social reality of the model and how important is it to the sim because ENV events are scored differently than other events. Rules for ENV events

1. None of the other groups affiliate with ENV.
2. Every ENV event must have an influence edge going to an event in another group's subgraph, or coming from another group's subgraph.
3. No ENV event should appear in any other subgraph.
4. Agented events in other subgraphs can influence ENV events (e.g., govt monetary actions could impact ENV's inflation and economic decline).
5. ENV events are not agented by any additional groups.

6. ENV events are not zipped to HGNs.
7. The ENV subgraph has an event called “The Land has Rest,” inspired by a common expression throughout Israel's ancient history describing periods of peace and relative prosperity [4]. The idea is that a realistic environment will not always be malicious and there could be some periods of calm.

Nodes with one color indicate the group of agents who control the event represented by the nodes, while nodes with two colors indicate events involving multiple groups of agents or resulting from impersonal forces of nature.

1. To color a node, select it and then select Window, then Show Style Palette.
2. Use the Color drop-down menu to select a color. Colors have no effect on the sim, but they do facilitate construction of the CEG. Table 5 shows the colors used in Ground Truth.

Table 5. Node colors are used to represent different groups.

Group	Color	Group	Color
Government		Government + Violent Extremists	
Armed Opposition Forces		Government + Military	
Relief Agencies		Government + Military + Armed Opposition Forces	
People		Government + Military + Violent Extremists	
Violent Extremists		Government + Military + People	
Military		Armed Opposition Forces + Violent Extremists	
Environment		Government + Armed Opposition Forces + Violent Extremists + People	
Government + Armed Opposition Forces			

3.2.2 Adding Agency Edges

Agency edges represent sequential causal constraints between successive events in which agents might participate. Different groups have agency for different events. For example, the government can do some things that citizens cannot. Agency edges connect events for which the same groups have agency, so that it makes sense for an agent, having participated in the first event, to move to the second and participate in it. The agency edge indicates whether this movement is causally coherent and every directed path of agency edges through the graph should represent a coherent narrative [5].

The type of edge details the causal relations among events. There are two types of agency edges

1. `then` joints a single antecedent event to a single consequent. A consequent may have multiple antecedents by way of then causal relations, meaning that many different causes may have the same result.
2. `thenGroup` joins a single antecedent event to multiple consequents. The agent participates in all the events concurrently, for as long as the maximum duration of any of the concurrent events. Since an agent can only be one place at a time, only one of the events can be a geospatial event.

Rules for agency edges

1. Each group type defines a subgraph of the CEG.
2. Agency edges satisfy the lattice rule

a. Each subgraph starts at START, and ends at <Group> STOP. (ENV does not require s STOP node.) The string STOP, like START, must be all upper-case. By default, when an agent reaches STOP, it returns to START.

Figure 4. Double arrow above START node.

b. Every node in a subgraph must be accessible by agents. In most models, this means that there must be a path of agency edges from START to the node. The group change feature described in Section 6.2 can move agents to a node without following agency edges, so if you are using this feature, you may have nodes with no incoming agency edges, but as a rule of thumb, be sure that all your nodes are reachable from START.

c. <Group> STOP must be reachable from every node in the group's subgraph.

3. Avoid causal paths with little or no branching.
4. Subgraphs may (should) share nodes
5. Agency edges may be either within or between subgraphs

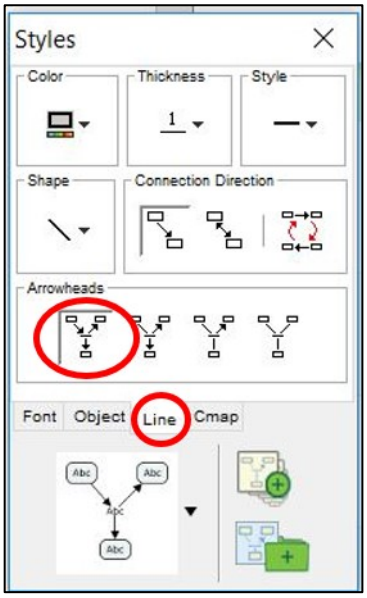

Figure 3. Select arrowhead design in the Style

To add agency edges to the CEG

1. Select the START node and then click on the double arrow that appears above it. See Figure 4.
2. Drag the cursor to the inside of the other node and then click again to create the relationship edge.
3. Select the lines of the edge on either side of "????"
4. Open the Style Palette by selecting Window, then Show Style Palette
5. In the Style Palette, select Line then select the first type of arrowheads, as indicated by the red circles in Figure 3. The edge should look like the one depicted in Figure 6.
6. Label the edges by double-clicking on "????" in the middle of the arrow and type in the name of the relation, being careful not to add any additional spaces or characters. All edge labels are in lowercase.
7. To match the edge color of the node color, Select Window, then Show Style Palette. See Figure 5.
   a. Select both parts of the edge.
   b. Select the Line tab
   c. Select the Color
   d. Select Thickness (3). The result should look similar to Figure 7.

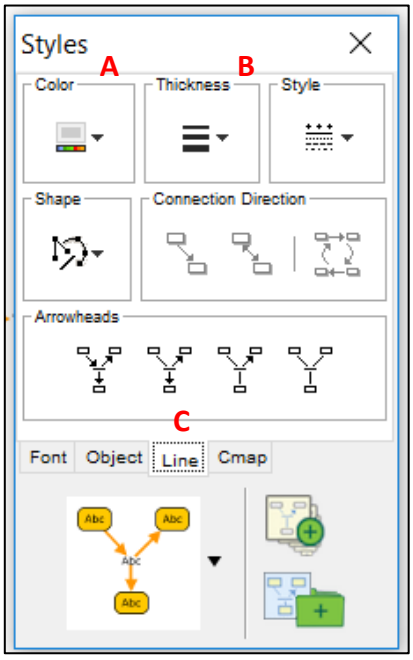

Figure 5. The Style Palette is used for edge color and thickness.

3.2.3 Adding influence edges

Influence edges are used between events that are agented by different groups, affecting the likelihood that an event will occur. Influence edges indicate that activity on the source of the edge influence the selection of the node at the destination, either by affecting whether it is included among the options

available to agents (enable, prevent), or by changing its attractiveness (enhance, inhibit). Agents do not travel over influence edges. It is important to remember that influence edges can only act when agents have participated in the events from which the edges originate. Thus these edges impose a temporal constraint between their source and their destination.

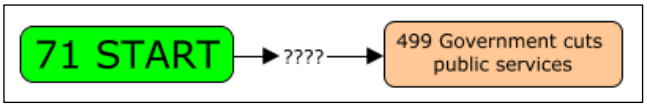

Figure 6. Example from Ground Truth: Edge arrowheads.

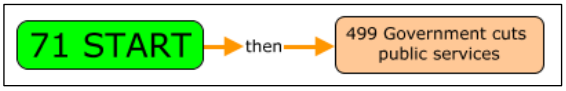

Figure 7. Example from Ground Truth: Properly labeled edge.

There are two categories of influence edges, each with two types:

1. Soft influence edges
   a. `enhance` asserts that occurrence of the antecedent renders the consequent more likely to happen.
   b. `inhibit` asserts that occurrence of the antecedent renders the consequent less likely to happen.
2. Hard influence edges
   a. `enable` asserts that occurrence of the antecedent requires the consequent to happen.
   b. `prevent` asserts that occurrence of the antecedent blocks the consequent from happening.

Rules for influence edges

1. Any event with an outgoing `inhibit` or `prevent` edge must also have an outgoing `then` edge.
2. Influence edges may be either within or between subgraphs.

⚠ Too many hard influence edges, or edges in the wrong places, can lock up the model.

To add influence edges to the CEG

1. Label the edges by double-clicking on "????" in the middle of the arrow and type in the name of the relation, being careful not to add any additional spaces or characters. All edge labels are in lowercase.
2. To distinguish them from agency edges, influence edges are drawn as red dashed arrows. See Figure 8.
   a. Select Window, then Show Style Palette
   b. Select both parts of the edge
   c. Select the Line tab
   d. Select the Color
   e. Select Thickness (3)
   f. Select Style (dashed)
3. Save the CEG as **<ModelName>.cmap**.

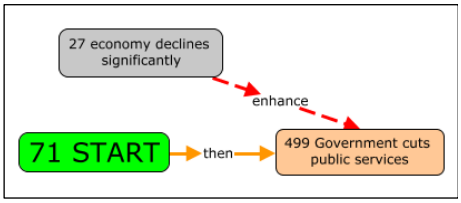

Figure 8. Influence edges have dashed lines; causal edges have solid lines.

3.2.4 Decreasing the .cmap file size

The file size of the *.cmap file can suddenly start increasing. If it reaches 10 MB or more, you may not be able to open it. For reference, a CEG with over 400 nodes is less than 1 MB. If saving the *.cmap file takes longer than usual, check the file size.

1. Open the CmapTools program

2. In the Views – CmapTools window, navigate to the file
   a. Select Export Cmap As, then CXL file
   b. Save file
3. In the Views – CmapTools window, select File, then Import, then Cmap from CXL file
   a. Navigate to the file
   b. Click Import
      i. The ***.cmap** file will open
   c. Save the file
4. Navigate to the location on your hard drive where the *.cmap files are saved.
5. Check the file size to confirm it has decreased.

3.2.5 Team CEG construction

Team members construct fragments of the CEG independently. The team lead acts as a curator, connecting the fragments and then all team members review the entire CEG. This process repeats each time the CEG is updated.

3.2.6 Reviewing the CEG

1. All nodes must have at least one incoming and outgoing edge (except for STOP, or nodes that are sources or destinations for group change rules, Section 6.1.)

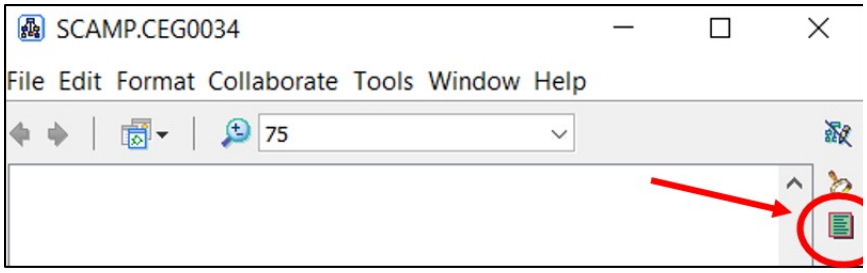

Figure 9. Cmap list view.

   a. Use Cmap List View. See Figure 9.
   b. Select the Concepts tab
   c. Sort by Concepts to be sure you don't have any blank concepts. CmapTools creates a new concept just by clicking, and it's very easy to have "orphan" concepts floating around, which will crash the simulator. You can quickly spot such problems in this view.
   d. Sort by Links In. See Figure 10.
      i. There should not be any nodes with zero links in or incoming edges.
      ii. Even the START node has one incoming edge from The Land has Rest.
   e. Sort by Links Out. Only the STOP nodes should have zero links out or outgoing edges.
   f. Select the Linking Phrases tab, and sort by linking phrase to bring identical links together. Be sure that the only linking phrases you see are

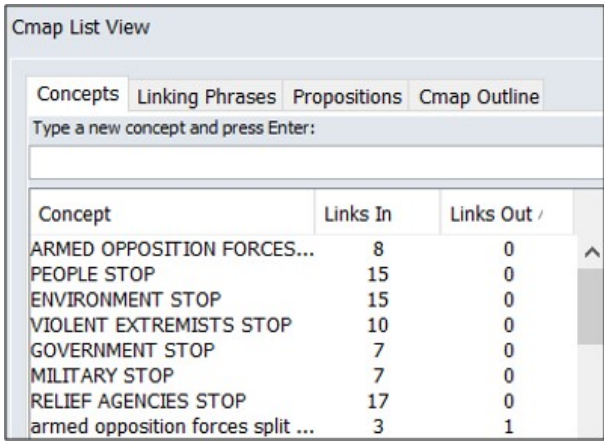

Figure 10. Cmap List View Links In and Links Out.

then, thenGroup, enable, enhance, inhibit, prevent, support, block, and, and or.

2. Confirm that all edges have a label by searching the CEG for "?"
3. Search CEG for "than" and change to "then"
4. Check for misspelled event and edge labels by going to the Tools menu and selecting Spelling
5. Confirm that there are no duplicate edges
   a. File, Export Cmap As, Propositions as text
   b. Open a new Excel file.
   c. Import the data from the *.txt file into Microsoft Excel.
   d. Check for duplicates. If there are any, delete them from the CEG.

3.2.7 Format of the CEG files

Two file formats of the CEG are needed for the model package.

1. *.cmap file format
   a. This is the file format of the CEG created in Cmap, and is used by analysts who wish to examine and modify the CEG.
   b. Save the CEG file as **<modelName>.cmap**
   c. Copy the final version of the file to the source subdirectory.
2. *.xml file format
   a. This is the format that the SCAMP system consumes.
   b. Open **<modelName>.cmap**
   c. Select File, Export CMap As, CXL file
   d. Change the name and extension of the file, so that it is called **CEG.xml**. Your file system may require you to approve changing the extension.
   e. Add the final version of the file to the model subdirectory.

3.2.8 Formatting the tabs in model.xlsx

⚠ Additional columns and tabs can be added to model.xlsx, as long as they do not have the same names as those read by the sim. Table 3 lists column names in the current version of model.xlsx ignored by SCAMP. All other column names are required.

Some of the columns in **model.xlsx** have names that depend on the groups in the model and the fixed features that those groups support. Section 3.4 tells how to define the abbreviations for the fixed features used in these headings, and Section 3.5 tells how to define the group abbreviations. In this section, headings that require extension indicate the required extensions with <Gp> or <Ff>. One instance of such a column is required for each combination of extensions. Thus for a model with three groups, with abbreviations G1, G2, and G3, Dest<Gp> stands for three columns: DestG1, DestG2, and DestG3.

Some entries in **model.xlsx** are numerical, and some are strings of characters. When SCAMP compiles **model.xlsx**, it reads them all as strings, and then converts any that it needs as numbers back into numerical form. Thus modelers are free to use strings as they desire. In particular, in some contexts, it may be convenient to use a string representation for identifiers (ids) so that they can be copied without automatically incrementing.

1. Create a new Excel workbook and save in the model subdirectory as **model.xlsx**.
2. Create a tab called "events" with columns labeled
   a. id

   b. Event
   c. Feat<Gp><Ff>
   d. Probability
   e. Ag<Gp>
   f. TransitTime
   g. EffectTime
   h. Region
   i. Dest<Gp>
   j. Sticky<Gp>
3. Create a tab called “symbols” with columns labeled
   a. Symbol
   b. Value
4. Create a tab called “fixedFeatures” with columns labeled
   a. id
   b. Feature
   c. Abbr
5. Create a tab called “groupChanges” with the following columns. Columns ending “Text” are for the convenience of the modelers, and are not used by the program.
   a. order
   b. Trigger
   c. TriggerText
   d. FromGroup
   e. FromGroupText
   f. FromLoc
   g. FromLocText
   h. FromLimit
   i. ToGroup
   j. ToGroupText
   k. ToLoc
   l. ToLocText
   m. ConditionalGroup
   n. ConditionalGroupText
   o. ConditionalLoc
   p. ConditionalLocText
   q. Prob
   r. Promoters
   s. PromotersText
   t. PromoterLoc
   u. PromoterLocText
   v. Blockers
   w. BlockersText
   x. BlockerLoc
   y. BlockerLocText
6. Create a tab called “groups” with columns labeled
   a. id
   b. Group

   c. GroupAbbr
   d. Pref<Ff>
   e. PrefGoal<Gp>
   f. PrefPresence<Gp>
   g. StartLocation1
   h. Feature1 Name
   i. Pct1
   j. StartLocation2
   k. Feature2 Name
   l. Pct2
   m. StartLocation3
   n. Feature3 Name
   o. Pct3
   p. MvmtDelay
   q. IndependentAgency
   r. ReceivesAffiliation
   s. MakesAffiliation
   t. GhostsPerShift
   u. ShiftsPerGen
   v. MaxGhostSteps
   w. GroupAvatars
   x. IndividualAvatars
   y. AffiliationThreshold
   z. PrefVariation
   aa. UseHGN
7. Create a tab called “defaults” with columns labeled
   a. type
   b. ColumnType
   c. Default
8. Create a tab called “namedAgents” with columns labeled
   a. Name
   b. Leader?
   c. Influencer?
   d. NumBackups
   e. Notes
9. Create a tab called “regions” with columns labeled
   a. id
   b. Name
   c. Layer
   d. Color
   e. Falloff
   f. DelayDiv
10. Create a tab called “zips” with columns labeled
   a. id
   b. Group
   c. Goal ID

   d. Goal Text
   e. Event ID
   f. Event Text
   g. Relation
11. Create a tab called “TLHR calculation” with columns labeled
   a. #
   b. Name
   c. env
   d. TT
   e. ET
   f. mean_median
   g. transitTime
   h. effectTime

3.2.9 Generating a list of nodes from the CEG

The events tab in the Excel workbook **model.xlsx** requires a list of all events in the CEG. To export the list of events from the CEG

1. Open CEG in CmapTools
2. Go to File, Export Cmap As, Propositions as text.
3. Save as a ***.txt**.
4. Separate the event ID numbers from the event labels.
   a. Import the data from the *.txt file into Microsoft Excel.
      i. Delete the column with the edge labels.
      ii. Cut and paste the events into a single column, sort, and remove the duplicates.
   b. On the Design tab, de-select Header Row and Banded Rows.
      i. Select any cell in the column.
      ii. Go to the Design tab, then Tools and Click the Convert to Range tool. Then click OK.
   c. Put the events in column C and, so that Columns A and B are blank
   d. To create a list with only the ID numbers, use the string function in Column A: =VALUE(LEFT(C1, SEARCH(" ",C1,1)-1)).
   e. To create a list with only the Event names, use the string function in Column B: =MID(C1,FIND(" ",C1)+1,1000)
5. Copy and paste the event ids into `id` and the event labels into `Event` of the **events** tab of **model.xlsx**. See Table 6 for the correct format.
6. Save **model.xlsx**.

Table 6. Example from Ground Truth: Partial list of CEG events.

id	Event
1	Government security Forces attack protesters
2	large numbers of people move to urban areas
3	government forces increase attacks on protesters
4	Protests increase across country
5	public demands democratic reforms

3.3 Symbols

Many of these parameters are numerical, in the range [-1, 1]. The modeling team developed a scale to avoid the precision and to facilitate the scoring process. The description of this process serves as an example and is not the only way to do it. We were striving for a way of looking at it that fits the cognitive preferences of the modelers and matches the cognitive preferences of the software team.

The **symbols** tab defines the nominal values associated with the symbols L(arge positive), M(edium positive), S(mall positive), 0, s(mall negative), m(edium negative), and l(arge negative), where L indicates that the agent is strongly attracted to events that present a positive value to the associated feature, and 1 indicates that the agent is strongly repelled. When we say 0.9, we are really saying there is a distribution with a mean of 0.9. It is what we would psychologically call a L(arge positive). 0 means there is no impact or it is not discernable. In some cases, negative preferences may not be meaningful.

The **symbols** tab has two columns:

- *Symbol* gives a single-character symbol
- *Value* gives a numerical value between -1 and 1, which is the nominal value associated with the symbol.

Table 7 shows the scale developed by the modeling team and the corresponding values that are entered in the **symbols** tab. Table 8 shows the **symbols** tab from Ground Truth.

The simulator replaces each symbol with a value selected uniformly and randomly around the symbol.

Table 7. Example from Ground Truth: Scale, values and symbols.

Scale	Value	Symbol
-1.0	-0.9	l
-0.9		
-0.8		
-0.7	-0.5	m
-0.6		
-0.5		
-0.4		
-0.3	-0.2	s
-0.2		
-0.1		
0	0	0
0.1	0.2	S
0.2		
0.3		
0.4	0.5	M
0.5		
0.6		
0.7		
0.8	0.9	L
0.9		
1.0		

3.4 Fixed Features

A SCAMP model can have a number of *exogenous features* with each event. These features are not modulated by the execution of the program, but are specified by the modelers. The model we constructed for Ground Truth had three such features: the impact of the event on the economic, physical, and psychological wellbeing of agents who participate in it. Other models could have different fixed features, or even a different number of them. These features are involved in defining many columns in model.xlsx, and this tab defines the abbreviations that are used elsewhere. It has three columns:

Table 8. Example from Ground Truth: Symbols tab.

Symbol	Value
l	-0.9
m	-0.5
s	-0.2
0	0
S	0.2
M	0.5
L	0.9

- *id* is a unique number for each feature
- *Feature* is a human-readable name for the feature
- *Abbr* is a short abbreviation used in naming columns elsewhere in model.xlsx. It can consist of uppercase and lowercase letters, as well as numbers, but it should not include punctuation, spaces, or special characters.

Table 9 shows this tab in our current model. Section 3.7 discusses how to score these features.

3.5 Groups

A group is a set of agents that share the same goal structure. A group has agency in an event when it is eligible to participate in that event. At least one group must have agency for each event. Agents of a given group can only visit events for which they have

Table 9: Example from Ground Truth: fixedFeatures tab

id	Feature	Abbr
0	Economic	Econ
1	Physical	Phys
2	Psychological	Psych

agency. Agents also can contribute to making events happen for which they have no agency, via influence edges.

First, list the groups and group abbreviations in the **groups** tab of **model.xlsx**. While there is no hard limit for the number of groups, most models will likely have ten or fewer. Table 11 is a partial list of the groups and abbreviations in Ground Truth.

Table 11. Example from Ground Truth: List of groups and group abbreviations.

id	Group	GroupAbbr
0	Government	Gov
1	ArmedOppositionForces	AO
2	ReliefAgencies	RA
3	People	Peo

Group preferences for each of the features that events present are listed in the **groups** tab of **model.xlsx**. Table 10 lists the columns, formats and definitions.

Table 10. Columns, format and definitions for preferences.

Column	Format	Definition
Pref<Ff>	"Pref" plus abbr from the fixedFeatures tab (Prefabbr with no additional spaces)	A measure of how much a group cares about that feature
PrefGoal<Gp>	"PrefGoal" plus GroupAbbr from the groups tab (PrefGoalGroupabbr with no additional spaces)	A measure of a group's attitude towards achieving its goal and how much each group cares about other groups achieving their goals
PrefPresence<Gp>	"PrefPresence" plus GroupAbbr from the groups tab (PrefPresenceGroupAbbr with no additional spaces)	A measure of how much a group wants to be around members of its own and other groups.

3.5.1 Fixed Features: Independent scoring

The scoring categories in the **symbols** tab of **model.xlsx** are used to score the group preferences.

1. `Pref<Ff>` is a measure of the group's concern for its overall well-being.
   a. An agent's preference for an event's impact on its well-being cannot be negative. Choices for the scores of the preferences for the static features are 0, S, M, L.
   b. A score of 0 means that a group does not care about a particular feature for any of the events.
   c. For the Environment group, the scores of the preferences for the fixed features are L, 0, 0. The meanings of the fixed features don't matter. (After all, the Environment agent stands for impersonal Nature, and doesn't have any preferences.) What matters is that the product of this preference and the value of the first fixed feature be greater than zero, to make the Environment agent work correctly.

   ⚠ The Environment group is not scored the same way as the other groups.

   d. Enter scores into `Pref` of the **groups** tab of **model.xlsx**.
2. `PrefGoal<Gp>` is a measure of how much a group supports the goals of another group
   a. Choices for the scores of the preferences for goal features are l, m, s, 0, S, M, L, where L indicates that the group strongly supports the goals of another group, and l indicates that the group strongly opposes the goal of another group.
   b. A score of 0 means that a group does not care about a particular feature for any of the events.
   c. For the Environment group, all of the scores of the preferences for the goal features are 0.
   d. Enter scores into `PrefGoal<Gp>` of the **groups** tab of **model.xlsx**.

3. `PrefPresence<Gp>` is a measure of the degree to which one group wants to be near another group. Keep in mind that a group may not support another group's goals (`PrefGoal<Gp>`), but may still want to be near that group. An example of this is two groups that are in a battle.
   i. Choices for the scores of the preferences for pheromones are l, m, s, 0, S, M, L, where L indicates that the group wants to be in close proximity to another group, and l indicates that the group wants to be far away from another group.
   j. A score of 0 means that a group does not care about a particular feature for any of the events.
   k. For the Environment group, all of the scores of the preferences for the pheromones are M.
   l. Enter scores into `PrefPresence<Gp>` of the **groups** tab of **model.xlsx**.

3.5.2 Fixed Features: Team scoring

1. If all scores match, there is no team discussion.
2. If two of the three scores match, and the third score is the same polarity and only one step from the other scores, the group score is the same as the majority scores and there is no team discussion (e.g., m m s, M L L).
3. If two of the three scores match, and the third score is the same polarity but more than one step from the other scores or is 0, there is a discussion to determine the group score (e.g., l s l, S S 0)
4. If two of the three scores are 0, and the third score is not 0, there is a discussion to determine the group score (e.g., 0 0 s, M 0 0).
5. If two of the three scores match, and the third score is the opposite polarity, there is a discussion to determine the group score (e.g., m m M, s L L).
6. If all of the scores are small but one has a different polarity than the other two, the group score is 0 (e.g., s s S, s S S).
7. If none of the scores match in magnitude and polarity, there is a discussion to determine the group score (e.g., m M L, s m l).

It is important to distinguish the three types of preferences: those over fixed features, those over goals, and those over presence. For example, People care most about the physical impact of an event, while the Violent Extremists care more about the psychological impact of an event. The People do not support the goals of the Violent Extremists and the Violent Extremists do not care about the goals of the People. In general, the preferences over presence pheromones are homophilic, meaning each group likes to be with others of its own group. The Violent Extremists are attracted to the People in order to attack them, but the People are trying to stay away from them. Table 12 shows the group scores for these two groups in Ground Truth.

Table 12. Example from Ground Truth: Fragment of groups tab with groups scores.

id	Group	GroupAbbr	PrefEcon	PrefPhys	PrefPsych	PrefGoalPeo	PrefGoalVE	PrefPresencePeo
3	People	Peo	M	L	M	L	l	L
4	ViolentExtremists	VE	M	M	L	0	L	M

3.5.3 Technical Execution Parameters

A number of columns in the **groups** tab define more technical parameters that distinguish the behavior of agents in the different groups. Some of these are Boolean (True or False), for which we enter the values 1 or 0, respectively. These columns are:

- *IndependentAgency*: Do agents in this group have agency for events in a subgraph of their own in the CEG (1), or must they affiliate with another group in order to run (0)? If a group has 0 in this column, it must have 1 in the MakesAffiliation column.
- *ReceivesAffiliation*: Can agents in other groups affiliate with this group?
- *MakesAffiliation*: Can agents in this group make affiliations with other groups?
- *GhostsPerShift*: The number of ghosts that an avatar sends out per shift (see Chapter 1). In Ground Truth, we used 4.
- *ShiftsPerGen*: The number of shifts that an avatar executes before taking a step (see Chapter 1). In Ground Truth, we used 6.
- *MaxGhostSteps*: How many events ahead a ghost explores. In Ground Truth, we used 2.
- *GroupAvatars*: How many group avatars (avatars with pure group preferences, who do not affiliate with other groups) are initialized. In Ground Truth, we assigned one for each group.
- *IndividualAvatars*: How many individual avatars (avatars that can affiliate with other groups if MakesAffiliation is 1) are initialized
- *AffiliationThreshold*: An avatar affiliates with a group other than its own if the cosine distance (normalized dot product) of its preferences and the preferences of the other group exceeds this threshold. AffiliationThreshold is between 0 and 1. 0 means that the avatar will affiliate with any group yielding a positive cosine distance, while 1 means that affiliation requires perfect agreement in preferences. Low values indicate a group whose members readily affiliate with groups with differing preferences; high values indicate a group whose members are more exclusive in their affiliations.
- *PrefVariation*: A number between 0 and 1 indicating how much the preferences defined by the group vary from one member of the group to another. In creating a new member of the group, each of its preferences is computed by selecting a number uniformly at random between the group preference – PrefVariation and the group preference + PrefVariation.
- *UseHGN*: Whether agents in this group should use the group's HGN (1), or ignore it (0).

3.6 Agency

Determine agency for the CEG events. A group has agency in an event when it is eligible to participate in that event. Each event must be agented by at least one group. The agency scores should match the subgraph events with incoming or outgoing `then` edges. List the agency scores in the **events** tab of **model.xlsx**. Table 13 lists the columns, formats and definitions.

Table 13. Column, format and definition for agency.

Column	Format	Definition
Ag<Gp>	"Ag" plus GroupAbbr from the groups tab (AgGroupAbbr with no additional spaces)	Agency is eligibility to participate in an event.

3.6.1 Independent scoring

For each of the groups, score each event as having agency or not having agency

1. 1 = group has agency
2. 0 = group does not have agency
3. If an event is agented by multiple groups, the agency score is 1 for each of the groups.
3. START gets a score of 1 for every single group.

⚠ START and STOP nodes are not scored the same way as the other events.

4. If you have a STOP node for each group, the score is 1 for the group that agents that event. If you have a single STOP node for all groups, the score is 1 across all groups.

3.6.2 Team scoring

For all events, team members individually score agency and then a group score was determined as follows

1. If all scores match, no team discussion is needed.
2. If two of the three scores match, the group score is the same as the majority scores and no team discussion is needed (e.g., 0 0 1, 0 1 1).

Table 14 shows the correct format of the agency columns.

Table 14. Example from Ground Truth: Fragment of event tab with agency scores.

id	Event	AgGov	AgAO	AgRA	AgPeo	AgVE	AgMil	AgEnv
1	Government security Forces attack protesters	1	0	0	0	0	0	0
2	large numbers of people move to urban areas	0	0	0	1	0	0	0
3	government forces increase attacks on protesters	1	0	0	0	0	1	0
4	Protests increase across country	0	0	0	0	0	0	1
5	public demands democratic reforms	0	0	0	1	0	0	0
71	START	1	1	1	1	1	1	1
191	VIOLENT EXTREMISTS STOP	0	0	0	0	1	0	0

3.7 Scoring Fixed Features

Each event in the CEG has a set of fixed features to which agents respond while traversing the graph. Features are numerical characteristics of an event that define how attractive or repulsive that event is to a given group.

3.7.1 Creating a scoring rubric

Feature scores are assigned to an event and are based on agency. To facilitate the scoring process, a rubric was created. The rubric consists of scoring categories and definitions.

First, describe the impacts of the different features on the groups. Table 16 is an example from Ground Truth. The Table refers to the objectives of each group.

- The objectives of the government are to maintain political power, or territorial control, or socioeconomic influence.
- The objectives of the armed opposition are either to reform the government non-violently or, if necessary, overthrow it by force.
- The objectives of relief agencies are to provide basic needs or long-term support, or to conduct search and rescue operations.

Next, refine the descriptions of the impacts of the features for each of the scoring categories listed in the **symbols** tab of **model.xlsx**. A score of L means that the event is good for the indicated group along the indicated feature. For example, a score of L on physical for a group means that the event protects and enhances the physical well-being of that group. A score of l means that the event is bad for that feature. A score of 0 is used if a particular feature has no

Table 16. Example from Ground Truth: Examples of group-specific definitions of features used in the scoring rubric.

	Feature			
Group	**Economic**	**Physical**	**Psychological**	**Group Definition**
GOV	Resources needed to achieve objectives	Ability to achieve objectives	Belief that it can achieve objectives	A UN-recognized political state; its goal is to maintain continuity of rule as an authoritarian regime. Government security forces are part of the Government group.
AOF	Resources needed to achieve objectives	Ability to achieve objectives	Belief that it can achieve objectives	Armed, organized groups that challenge government policies. Former members of the AOF, the new government and local administrative councils are included in this group. Their goal is to have a government that recognizes civil liberties and legitimizes all sectors of the population; armed civilians are not part of the AOF.
RA	Resources needed to achieve objectives	Ability to achieve objectives	Belief that it can achieve objectives	Non-governmental organizations that provide food, water, shelter and medical aid. International organizations, such as Human Rights Watch and elections monitors, are not included in this group.

impact on a group's decision to visit an event. Table 15 shows a fragment of the scoring rubric used in Ground Truth.

Table 15. Example from Ground Truth: fragment of the scoring rubric used to score the features.

		Negative (-) Impact			
Group	**Feature**	**l**	**m**	**s**	**0**
GOV	Econ	Resources needed to achieve objectives increased by about 50%.	Resources needed to achieve objectives increased by about 25%.	Resources needed to achieve objectives increased by about 10%.	The event has almost no impact on the resources needed to achieve objectives .
	Phys	Government's ability to achieve objectives decreased by about 50%.	Government's ability to achieve objectives decreased by about 25%.	Government's ability to achieve objectives decreased by about 10%.	The event has almost no impact on the Government's ability to achieve objectives
	Psych	The Government's belief that it can achieve objectives decreased by about 50%.	The Government's belief that it can achieve objectives decreased by about 25%.	The Government's belief that it can achieve objectives decreased by about 10%.	The event has almost no impact on the Government's belief that it can achieve objectives.

Over the course of the project, if there is a qualitative change, the scales and scores should be revisited because scale definitions may need to be changed. For example, when a new group was added, the features scoring rubric was edited and the definitions were updated, based on the HGNs (Section 4). For more information on the use of rubrics in the field of evaluation, see the work of Jane Davidson [6].

3.7.2 Scoring

Table 17 lists the column, format and definition for the **events** tab of **model.xlsx**.

Table 17. Columns, format and definitions for features.

Column	Format	Definition
Feat<Gp><Ff>	"Feat" plus GroupAbbr from the groups tab and Abbr from the fixedFeatures tab (FeatGroupAbbrAbbr with no additional spaces)	Numerical characteristics of an event that define how attractive or repulsive that event is to a given group.

Features are scored for each event, but only for the group that agents that event. All other groups receive a score of 0 across all the features for that same event. See Table 21 for examples.

⚠ It is easy to lose your place when scoring the features. Highlighting which cells need to be scored and which will get a 0 will aid this process.

Here is how we highlighted the cells that needed scores: In the events tab of **model.xlsx**, there are Feat<Gp><Ff> columns for each group. Hide the columns for all but one of the groups. Also hide the Probability column. Select the Ag<Gp> column that has the same group as the Feat<Gp><Ff> columns that are visible and in the Excel Home tab, in the Styles section, go to Conditional Formatting, Highlight Cell Rules, Equal To. In the dialog box that opens, change the value of Format cells that are EQUAL TO 1. All values equal to 1 will be filled with the color listed. Then highlight those cells in the Feat<Gp><Ff> columns. Repeat for the remaining Feat<Gp><Ff> columns. Fill in all the cells that are not highlighted with 0.

Modeling team members individually score the features for a large portion of the events and then a group score is determined based on the following

1. If all scores are the same letter and the same polarity, there is no team discussion (e.g., L L L, m m m).
2. If all scores are 0, there is a team discussion.
3. If two of the three scores match, and the third score is the same polarity and only one step from the other scores, the group score is the same as the majority scores and there is no team discussion (e.g., m m s, M L L)
4. If two of the three scores match, and the third score is the same polarity but more than one step from the other scores or is 0, there is a discussion to determine the group score (e.g., l s l, S S 0)
5. If two of the three scores are 0, and the third score is not 0, there is a discussion to determine the group score (e.g., 0 0 s, M 0 0)
6. If two of the three scores match, and the third score is the opposite polarity, there is a discussion to determine the group score (e.g., m m M, s L L). However, if all of the scores are small but one has a different polarity than the other two, the group score is 0 (e.g., s s S, s S S)
7. If two of the scores are small but one has a different polarity than the other, and the third score is 0, the group score is 0 (e.g., s S 0).
8. If none of the scores match in magnitude and polarity, there is a discussion to determine the group score (e.g., m M L, s m l)

⚠ The Environment group is not scored the same way as the other groups.

1. For all Environment events, economic features are scored L, while physical and psychological features are scored 0. Again, the semantics don't matter here, because the environment is impersonal. We just need the product of one feature and its preference to be greater than 0.

⚠ START and STOP nodes are not scored the same way as the other events.

2. For START, the score is L for all features across all groups.
3. If you have a STOP node for each group, the score is L for all features for the group

with agency. If you have a single STOP node for all groups, the score is L for all features across all groups.

The above process ensures that members of the team have a common mindset for scoring. Then, for the sake of efficiency, team members individually score the features for the remaining events. The team lead reviews the scores and decides the group score on behalf of the team. The same scoring rules are applied, with the following exceptions

1. If an event is similar in nature to other events, the events receive the similar scores.
2. If the event cannot be scored by the team lead because the individual scores varied in magnitude and/or polarity, and/or there were no similar events on which to base the group score, a team discussion determines the group score.
3. The team lead then reviews all scores by row and column to ensure the scores are consistent across the features for all groups (row), as well as for each feature of an individual group (column).

Examples from Ground Truth

1. A score of L for the event "armed opposition forces gain full control over functional aspects of governance & territorial rule" means that this event protects and enhances the well-being of the Armed Opposition Forces.
2. A score of l for the event "armed opposition forces surrender to government forces" means that the Armed Opposition Forces physical well-being is negatively impacted.

Table 18 shows the correct format of `Feat<Gp><Ff>` in the **events** tab of **model.xlsx**.

Table 18. Example from Ground Truth: Fragment of events tab with features scores.

id	Event	FeatGovEcon	FeatGovPhys	FeatGovPsych	FeatAOEcon	FeatAOPhys	FeatAOPsych
1	Government security Forces attack protesters	s	S	M	0	0	0
2	large numbers of people move to urban areas	0	0	0	0	0	0
71	START	L	L	L	L	L	L
179	GOVERNMENT STOP	L	L	L	0	0	0
191	VIOLENT EXTREMISTS STOP	0	0	0	0	0	0
280	armed opposition forces surrender to government forces	0	0	0	l	l	l

3.8 Stickiness

Each time an agent (either ghost or avatar) decides to move, it chooses from a set of events. An event is sticky if an agent's current event is included in that set of events (thus allowing it to repeat that event immediately). Stickiness may make the sim more meaningful, but can lead to agents spending a long time on a single event, or even getting stuck there for the duration of the run, if this event is more attractive than any of its successors.

Table 19 lists the column, format and definition for stickiness in the **events** tab of **model.xlsx**.

Table 19. Column, format and definition of stickiness.

Column	Format	Definition
Sticky<Gp>	"Sticky" plus GroupAbbr from the groups tab (StickyGroupAbbr with no additional spaces)	An event that can be immediately repeated

For all events, team members individually score stickiness and then a group score was determined as follows:

1. Choices for stickiness scores are 1 for sticky or 0 for not sticky.
   a. Events with a `TransitTime` of 0 also have stickiness scores of 0.
   b. Events experienced by society as a whole have stickiness scores of 0.
2. If all scores match, there is no team discussion.
3. If two of the three scores match, the group score is the same as the majority scores (e.g., 0 0 1, 0 1 1).
4. For START and STOP nodes, the score is 0 across the groups.
5. Scores are entered in the **events** tab of **model.xlsx**.

⚠ START and STOP nodes are not scored the same way as the other events.

Table 20 shows the correct format of the stickiness columns.

Table 20. Fragment of events tab with stickiness scores.

id	Event	StickyGov	StickyAO	StickyRA	StickyPeo	StickyVE	StickyMil
1	Government security Forces attack protesters	1	0	0	0	0	0
2	large numbers of people move to urban areas	0	0	0	1	0	0
3	government forces increase attacks on protesters	1	0	0	0	0	0
71	START	0	0	0	0	0	0
179	GOVERNMENT STOP	0	0	0	0	0	0

3.9 Delays

Events in the real world do not happen instantly, so the modeling team assigns a notional completion time to each event, which is used to sample the actual delay from an exponential distribution. This distribution reflects the inter-event times for events that are Poisson distributed. There are two delays associated with an event: `TransitTime` and `EffectTime`.

`TransitTime` is the duration of the event, how long it takes an agent to participate in an event and then move on. It is based on what an observer on the ground would see. This means that for the event "government bombs hospitals," the transit time is not a government policy decision to implement a hospital bombing campaign as part of its war strategy, which could last for years. Instead, it is the time it took the government to physically bomb the hospitals (7 days).

`EffectTime` is how long the effects of that event persist (needed for pheromone deposits, influence edges, and stimulating the HGNs). For the event "government bombs hospitals," `EffectTime` is 1 year. The impact of losing access to hospitals lasts longer than the physical loss of the buildings.

⚠ If an event is expected to influence other events, either through an influence edge or by zipping it to an HGN, it must have a non-zero delay.

For some events, such as "people reside at IDP camp", the two delays are the same (5 years), but for other events like "government bombs hospitals," `EffectTime` (1 year) is much larger than `TransitTime` (1 week).

Table 21 lists the columns and definitions for the **delays** columns of the **events** tab of **model.xlsx**. Table 22 shows the correct format of the **delays** columns.

Table 21. Columns and definitions for delays.

Column	Definition
TransitTime	The duration of the event
EffectTime	How long the effects of that event persist

3.9.1 Team scoring

1. Each member of the team independently estimates orders of magnitude, using "D" for day, "W" for week, "M" for month and "Y" for year.
2. If the majority of the individual orders of magnitude match, use the majority score as the team score.
3. If none of the individual orders of magnitude match, discuss the event in question to decide the team score.
4. As a team, assign a numerical value to the order of magnitude.
   a. The numerical value options are
      i. 0
      ii. 1-6 for D
      iii. 1-3 for W
      iv. 1-11 for M
      v. 1- 15 for Y
   b. The numerical values must be integers.
   c. If short on time, the team lead finishes assigning the numerical values.
   d. The team lead reviews all of the scores for consistency.
   e. Convert all numerical value into days
5. For START and the STOP nodes, `TransitTimes` and `EffectTimes` are 0.

Table 22. Example from Ground Truth: Fragment of events tab with delays scores.

id	Event	TransitTime	EffectTime
1	Government security Forces attack protesters	7	30
2	large numbers of people move to urban areas	60	5475
3	government forces increase attacks on protesters	60	330
4	Protests increase across country	180	3650

⚠ START and STOP nodes are not scored the same way as the other events.

3.9.2 Calculating The Land Has Rest delays

The Environment subgraph of the CEG identifies events of nature that may influence the rest of the simulation, but are not carried out by any of the groups in the model. But sometimes we would like the environment not to do anything, so we identify a special event, "The Land has Rest" (TLHR), which requires separate delays calculations.

Table 23 lists the columns and formats for the **TLHR calculation** tab in **model.xlsx**. None of the information in this tab is consumed by the sim. All columns in this tab are optional and for the convenience of the modeling team.

Table 23. Columns and formats for TLHR.

Column	Format
#	id from the events tab
Name	Event from the events tab
env	Events agented by the Environment group
TT	TransitTime from the events tab
ET	EffectTime from the events tab
mean_median	"ENV mean"
	"ALL mean"
	"ENV median"
	"ALL median"
transitTime	ENV mean calculation
	ALL mean calculation
	ENV median calculation
	ALL median calculation
effectTime	ENV mean calculation
	ALL mean calculation
	ENV median calculation
	ALL median calculation

Calculate the delays.

1. In **model.xlsx**, copy `id`, `Event` and `AgEnv` from the events tab and paste under the `#`, `Name` and `env` columns in the **TLHR calculation** tab.
   a. Sort the columns by `env` to identify the events agented by the Environment group.
   b. Change the 1 to "yes"
2. Calculate mean `transitTime` and median `effectTime` for all of the Environment group events.
3. Calculate median `transitTime` and median `effectTime` for all events.
4. Do not include START and STOP nodes in the calculations.
5. Do not count "The Land has Rest" in the divisor of the calculations.
6. Use whichever numbers are larger.
7. Enter the values for "The Land has Rest" `transitTime` and `effectTime` in the `TransitTime` and `EffectTime` columns in the **events** tab of **model.xlsx**.
8. Add the final version of **model.xlsx** in the model folder of the model directory.

3.9.3 The problem of too many transit time scores equal to 0

`TransitTime` ages the agent by advancing its individual calendar. If more than about 5% of `TransitTimes` equal 0, the agent may get caught in a loop and its individual calendar will not advance, which means no other agents get a chance to run through the sim. The modeling team initially assigned numerical values of 0 for some `TransitTimes` because the smallest order of magnitude was not small enough. The smallest order of magnitude was weeks, but was later changed to days. If an event occurs in an instant, use the smallest non-zero order of magnitude for `TransitTime`. For example, `TransitTime` is 1 day for the event "suicide bomber kills civilians." Even though the bombing is instantaneous, there was preparation that went into the attack and travel to the site, and some of the victims may not have succumbed to their injuries immediately.

3.10 Probability

The Probability column allows the modelers to include their subjective sense of the likelihood that an agent on one event will move to another. It can be used to construct a CEG with no features and preferences, but with all transition probabilities specified by the modelers, or it can be used to adjust the probabilities computed by features and preferences.

These numbers are not strictly probabilities, since they do not necessarily sum to 1. Recall from Chapter 1 that an agent, in choosing its next event, constructs a roulette wheel over the events that it might visit next. The last step in constructing this roulette wheel, before spinning it to make a selection, is to multiply each segment (corresponding to an event) by the value specified in this column. If the probabilities specified for all the events are the same, the proportional sizes of the segments of the roulette wheel will be unchanged. However, if the probabilities differ, the relative sizes of the alternatives will be changed.

4 Hierarchical Goal Network (HGN)

Each group in SCAMP may have one or more HGNs summarizing its goals, though groups can also operate without any HGN at all.

HGNs provide an explicit representation of agents' motivations and capture goal-driven causality. They are used to compute the urgency of different events to a group in light of events that have already occurred. Agents do not move within the HGNs, the way they do through the CEG. Rather, each HGN maintains an estimate of the degree to which each goal is satisfied, based on the number of times agent visit events that are relevant to the network. The more activity on an event, the higher the satisfaction of a supported goal and the lower the satisfaction of a blocked goal. Satisfaction propagates through the HGN up to the root. Then urgency propagates back down to the marginal goals.

The goal hierarchy in the HGN continually monitors the state of satisfaction (quality) of each non-terminal goal, and deposits a priority pheromone on marginal nodes descending from that goal. An agent who has already achieved one branch of an `or` goal in its HGN has less motivation to pursue other branches, and thus their priority is reduced. An agent who has already achieved some branches of an `and` goal is motivated to achieve the others to realize the higher-level goal.

4.1 HGN Model Package Files

The HGN portion of the model package includes two files and one or two tabs in **model.xlsx**. Table 24 summarizes their contents.

Table 24. HGN model package components.

Model Package Component	Location	Purpose
HGN Cmap file	Separate file in source subdirectory	Allows modeler to construct and modify structure of the HGN
HGN xml file	Separate file in model subdirectory	Version of the HGN that Ground Truth consumes
zips	model.xlsx tab	Defines the different groups support and block zips from the CEG marginal event nodes to the HGN marginal goals.

4.2 HGN Construction

Similar to the CEG, an HGN consists of nodes and edges, but in an HGN, nodes are "goals" rather than "events". In the sample model used in Ground Truth, every group has its own HGN, except for the Environment. However, groups do not have to have an HGN. Additionally, HGNs can be turned on and off for each sim run. For example, you may want to model a situation where members of that group behave entirely tactically, with no regard for strategic goals. This is managed in the `UseHGN` column of **model.xlsx**.

Table 25 summarizes the differences between CEG events and HGN goals.

Table 25. Events versus goals.

CEG	HGN
An event describes an *action* in the world.	A goal describes a desired *state* of the world.
An event describes *what* happens.	A goal describes *why* someone wants an event to happen.
An event has a duration.	A goal does not have a duration.
Events are organized *sequentially*.	Goals are organized *hierarchically*.

4.2.1 Basic HGN Construction

Each group that has an HGN needs its own separate HGN file, created in Cmap and exported to an xml file. If a group has multiple HGNs, they all go in a single file.

1. Using CmapTools, create a new HGN by selecting File, then New Cmap, which opens a new window.
2. Each HGN file has one or more top-level goals with no outgoing edges. All other goals in the hierarchy must be part of a pathway that can access a top-level goal. Most HGN files have only a single top-level goal, but SCAMP can support multiple goal hierarchies for a single group, as long as they are all in the same Cmap file.
3. Double-click in the window to create a node. For all HGN nodes, we used rectangles, to distinguish them from nodes in the CEG.
4. Double-click on the ???? inside the node and type a unique goal id number and goal name.
5. Beneath this top-level goal, add new nodes for the subgoals and replace ???? with the subgoal id and name.
6. When the goals cannot be divided into additional subgoals, these bottom-level goals are called “marginal goals.” Think of marginal goals in relation to CEG events that need to be zipped.
7. All goals must have unique names.
8. Each node in the HGN must have only one goal.
9. Each goal must have an id number associated with it that is unique across all HGNs. The HGNs used in Ground Truth have numbers between 2000 and 3000. While you can use any unique goal id number, we recommend numbers of the form *ggsnnn*, where
   a. *gg* is the group number
   b. *s* is 1 if the goal is a leaf goal (that is, with no subgoals) and 0 if it has subgoals
   c. *nnn* is a goal number, unique within a given *gg* and *s*.
10. There should be only a single space between the goal id number and the goal label. Examples
    a. 2091 Maintain control of political process
    b. 2030 achieve free & fair elections
11. Each group has its own HGN, but all groups do not have to have an HGN.
12. To color the node, select it and then select Window, then Show Style Palette.
    a. Use the Color drop-down menu to select a color.
    b. We used yellow for the top-level goal in all of our HGNs.
    c. For the marginal goals, we used the same color for each of the groups that we used for the CEG.
13. Add edges
    a. Drag the cursor to the inside of the other nodes and then click again to create the relationship edge.
    b. In the Window menu, select Show Style Palette.

c. Select the lines of the edge on either side of the ???? and in the Style Palette, select Line then select the first type of arrow heads.
d. Label the edges by double-clicking on "????" in the middle of the arrow and typing in the name of the relation, being careful not to add any additional spaces.
   i. *and* connects two or more subgoals to a higher-level goal, indicating that all of these subgoals must be satisfied in order to satisfy the higher-level goal.
   ii. *or* connects a single subgoal to one higher-level goals. If a goal has several *or* subgoals, any one of them suffices to satisfy it.
   iii. If a goal has multiple incoming edges from multiple *and*s or *or*s, any one incoming relation can satisfy it; the edges are *or'*ed together.

⚠ A rigid goal hierarchy that contains a majority of *and* relations can result in a group never having satisfaction. However, that may be the social reality of the scenario.

14. Save file in the source subdirectory as **<group>.hgn.cmap**.

Figure 11 shows a fragment of the Government HGN from the sample model used in Ground Truth. There are three marginal goals leading to a single subgoal. Note the combination of *and* and *or* relations. Goal 2087 will be satisfied if either of two conditions takes place:

1. 2043 is satisfied.
2. Both 2034 and 2042 are satisfied.

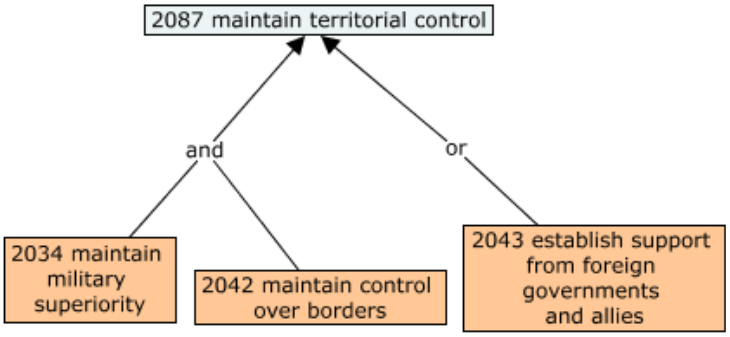

Figure 11. Example from the sample model used in Ground Truth: Fragment of Government HGN.

4.2.2 Team construction

Team members construct HGNs independently. The team meets to discuss the HGNs and then works collaboratively to construct the final HGNs.

4.2.3 Format of the HGN files

For the HGN, two file formats are needed for the model package

1. *.cmap file format
   a. This is the file format of the HGN created in CmapTools.
   b. Save the HGN file as **<group>.hgn.cmap**, where <group> is the name of the group to which the HGN belongs.
   c. Add the final version of the file to the source subdirectory.
2. *.xml file format
   a. This is the format that the SCAMP system consumes.
   b. Open **<group>.hgn.cmap**
   c. Select File, Export CMap As, CXL file
   d. Change the name and extension of the file, so that it is called **<group>.hgn.xml**. Your file system may require you to approve changing the extension.
   e. Add the final version of the file to the model subdirectory.

4.2.4 Generating lists of nodes from the HGNs

The **zips** tab in the Excel workbook **model.xlsx** requires a list of the marginal goals from the HGNs to which events in the CEG will be zipped. Here's how to generate this list.

1. Open HGN in CmapTools
2. Go to File, Export Cmap As, Propositions as text.
3. Save as a ***.txt**.
4. Separate the goal ID numbers from the goal labels.
   a. Import the data from the *.txt file into Microsoft Excel.
      i. Delete the column with the edge labels (Column B).
      ii. Cut the goals from Column A and paste them to the end of the goals in Column B.
      iii. Sort the spreadsheet by Column B to bring duplicates together, and remove all but one copy of each .
   b. Move the goals to column C, so that Columns A and B are blank
   c. To create a list with only the ID numbers, use the string function in Column A: =VALUE(LEFT(C1, SEARCH(" ",C1,1)-1)).
   d. To create a list with only the Event names, use the string function in Column B: =MID(C1,FIND(" ",C1)+1,1000)
5. You now have a list of the goal IDs and goal text labels for all of the nodes in the HGN, but you only want a list of the marginal goals.
6. Refer to the HGN and delete all of the goal IDs and goal text that are not marginal goals.
7. Copy and paste the node id numbers into the `Goal ID` column and the node labels into the `Goal Text` column of the **zips** tab of **model.xlsx**.
8. Refer back to the HGN and keep only the `Goal IDs` and `Goal Text` from the marginal goals.
9. For each `Goal Text`, add the appropriate `id` and `GroupAbbr` (from the **groups** tab) to the `id` and `Group` columns.
10. Repeat this process for the other HGNs.
11. Save **model.xlsx**.

4.3 Zipping the HGN to the CEG

Edges between events in the CEG and marginal goals are called "zips." There are 2 types of zips: `support` and `block`. A `support` zip asserts that the occurrence of the event supports the achievement of the marginal goal. A `block` zip asserts that the occurrence of the event blocks the achievement of the marginal goal.

Table 26 lists the columns, formats and definitions for the **zips** tab of **model.xlsx**.

Rules for zips

1. Goals and events do not have to have the same group.
2. Every marginal goal must have at least one zip to the CEG.
3. A single event can be zipped to multiple marginal goals of different groups.
4. A single marginal goal can be zipped to multiple events of different groups.
5. Only events with a non-zero delay can be zipped to marginal goals.
6. Not all events in the CEG need to be zipped to marginal goals in the HGNs.

Determining zips

1. Be sure the zips tab has at least the `id`, `Goal ID`, `Event ID`, and `Relation` headers from Table 26. The `Group`, `Goal Text`, and `Event Text` headers are for your convenience, and not required.

Table 26. Columns and definitions for zips.

Column	Format	Definition
id	id from the groups tab	The unique number assigned to the group.
Group	GroupAbbr from the groups tab	The abbreviation for Group. This is an optional field for the convenience of the modeling team.
Goal ID	--	The number used to identify the goal in the HGN.
Goal Text	--	The full text of the HGN goal label This is an optional field for the convenience of the modeling team.
Event ID	id from the events tab	The number used to identify the event in the CEG.
Event Text	Event from the events tab	The full text of the CEG event label This is an optional field for the convenience of the modeling team.
Relation	--	The type of relation or zip. A *support* zip asserts that the occurrence of the event supports the achievement of the marginal goal. A *block* zip asserts that the occurrence of the event blocks the achievement of the marginal goal.

2. We found it useful to copy `id` and `Event` from the **events** tab of **model.xlsx** into two blank columns at the end of the **zips** tab, so that you can consult the complete list without switching tabs. Once you complete the rest of the steps in this list, delete these columns. Note that these ids are event ids, and completely different from the `id` column
3. For each `Goal ID` and `Goal Text`, select the appropriate CEG event(s) that support(s) or block(s) its achievement, and enter them in `Event ID` and `Event Text`.
4. Add the zip type (*support* or *block*) to the `Relation` column.
5. Each goal can have more than one zip, and more than one type of zip. Add additional rows as needed for each goal to accommodate all of its zips.
6. Save **model.xlsx**.

Table 27 shows the correct format for the **zips** tab of the **model.xlsx**. Add the final version of **model.xlsx** in the model folder of the model directory.

Table 27. Example from Ground Truth: Fragment of the zips tab.

id	Group	Goal ID	Goal Text	Event ID	Event Text	Relation
1	AOF	2023	acquire & maintain adequate finances & supplies	186	foreign governments provide armed opposition forces with lethal aid	support
1	AOF	2023	acquire & maintain adequate finances & supplies	270	government disrupts armed opposition forces' supply lines	block
1	AOF	2024	develop effective working coalitions	150	armed opposition forces form coalition	support
1	AOF	2024	develop effective working coalitions	248	armed opposition forces & violent extremists unify in spite of differing ideologies	support

In experimenting with a model, you may want to turn a given zip off without deleting its row. If you replace the `id` cell with the number of a group for which no HGN exists (even a group that is not in the model), SCAMP will quietly ignore the row.

4.4 Connections among Different HGNs

Sometimes a goal in one HGN may contribute to higher-level goals in other HGNs, either for the same group or for other groups. The latest version of SCAMP supports such cross-linking with an optional tab in model.xlsx. This tab is only necessary if you want to cross-link HGNs. Otherwise it can be omitted.

The name of the tab should be **hgnLinks**.

The columns are:

- `fromGroup`: number of the group containing the lower-level subgoal.
- `fromGoal`: the unique number of the lower-level subgoal.
- `toGroup`: number of the group containing the higher-level goal.
- `toGoal`: the unique number of the higher-level goal.
- `type`: either *and* or *or*.

Each line in *hgnLinks* adds an HGN relation of the specified type from the `fromGoal` to the `toGoal`.

You are free to add other columns, such as `fromGroupName` or `toGoalName`, to help you keep track of the meaning of the connections, but the only columns that matter to the program are the five listed above.

5 Geospace

Geospace describes the physical places that our actors can be located as they move through the CEG. This perspective enables data to be generated that reflects movement decisions, temporal progress, and travel burden. It is represented as a hexagonal lattice in which agents interact with one another and with physical constraints of the environment. Because agents are always somewhere in geospace, each time they deposit presence pheromone, they do so on their current geospace hex as well as on their current event. As a result, even agents that do not move build up a strong pheromone field that can influence other agents who might be motivated to come to or avoid the location. Geospatial maps simulate movement constraints for events, to estimate the probability of success and duration of event nodes involving movement more realistically than fixed values.

5.1 Geospatial Model Files

The GEO portion of the model package includes two files and three tabs in the model workbook. Table 28 summarizes their contents.

5.2 GIMP

The geospatial map is constructed using GNU Manipulation Program (GIMP) [7].

5.2.1 Encoding geospatial data using GIMP

Every geospatial map consists of

1. The Base layer, which is a height map. A height map is a grey-scale image where black is the lowest elevation in the region and white is the highest altitude in the region.
2. The Labels layer can contain any text, icons, or other details we want to display to a human viewer. The sim will ignore this layer.

Table 28. GEO model package components.

Model Package Component	Location	Purpose
GEO ora file	Separate file in model subdirectory	Allows modeler to construct and modify structure of the CEG
GEO codebook file	Separate file in source subdirectory	Version of the CEG that SCAMP consumes
regions	model.xlsx tab	Defines the regions in GEO, including id, Name, Layer, Color, Falloff and DelayDiv.
events	model.xlsx tab	Defines destinations for GEO events.
groups	model.xlsx tab	Defines the start locations and movement delays of the different groups in the model.

3. Other layers that contain lines and regions whose semantics depends on the layer in which they occur.
   a. Features are identified by Layer name and Color name.
   b. Use the GIMP default palette (Windows/Dockable Dialogs/Palettes/Default), and use the color names that it provides.

5.2.2 Creating a height map

A height map or "difficulty map" is the basis for the geospatial map. Dark areas indicate low elevations and bright areas represent high elevations. SCAMP uses these elevations to create a gradient that affects the movement of all groups. High gradients are more difficult to traverse than more level terrain.

For Ground Truth, we used the Land Processes DAAC USGS Earth Resource Observation and Science (EROS) Center Global Data Explorer (GDEx) website [8], which has since been retired. Try https://AppEARS lpdaacsvc.cr.usgs.gov or https://search.earthdata.nasa.gov. The image we used is from the NASA Shuttle Radar Topography mission 2/11/00 to 2/21/00.

Figure 12 shows the height map used in Ground Truth. If your domain does not involve terrain of varying elevations, you can use a uniform gray layer in GIMP as your height map.

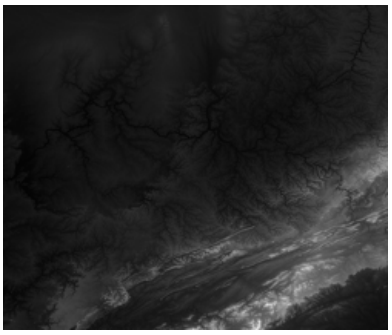

Figure 12. Sample height map from [8]

5.2.3 Importing height map into GIMP

1. Import height map into GIMP as an image file.
2. Compress the height map.
3. Adjust brightness and contrast.

5.2.4 Adding a grid overlay to image

The modeling team decided to create a map that is 400 x 400 km, based on the approximate area of Syria [9]. Adding a grid aids in approximating the size of geospatial features. We used a 20 x 20 grid to approximate the size of the different regions we added to the map. It is helpful to add a scale marker (using the grid) to the map for reference.

Note: the grid is only visible while using the program. Use Configure Grid in GIMP to edit the appearance of the grid and the spacing of the cells. Figure 13 is the height map used in Ground Truth with a grid overlay.

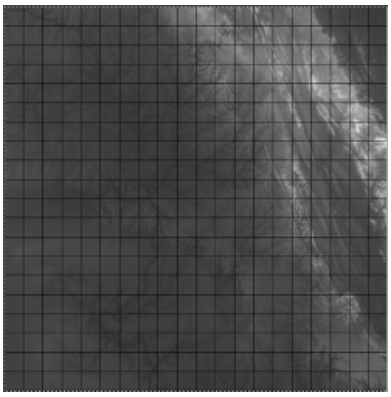

Figure 13. Height map with grid.

5.2.5 Adding a labels layer

A labels layer with the names and/or symbols for the different regions can be added to the geospatial map. This layer is not used by the sim. Figure 14 is the geospatial map developed in Ground Truth.

5.2.6 Creating geospatial regions

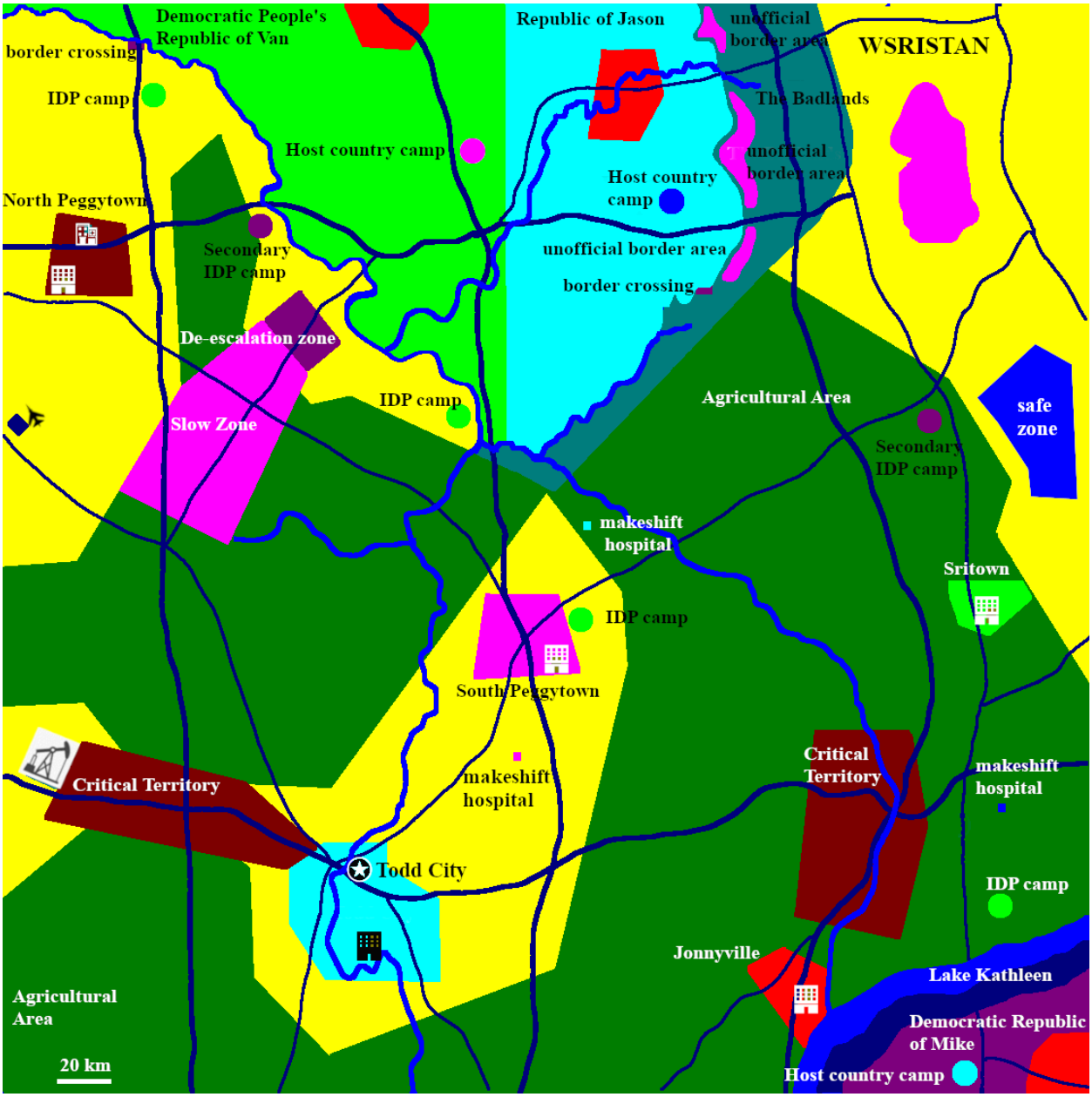

Figure 14: Geospatial map from the sample model used in Ground Truth.

The modeling team created a notional map in PowerPoint prior to adding the regions as layers in GIMP. The smallest region size is 20 km x 20 km. Instead of a grid, the sim uses a hexagon lattice because all adjacent hexes are the same distance, center to center, from each other. If a region is smaller than 20 km x 20 km, any hex that contains any part of it is considered to be part of the region. Table 29 lists the rationale for the geospatial regions in our model.

Table 29. Rationale for the geospatial regions.

Region	Number	Rationale
roads	5	Syria: Three main motorways [10] Five primary roads were added: three running north-south, two running east-west
rivers	NA	An additional branch to the river was added to create a partial water boundary between the home country and the republic of jason. Part of the river in the NE section of the map was deleted so that there was a way to cross into the badlands from the northern part of the republic of jason without crossing the river.
lake	1	lake kathleen is a water barrier between the home country and the democratic republic of mike. A boundary in the middle of the lake distinguishes which portion belongs to each country.
agricultural areas	NA	Based on the 2014 estimate of agricultural land in Syria [11]
other countries	3	North: used river as natural boundary to create two countries South: used lake as a boundary to create a country
large cities (n. peggytown, s. peggytown, jonnyville)	3	Based on the area and population of Lexington, KY [12] All large cities have roads running through them
small city (sritown)	1	Based on the area and population of Chicago, IL [13]
host country camps	3	Based on the average area of the Zaatari refugee camp [14], the Mrajeeb Al Fhood refugee camp [15], the Dadaab Refugee Complex [16] and and the Ifo Camp Profile [17]. One per host country
idp camps [18]	3	Based on the same rationale used for host country camps
secondary idp camps [18]	2	Based on the same rationale used for host country camps
border crossings [19]	2	One where a primary road crosses over a river and into another country One off the beaten path, at the border of the badlands and another country
unofficial border areas	2	Located in areas where there is no river to cross
hospitals	4	Hospital: Located in a large city (north peggytown). Three makeshift hospitals, one each for the People, Armed Opposition Forces, Violent Extremists
slow zone	1	Positioned between two large cities and adjacent to the de-escalation zone
critical territory	2	Area between the capital and the oil field Area that contains the intersection of primary roads and the river
government security forces hq	5	Located in the large cities, the small city and the capital
military airbase	1	Based on the average area of Abu al-Duhur Military Airbase, Idlib, Al-Qusayr Military Airbase, Homs and Al-Dumayr Military Airport, Damascus [20] Serves as the hospital for the Government and Military groups
de-escalation zone	1	Based on the area of the Eastern Ghouta de-escalation zone [21, 22]
oil field	1	Based on the area of oil fields in Syria [23]
the badlands	NA	Part of wsristan; area of lawlessness

5.2.7 Importing Features from Google Earth

GoogleEarth can be used as a source of features such as rivers and roads.

1. Open Google Earth Pro [24]
2. Select Add, then Image Overlay
3. On Location tab, add longitude and latitude for the four corners of the height map. See Figure 15.
4. Adjust Transparency using the slide bar, so that the image is semitransparent to allow visualization of the Google Earth image.

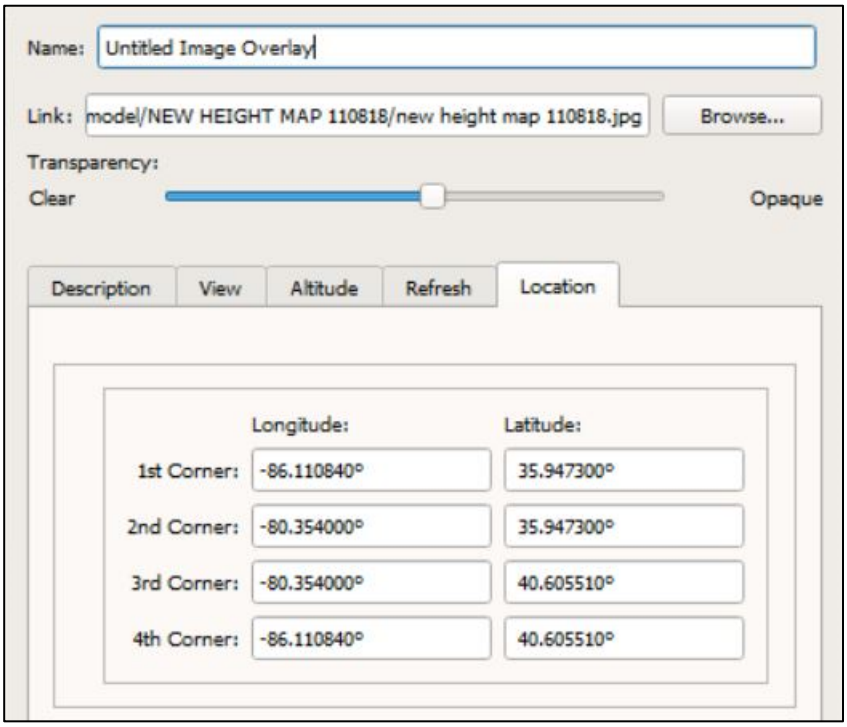

Figure 15. Google Earth Location tab in New Image Overlay pop-up menu.

5. Click OK to align height map in Google Earth.
6. On the Layers panel in Google Earth, select the desired feature, such as Roads or Water Body Outlines.
7. Zoom in or out to get the desired level of detail.
8. Save the image using the Save Image button at the top of Google Earth.
9. Select Map Options and de-select Title and Description, Legend, Scale and Compass.
10. Click Save Image.
11. Import the image into GIMP.
12. Adjust opacity, so that the height map underneath is sufficiently visible to align the image with the roads or rivers.
13. Add and name a new layer.
14. Trace the feature from the image, extending or adding to it as desired.
15. Figure 16 shows an image of roads from Google Earth and the corresponding roads layer that was added to the geospatial map.
16. Figure 17 shows an image of rivers from Google Earth and the corresponding rivers and other water features layer that was added to the geospatial map.

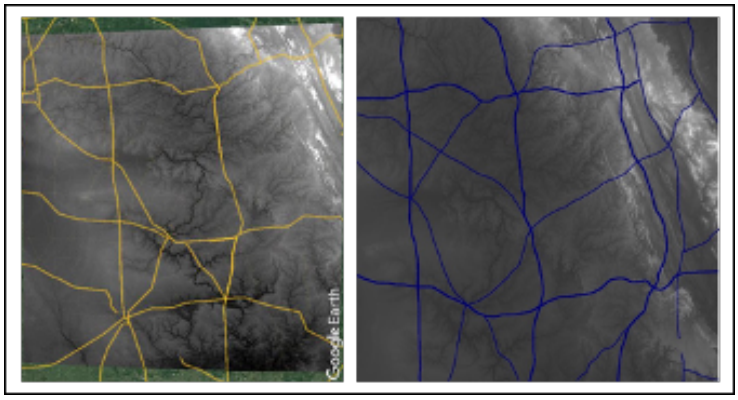

Figure 16. Google Earth image with roads (L) and height map with roads layer (R).

17. When finished, delete the image, so that only the height map and the roads layer remain.
18. Save as **map.xcf** in the source subdirectory so that it is available for human examination and modification in GIMP. Note: The xcf file can be large (~5 MB) and can take 20-30 minutes to save.
19. Save as **map.ora** in the model subdirectory.

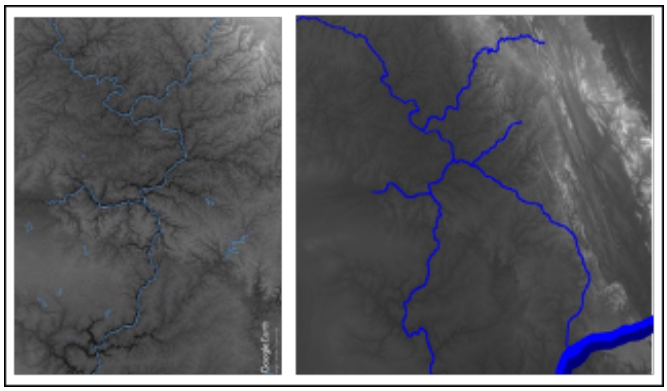

Figure 17. Google Earth image with rivers (L) and height map with rivers and other water features layer (R).

5.2.8 Combining geospatial regions

Regions can be combined by creating a new layer, duplicating the layers that contain the regions that are to be combined and then merging those layers. For example, large cities can be individual layers or can be combined into a single layer.

⚠ Any new layer must be added to the regions tab of model.xlsx and codebook.doc files.

5.3 Codebook

The codebook contains region name, layer in GIMP, color name, HTML code for color, design, and the GIMP tool used to create feature.

5.3.1 Contents of Codebook

The Codebook has three sections

1. Individual regions
2. Grouped regions
3. Labels

For **codebook.docx**, Table 30 lists the columns and definitions for individual regions, Table 31 lists the columns and definitions for grouped regions, and Table 32 lists the columns and definitions for labels.

List all regions, the corresponding layer in GIMP and the design details.

1. Keeping the same order, copy and paste the names from the layers in GIMP.
2. The spelling of the regions must be the same in all files.
3. Save in the source subdirectory as **codebook.docx.**

Table 30. Columns and definitions for individual regions.

Column	Definition
Region	Name of geospatial region
Layer	Layer in GIMP for Region
Design	Shape of Region
Color	Color from the GIMP default colors palette
Tool	Tool used in GIMP to create Region
HTML Code	HTML notation from GIMP default colors palette
Line Width (px)	Pixel size of line, if used

Table 31. Columns and definitions for grouped regions.

Column	Definition
Geo ID	R plus a unique 3-digit number assigned to the region
Region	Name of geospatial region
Layer	Layer in GIMP for Region
Design	Shape of Region
Color	Color from the GIMP default colors palette
HTML Code	HTML notation from GIMP default colors palette

Table 32. Columns and definitions for labels.

Column	Definition
Layer	Layer in GIMP for Region
Region	Name of geospatial region
Symbol/Design	Symbol/Design for Region
Label	Name of Region for Labels layer
Unicode Character	Unicode notation for Symbol/Design

Table 34. codebook.doc file format: Individual regions.

Individual Regions							
Geo ID	**Region**	**Layer**	**Design**	**Color**	**Tool**	**HTML Code**	**Line Width (px)**
R001	rivers	rivers and other water features		blue	pencil	0000ff	5
R004	primary roads	primary roads		dark blue	pencil	00007f	5
R007	critical territory	critical territory		dark red	pencil, bucket fill	7f0000	
R009	host country 1 camp	host country 1 camp		magenta	ellipse select tool, bucket fill	ff00ff	

Table 35. codebook.doc file format: grouped regions.

Grouped Regions					
Geo ID	**Regions**	**Layer**	**Design**	**Color**	**HTML Code**
R035	north peggytown, south peggytown, jonnyville	large cities in wsristan		dark red, magenta, red	7f0000, ff00ff, ff0000
R037	host country 1 camp, host country 2 camp, host country 3 camp	host country camps		magenta, blue, cyan	ff00ff, 0000ff, 00ffff

5.3.2 Format of the codebook file

Table 34, Table 35 and Table 36 show the correct format for the Individual Regions, Grouped Regions and Labels, respectively. Add `Geo ID` and `Region` to `id` and `Name` in the **regions** tab of **model.xlsx**.

Table 36. codebook.doc file format: Labels.

Labels			
Region	**Symbol/Design**	**Label**	**Unicode Character**
capital city		Capital	272A
government security forces hq		Govt security forces HQ	1F3E2
hospital		Hospital	1F3E5
military airbase		Airbase	1F6E6
oil field		Oil field	NA

5.4 Destinations

Every geospatial event defines a goal feature or destination for each group that has agency for that event. In each case, some fields are blank, because all groups do not have agency for the event in question. When multiple destinations are required, they are combined into a single region that is the overall destination. Agents choose which part of the combined destination to go to according to ease of travel, what other agents of the same group did, etc. These grouped features are added to GIMP as a new layer, and to **codebook.doc** and the **regions** tab of **model.xlsx**.

Table 33 lists the columns, formats and

Table 33. Column and definitions for destinations

Column	Format	Definition
Region	R#_plus text name of region	A named area on the geospatial map. This is an optional field for the convenience of the modeling team.
Dest	Dest<Gp> from the groups tab	A goal feature or destination for each group that has agency for a geospatial event

definitions of `Dest` in the **events** tab of **model.xlsx**.

First, decide which CEG events are geospatial. You can add columns to the events tab of **model.xlsx** to facilitate the decision-making process. We used the following criteria:

⚠ Environment events are never geospatial.

1. Does the event involve or imply movement? Keep in mind that there is a difference between location and movement. Geospatial does not mean coordinates on a map. Here are some examples:
   a. Violent extremists flee country: YES
   b. Government blocks humanitarian food distribution: YES
   c. Government refuses to negotiate with armed opposition forces: NO. The reason is that the event does not involve movement.
2. If YES, can we map it? Here are some examples:
   a. Violent Extremists flee country: YES
   b. Government blocks humanitarian food distribution: NO. The reason is that the blocking is done by local forces, which is too small of a scale for our map.

Table 37. Example from Ground Truth: Partial list of geospatial events.

Node ID	Event
1	Government security Forces attack protesters
2	large numbers of people move to urban areas
3	government forces increase attacks on protesters
16	government bombs hospitals
17	government kills large numbers of armed opposition forces
37	government arrests anti-government protesters
38	government kills & injures large numbers of protesters
45	govt & armed opposition forces kill large numbers of civilians
52	government threatens protesters
56	armed opposition forces kill large numbers of govt forces

Table 37 lists some of the geospatial events from the sample model used in Ground Truth.

Next, determine the destinations for each geospatial event, referring back to previous destination decisions to ensure that your story is consistent. Refer to `id` and `Name` in the **regions** tab of **model.xlsx**. If new regions are needed be sure to add them to the **regions** tab of **model.xlsx**, **codebook.doc**, **map.xcf** and **map.ora**.

Add a `Dest` column for each group, using the format in Table 33. For each geospatial event, add `id` and `Name` from the **regions** tab to the `Dest` and `Region` columns in the **events** tab for the group that agents that event. If an event is agented by more than one group, `id` goes into the `Dest` column for each of those groups.

Table 38 shows the correct format of the `Dest` columns.

Table 38. Example from Ground Truth: Fragment of events tab with destinations.

id	Event	Region	DestGov	DestAO
1	Government security Forces attack protesters	R1_northpeggytown_jonnyville_toddcity	R038	0
2	large numbers of people move to urban areas	R1_northpeggytown_jonnyville_toddcity	0	0
3	government forces increase attacks on protesters	R1_northpeggytown_jonnyville_toddcity	R038	0

5.5 Regions: Falloff

Falloff simulates how well known a location is. It refers to the visibility of the feature itself (something sensed through static preferences), independent of whether any agents are there. Each region has a single Falloff score that applies to all groups. Falloff is only measured for a region that is a destination.

Table 39. Columns, formats and definitions for Falloff.

Column	Format	Definition
id	--	R plus a unique 3-digit number assigned to the region.
Name	Region from events tab	A named area on the geospatial map
Layer	Layer from Codebook	The GIMP layer for the region
Color	Color from Codebook	The color assigned to the region in GIMP
Falloff	--	A measure of the visibility of a region.

Table 39 lists the columns, formats and definitions for `Falloff` in the **regions** tab of **model.xlsx**.

The values used for Falloff are 0, 0.25, 0.5, 0.75, 1.0. A feature with a Falloff of 0 has no signal strength and cannot be used as a destination, since an agent cannot see it until it is on top of it. A region with a Falloff of 1.0 has a signal that can be felt from anywhere on the map and agents in any location can easily find their way there.

If a destination is accessible to multiple groups and the groups have different Falloffs, a new region is defined based off that same feature. For example, the example "government & armed opposition forces wage fierce battle for control of critical territory," has critical territory as its destination for the Government, Armed Opposition Forces and the Military. Critical territory also is the destination for the event "violent extremists gain control of large swaths of critical territory," which is agented by the Violent Extremists. The rationale is that the Government, Armed Opposition Forces and Military are familiar with the location, while the Violent Extremists are somewhat less knowledgeable. Therefore, the Government, Armed Opposition Forces and Military have the same Falloff score (1.0), while the Violent Extremists have a slightly lower Falloff (0.75). A new region is defined for the Violent Extremists and based off this same feature, and a new row is added to the regions tab of **model.xlsx**, using the same layer. The newly defined region also is added to the `Dest` columns of the **events** tab of **model.xlsx**.

Table 40. Fragment of regions tab with falloff scores.

id	Name	Layer	Color	Falloff
R002	lake kathleen (wsristan)	rivers and other water features	blue	1
R003	lake kathleen (democratic republic of mike)	rivers and other water features	dark blue	1
R007	critical territory	critical territory	dark red	1
R008	critical territory_Violent Extremists	critical territory	dark red	0.75

Table 40 shows the correct format for `Falloff` in the **regions** tab.

5.6 Regions: DelayDiv

Unlike the height map, regions can affect the groups differently. `DelayDiv` defines how a geospatial region modifies the travel speed for agents moving through it. It is a divisor for the nominal delay experienced by the agent in moving through a tile with this region. Values less than one indicate that the region slows agents down (increasing their delay), while values greater than one indicate that the region speeds them up. `DelayDivs` do not take into consideration the context of the scenario.

`DelayDiv` is group- and region-dependent. If two regions occupy the same tile and apply to the same group, `DelayDiv` for both regions affect the agent and are proportionate to the area of the regions that originally occupied a tile. If two regions occupy different tiles and are grouped to form a new region, the grouped region has its own `DelayDiv`, which applies as an individual region. Each individual region still has its own `DelayDiv`. If a user specifies a `DelayDiv` of 0, SCAMP will assign a value of 0.1 to avoid execution errors, but an explicitly specified `DelayDiv` smaller than that but larger than 0 will be respected. For a very small `DelayDiv`, agents virtually stand still in the region.

Table 41. Columns, format and definition for DelayDivs.

Column	Format	Definition
DelayDiv	"DelayDiv" plus GroupAbbr from the groups tab (DelayDivGroupAbbr with no additional spaces)	The amount by which the delay for an agent moving through the region is divided

Table 41 lists the column, format and definition for `DelayDiv` in the **regions** tab of **model.xlsx**.

5.6.1 Team scoring

The modeling team discussed all of the `DelayDivs`. Table 42 lists the rationale for the scores of the different regions in Ground Truth.

For each geospatial region, determine the scores for all groups. Continue scoring the remaining geospatial features.

Table 43 shows the correct format for `DelayDiv` in the **regions** tab of **model.xlsx**.

Table 42. Rationale for DelayDiv scores.

Region	Rationale
roads	Same for all groups
critical territory	Relief Agencies are slower because they are security conscious People are more cautious Same for Violent Extremists and Armed Opposition Forces
host country camps	Relief Agencies—that's their environment, neither slower or faster Government, Armed Opposition Forces and Violent Extremists do not pass through
hospital	Does not change any group
slow zone	0.67 for all groups, except People, who are slower (0.5)
government security forces hq	No change for Government or Relief Agencies Slower for Armed Opposition Forces, People, Violent Extremists
military airbase	Government—ease of movement Armed Opposition Forces and Violent Extremists would not go there Relief Agencies are slower due to checkpoints People are slower than Relief Agencies
sritown	Slower for the Government Slightly faster for Armed Opposition Forces and Violent Extremists A little slower for People and Relief Agencies
idp camps	Lots of activity in that area
de-escalation zone	Ease of movement for People Slightly slower for Violent Extremists, Armed Opposition Forces and Government
oil field	There's security around it No change in speed for the Government People and Relief Agencies are slowed down a little bit due to checkpoints Violent Extremists and Armed Opposition Forces are slowed down a lot
unofficial border areas	Conflict going on there Dangerous territory
republic of jason	Sympathetic to Violent Extremists—no change in speed
wsristan	Government and Relief Agencies are unaffected People are a little slower Armed Opposition Forces and Violent Extremists are slower than People

Table 43. Example from Ground Truth: Fragment of regions tab with DelayDiv scores.

id	Name	Layer	Color	DelayDivGov	DelayDivAO
R002	lake kathleen (wsristan)	rivers and other water features	blue	0.67	0.5
R003	lake kathleen (democratic republic of mike)	rivers and other water features	dark blue	0.67	0.5
R007	critical territory	critical territory	dark red	1.33	0.67
R008	critical territory,_Violent Extremists	critical territory	dark red	1.33	0.67

5.7 Start locations

Table 44. Columns, formats and definitions for start locations.

Column	Format	Definition
StartLocation(n)	id from regions tab	The location of a group when SCAMP initializes
Feature(n)Name	Full text for StartLocation from regions tab	An optional field for the convenience of modelers
Pct(n)	--	The proportion of agents in a group at StartLocation(n)

Start Locations are the locations of the groups when SCAMP initializes, which is modified as agents participate in geospatial events. A group can have more than one start location. The modeling team assigns the start location and proportion of agents at those locations for each group. The start locations are listed in the **groups** tab of **model.xlsx**.

Table 44 lists the columns, formats and definitions for the `StartLocations` in the **groups** tab of **model.xlsx**.

There are about 10-20 agents in each group, so it is important not to spread them too thinly since it is possible that some locations will not have any agents.

1. Determine all of the start locations and percentages and enter in the appropriate columns in the **groups** tab.
2. * can be used to have agents distributed uniformly over the entire map. Note that the use of * could lead to agents in locations that are inconsistent with the events for which they have agency.
3. New regions can be created that contain “holes” for regions associated with other groups or places that do not exist at the start of the sim. Example from Ground Truth: “Holes” were made for the oil field, military airbase, and government security forces HQ, hospital, river, idp camps and de-escalation zones. Figure 18 shows an example from Ground Truth.

Table 45 shows the correct format of `startLocations`.

5.8 Movement Delays

An agent’s group not only defines where it starts, but also how fast it moves. Movement Delays are analogous to the transit time for the CEG. They define the movement speed for each group, which is the time (in days) it takes an agent to traverse one tile, which is 20 km.

The `MvmtDelay` column in the **groups** tab of **model.xlsx** is the movement speed for each group, which is the time it takes an agent to traverse one tile.

Timelines from the conflicts in Syria and Iraq served as the basis for the team movement delays scores in Ground Truth.

Table 45. Example from Ground Truth: Fragment of groups tab with start locations.

id	Group	GroupAbbr	StartLocation1	Feature1 Name	Pct1	StartLocation2	Feature2 Name	Pct2
0	Government	Gov	R017	todd city	0.3	R019	Large cities in wsristan	0.7
1	Armed Opposition Forces	AO	R046	R10_ wsristan_ with holes	1			
2	Relief Agencies	RA	R017	todd city	0.7	R006	cities in other countries	0.3
3	People	Peo	R046	R10_ wsristan_ with holes	1			

- The Government moves slightly faster than the Military.
- The Armed Opposition Forces travel at a similar speed as the Violent Extremists.
- The `MvmtDelay` for the Relief Agencies is between that of the People and the Government.
- The `MvmtDelay` for the People is based on a family fleeing Syria and walking to Serbia. [31]
    - Distance from Aleppo, Syria to Serbia-Hungarian border = 1400 mi [31]
    - Time spent walking each day: 8 h
    - Distance walked each day: 25 mi = 40 km
        - 40 km/24 h = 20 km/ x h
        - 0.5 day to travel 20 km

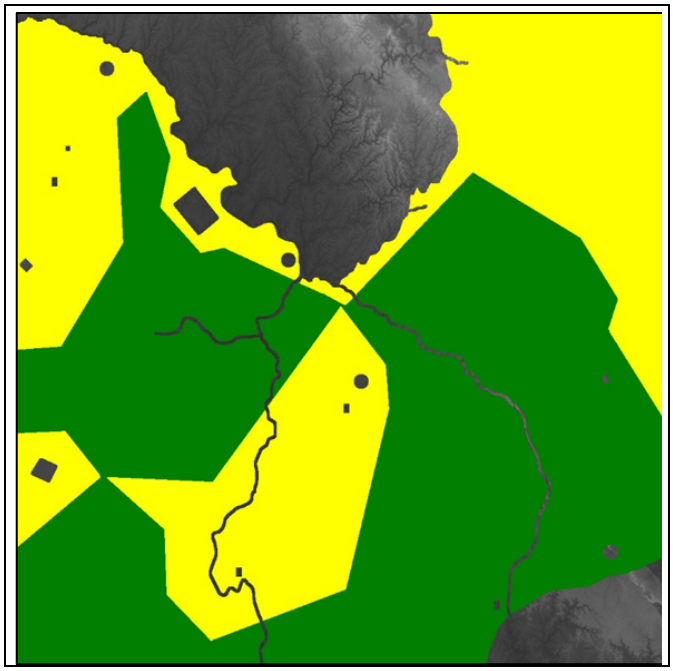

Figure 18. Example from Ground Truth: Geospatial map with "holes."

- The `MvmtDelay` for the Violent Extremists is based on the 2014 Northern Iraq offensive [32, 33]
    - June 2014: ISIL attacked Mosul
    - 11 June 2014: 5 days later, ISIL advanced to Baiji
    - Distance from Mosul to Baiji = 175 km [34]
        - 35 km/24 h = 20 km/ x h
        - 0.6 day to travel 20 km
    - 11 June 2014: ISIL took full control of Tikrit
    - Distance from Mosul to Tikrit = 231 km [35]
        - 46 km/24 h = 20 km/ x h
        - 0.4 day to travel 20 km
    - The two times were averaged to get 0.5 d to travel 20 km.
- The `MvmtDelay` for the Military is based on the 2018 Southern Syria offensive [28, 29]

- 26 June 2018: the Syrian Army took control of Busra al-Harir
- 27 June 2018: the Syrian army arrived at the eastern outskirts of Al-Hirak
- Distance from Busra al-Harir to Al-Hirak = 12 km [30]
  - 12 km/24 h = 20 km/x h
  - 0.4 d to travel 20 km

- There is no `MvmtDelay` for the Environment.

⚠ The Environment group is not scored the same way as the other groups.

Table 46 shows the correct format of `MvmtDelays` in the **groups** tab of **model.xlsx.** Be sure the final version of **model.xlsx** is in SCAMP/data/<modelName>/model.

Table 46. Example form Ground Truth: Movement delays.

id	Group	GroupAbbr	MvmtDelay
0	Government	Gov	0.3
1	ArmedOppositionForces	AO	0.5
2	ReliefAgencies	RA	0.4
3	People	Peo	0.5
4	Violent Extremists	VE	0.5
5	Military	Mil	0.4

6 Social Networks (SN)

The SN portion of the model package includes two tabs in the model workbook. Table 47 summarizes their contents:

Table 47. Social network model package components.

Model Package Component	Location	Purpose
groupChanges	model.xlsx tab	Defines the conditions required for a group change to be triggered.
namedAgents	model.xlsx tab	Defines the named agents associated with the group changes, as well as the group changes themselves.

6.1 Group changes

There are five types of group changes

1. Birth
2. Death
3. Change in location
4. Change in agency
5. Suspend/resume

For each domain day that an event has participation and the event's condition is satisfied, the transition fires with the specified probability.

⚠ Long TransitTimes can result in a large number of births, which could overload the sim.

Changes in birth and death utilize a null group called "Guf" [25], the repository of souls in Jewish mysticism. Guf is both a group and an event. For births, the transition is from Guf to <group> and Guf to an event in the CEG. When an agent is born, it has a brand-new set of preferences, based on its group, and a history containing only the event where it originated.

⚠ Long TransitTimes can result in a large number of deaths, which may or may not be the intent of a particular group change.

For deaths, the transition is from <group> and location in the CEG or region in geospace to Guf (group) and Guf (location or region). When an agent dies, its identity is obliterated, and it no longer chooses events. Guf does not require a CEG subgraph or an HGN.

An agent can be picked up from one region in geospace and moved to different region.

⚠ An agent that is not in geospace will not drop into geospace due to an agency change

When an agent changes agency, its old preferences and affiliations vanish, and are replaced by a new set. It will either be moving in geospace or CEG space when that happens. It may later switch into geospace, but that would be because of a normal decision with its new agency. Changes in agency cannot occur when an agent drops down into geospace.

For suspensions, the transition is from a group's location in the CEG to Limbo. Limbo is equivalent to a CEG Event. Suspension is like death in that an agent stops choosing events, but it retains its identity, including its preferences and history. If there is Limbo, there also must be resume. A group change rule must be created to transition the agent from Limbo back to an event in the CEG.

Table 48 lists the columns and definitions used in the **groupChanges** tab of **model.xlsx**.

Table 48. Columns and definitions for group changes.

Column	Format	Definition
order	Google Coin Toss [26] to intersperse rows for two groups Random Sequence Generator [27] to intersperse rows for more than two groups	The sequence used for the opposing rows to keep the populations of opposing groups as constant as possible over the course of several rows. This is an optional field for the convenience of the modeling team.
Trigger	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) where group changes can be initiated
TriggerText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for Trigger This is an optional field for the convenience of the modeling team.
FromGroup	id from the groups tab of model.xlsx	The group(s) from which agents can be drawn or a list of avatar ids
FromGroupText	Group from the groups tab of model.xlsx	Full text for FromGroup This is an optional field for the convenience of the modeling team.
FromLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the agents are drawn
FromLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for FromLoc This is an optional field for the convenience of the modeling team.
FromLimit	A positive integer value or * (meaning no limit)	The maximum number of potentially affected agents
ToGroup	id from the groups tab of model.xlsx	The group(s) to which an agent will be switched
ToGroupText	Group from the groups tab of model.xlsx	Full text for ToGroup This is an optional field for the convenience of the modeling team.
ToLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) to which an agent can be moved
ToLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for ToLoc This is an optional field for the convenience of the modeling team.

Column	Format	Definition
ConditionalGroup	id from the groups tab of model.xlsx	The group(s) that are required to be present for any agents to be affected
ConditionalGroupText	Group from the groups tab of model.xlsx	Full text for ConditionalGroup This is an optional field for the convenience of the modeling team.
ConditionalLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) where Conditional Groups have to be located for any agents to be affected
ConditionalLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for ConditionalLoc This is an optional field for the convenience of the modeling team.
OLDProb	A value from 0.0 to 1.0	The base Probability used in the prior sim run This is an optional field for the convenience of the modeling team.
Prob	A value from 0.0 to 1.0	The base Probability required for an agent to be affected
Promoters	id from the groups tab of model.xlsx	The group(s) that increase the likelihood that the group change will occur
PromotersText	Group from the groups tab of model.xlsx	Full text for Promoters This is an optional field for the convenience of the modeling team.
PromoterLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the Promoters are drawn
PromoterLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for PromoterLoc This is an optional field for the convenience of the modeling team.
Blockers	id from the groups tab of model.xlsx	The group(s) that decrease the likelihood that the group change will occur
BlockersText	Group from the groups tab of model.xlsx	Full text for Blockers This is an optional field for the convenience of the modeling team.
BlockerLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the Blockers are drawn
BlockerLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for BlockerLoc This is an optional field for the convenience of the modeling team.

Table 49 lists the symbols that are used in the **groupChanges** tab.

Table 49. Symbols used for group changes.

Symbol	Uses
*	Generally, this means all values. For From Loc or To Loc, this means all possible locations, both in CEG and geospace, though this may be limited to one or the other. For groups, this means all defined groups.
=	Generally, this means "the same". For From Loc, this means use the same locations as the triggering event or hex grid to which this particular change is attached. For To Loc, this means keep the same location. For To Group, this means keep the same group.
S	This is shorthand for the START event, which is assumed to be alone, with no other events or regions. If you need to use the start event as part of a group, list it by its id.
Gu	This short for Guf. [25] Gu is used for birth or death events. As the From Group value, this indicates a birth event. As the To Group or To Loc value, this is a death event. Gu is a group AND an event.
-	The dash/minus sign is equivalent to Gu, but is meant to be used as a filler to say that promoters or blockers aren't to be used.
A#	Things that start with A in From Groups are assumed to be agents, which are only found in From Groups.
L#	Locations starting with L are assumed to be Limbos. If a limbo event doesn't exist, it will be created at startup. Note that the only way to get an agent out of a limbo event is by creating a group change rule. Limbo is equivalent to a CEG Event. If Limbo is the From Loc, a CEG Event must be used as the To Loc.
E#	The id of a CEG event. E followed by the node id.
R#	The id of a geospatial region. R followed by the id.

Rules for determining group changes:

1. For CEG events that trigger group changes
   a. Copy and paste `id` and `Event` from the **events** tab into the **groupChanges** tab under the columns `Trigger` and `TriggerText`.
   b. Review the list and delete the events that do not trigger group changes.
2. For regions that trigger group changes.
   a. Copy and paste `id` and `Name` from the regions tab into the **groupChanges** tab under the columns `Trigger` and `TriggerText`.
   b. Review the list and delete the regions that do not trigger group changes.
3. Rules for `Trigger`
   a. `Trigger` does not have to be the same as the `FromLoc`.
   b. `Trigger` may not occur if the population of agents is sparse.

c. Use the agent history log file to find which regions are likely to be stepped on by agents and then use those for `Trigger`.

4. Determine `FromGroup`, `FromGroupText`, `ToGroup` and `ToGroupText`
   a. A mix of group numbers and agents is not allowed.
      i. Named agents can be used as the `FromGroup`.
      ii. Named agents cannot be used in any other fields.
   b. Agency changes cannot occur if `FromLoc` is a geospatial event.
   c. A `single` Trigger can have multiple rows. Add additional rows, as appropriate.
5. Determine `FromLocation` and `FromLocText`
   a. `FromLoc` can be either a CEG event or geospatial region.
   b. `FromLoc` type must match `ToLoc` type
   c. `FromLoc` cannot be a geospatial event for agency changes.
   d. If `FromLoc` is a region, make sure it is not a large area because a large number of agents could be impacted.
6. Determine `ToLoc` and `ToLocText`
   a. This field has to match the type in `FromLoc`, either CEG event or geospatial region.
   b. If multiple events or regions are listed, one will be chosen randomly.
   c. `Trigger` can be the same as `ToLoc`, but keep in mind that this agent could affect the next time `Trigger` is called.
   d. For example, if a birth event is promoted by agents of group A on `Trigger`, and when triggered it creates an agent of group A on `Trigger` (it being the `ToLoc`), then the new agent could potentially encourage more births pretty quickly.
   e. If `ToLoc` is within `FromLoc`, the agent may already be at `ToLoc` when the group change is implemented.
   f. Figure 19 shows the hospital in North Peggytown. If a group change is initiated to move an agent from the hospital to North Peggytown, that agent will not move because it is already at its destination.

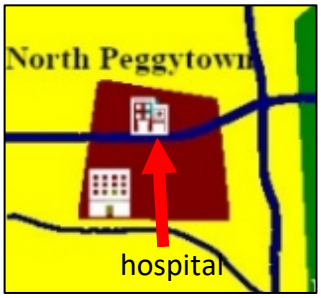

Figure 19. The hospital is in North Peggytown.

7. Determine `FromLimit`
   a. This cutoff is applied before the actual probability is run.
   b. This is either a positive integer value or * (meaning no limit).
   c. This field is required.
8. Determine `ConditionalGroup` and `ConditionalGroupText`
   a. Do not use =
   b. Agents of `Conditional Groups` must be participating in `ConditionalLoc` at the same time to enable the transition.
   c. The presence of these groups has a binary effect. Their level of participation does not matter. It is just about whether they are there or not.
   d. Nodes in the CEG are strictly speaking abstract event *types*, and only become events when one or more agents participate in them. This means that transitions can only take place on days that an event has some participating agents. So even if `ConditionalGroup` is empty, the transition only fires when *some* agent is there.

   e. Rules for `ConditionalLoc` and `ConditionalLocText`. `ConditionalLoc` is required. If this field is empty, no agents will ever be affected.
9. Determine `Prob`
   a. A value from 0.0 to 1.0
   b. Only a single agent is necessary at `FromLoc` (assuming other conditions are met, like `FromGroup`, `ConditionalGroup`, etc.), for it to be affected. Assuming a single potentially affected agent with a probability of 0.5, then it has a 50% change to be changed, equivalent to a coin toss; heads, it's moved/killed/etc., or tails, it is unchanged.
   c. If you need a lot of conditional groups on a rarely visited event, which would almost never happen, then a high probably could be called for, since just the fact the that preconditions were met is a pretty low probability circumstance.
   d. `Prob` will be modified if there are `Promoters` or `Blockers`.
10. Determine `Promoters` and `PromotersText`
   a. `Promoters` and `Blockers` act within the same row to oppose each other.
   b. If equal numbers of `Promoters` and `Blockers` are on a row, they cancel each other out, leaving the probability unchanged.
   c. When there are agents of the given group on `PromoterLoc`, the probability that agents from `FromGroup` will be affected increases, the more promoting agents, the greater the increase.
   d. The higher the participation of `Promoters`, the more likely the group change is to take place. If there are no `Promoters`, enter -.
11. Determine `PromoterLoc` and `PromoterLocText`. If there are no `Promoters`, then enter –
12. Determine Blockers and BlockersText
   a. Promoters and Blockers act within the same row to oppose each other. If equal numbers of Promoters and Blockers are on a row, they cancel each other out, leaving the probability unchanged.
   b. When there are agents of the given group on BlockerLoc, the probability that agents from the FromGroup will be affected decreases, the more blocking agents means the lower the probability.
   c. If there are no Blockers, enter –
13. Determine BlockerLoc and BlockerLocText. If there are no Blockers, enter -
3. If there are several rows for a single trigger, the group changes are implemented in order. Each row runs sequentially. Outcomes of previous rows can affect subsequent rows. For example, if there are 10 Group A agents on the first row, and 1 gets moved to a new location, the next row will be run against 9 Group A agents. When the subsequent rows that affect Group B run, the population of Group A has been potentially decreased twice.
   a. Intersperse opposing rows to keep the populations of opposing groups as constant as possible over the course of several rows.
   b. Intersperse death and injury rows.

Table 50 lists the data formats allowed in the **groupChanges** tab.

Table 50. Acceptable data formats for the group changes columns.

Field	Acceptable Data Format		Field	Acceptable Data Format
Trigger	* E52 R10 S		ConditionalLoc	= * E52 R10 S
FromGroup	0 2 blank = this event (list of) event IDs = all agents of correct group on those events * (list of) Axxx agent IDs = named agents		Prob	0.0 0.75 1.0
FromLoc	* E52 R10 = Gu S		Promoters	* 0 3 -
FromLimit	20 *		PromoterLoc	* = E52 R10 S -
ToGroup	5 = Gu		Blockers	* 0 3 -
ToLoc	* E52 R10 = Gu S		BlockerLoc	* = E52 R10 S -
ConditionalGroup	1 3 *			

6.2 Named agents

All agent names are generated automatically when the program initializes, but some aspects of the SN perspective will require the modeling team to name agents in advance (so they can be assigned to roles in a group, for example, or specified as targets in assassination events). This requires four steps:

1. Review the events in the **groupChanges** tab to determine which require named agents.
2. Generate usable agent names.
3. Add them the **groupchanges** tab for the appropriate events.
4. Instruct the sim to load agent names

Table 51 lists the columns and definitions for the **namedAgents** tab of **model.xlsx.**

Table 51. Columns and definitions for named agents.

Column	Definition
Name?	The individual agent name
Leader?	Group leader
Influencer?	Group influencer
NumBackup	The number of backups needed for the agent in case it dies or moves to another group
Notes	Descriptive information about the agent This is an optional field for the convenience of the modeling team.

Agent names must follow a certain pattern. An agent name in SCAMP consists of the letter 'A', followed by a string of digits. These digits fall into three groups, so the name looks like Aaaabc, where aaa is a number that may have more than one digit, while b and c are single digits.

1. Review `Triggers` in the **groupChanges** tab to determine which events require named agents. Highlighting them is helpful.
2. Create the named agents.
   a. Decide if the agent represents a group or an individual.
   b. Decide if the agent can affiliate with groups other than its home group.
   c. This is the "c" using the Aaaabc format. See Table 52.
   d. Select the group number from `id` in the **groups** tab of **model.xlsx** for named agent. This is "b" in the format Aaaabc. See Table 53.
   e. Select a value for "aaa" to distinguish multiples of the type of agent created in such a way that no other agent has the same overall name. Use 8 or 9 for any of the digits in the "aaa" format.

Table 52. Agent type and affiliatability numbers.

Last Digit	Group or Individual?	Can affiliate or not?
0	Group	no
1	Group	yes
2	Individual	no
3	Individual	yes

Table 53. Example from Ground Truth: Group numbers.

id	Group
0	Government
1	Armed Opposition Forces
2	Relief Agencies
3	People
4	Violent Extremists
5	Military
6	Environment

Note: Since no automatically generated agent ever has an 8 or 9 in its name, it is guaranteed not to conflict.

1. For a single agent with a given group and affiliation preference, use 8 or 9. Example: the head of state could be defined as A902.
2. For a group of influential agents, use multiple digits, as long as one of them is an 8 or 9. Example: a group of influential people in the Armed Opposition Forces (group 1) could be defined as A9012, A9112, A9212, and A9312.
3. Instruct the sim to load agent names.
   a. Enter named agent under `Name`.
   b. Decide if the agent is a Leader and/or Influencer in its group.
      i. All group Leaders are Influencers, but since there can be only one Leader and there can be many Influencers, most Influencers will probably not be Leaders.
      ii. Enter the appropriate number under `Leader?` and `Influencer?`
         1. Use 0 for no
         2. Use 1 for yes
   c. Decide how many backups need to be maintained for the agent in case it dies or moves to another group. Enter that number in `NumBackups` column.
   d. Add descriptive information about the agent under `Notes`.

Table 54 shows the correct format of the **namedAgents** tab in **model.xlsx**.

Table 54. Example from Ground Truth: Fragment of named agents tab.

Name	Leader?	Influencer?	NumBackups	Notes
A9100	0	1	4	government officials
A902	1	1	3	head of state/govt
A9003	0	0	1	police officer

4. Add the named agents as `FromGroups` for the appropriate events in the **groupchanges** tab of **model.xlsx**.

Add the final version of **model.xlsx** in the model folder of the model directory.

7 Defaults

This chapter describes a huge number of variables that can contribute to the performance of a SCAMP model, more than 20,000 in our current model of civil conflict.

A note on terminology:

- A *variable* is a cell in a worksheet.
- A *column* is a column in worksheet, which contains many variables, one for each row.
- A *column type* defines a set of columns that all contain variables of the same type.

While these variables provide a rich modeling language, it is unrealistic to expect modelers to define them all before getting some idea for how the model behaves. This section describes a mechanism by which a modeler can get a model running by defining only 22 default values that cover types of columns in the model workbook, then refining these into a larger set of up to 61 defaults for individual columns, and finally refining a subset of the remaining specific variables. The objective is to support an incremental approach to model construction.

This chapter describes the structure of the Excel workbook that contains these variables. Each sheet has a number of *entities* (identified in the rows), each with a number of *variables* (in the columns). In most cases the columns are grouped into *types*. For example, in the **events** tab, the entities are events. In our current model, each event has

- 21 columns of type Features with names like FeatGovEcon or FeatPeoPhys
- One Probability column
- Seven columns of type Agencies with names like AgGov and AgVE
- A TransitTime and EffectTime
- Seven columns of type Destinations with names like DestAO and DestRA
- Seven columns of type Stickiness with names like StickyGov or StickyVE

It also has an initial id variable, and three text fields for the convenience of the modeler: Event (name), and Region. Thus there are 21 + 1 + 7 + 2 + 7 + 7 = 45 columns to define for each event, but these columns are of six distinct types.

The defaults mechanism works in two stages.

1. The **defaults** tab of the worksheet defines a single default value for each column *type* that supports defaults. There are 22 such types.
2. Each tab that supports defaults has a final defaults row that can specify a default value for a single *column*. As we have seen, there are 45 possible defaults to specify on the **events** tab, and others on other tabs.

Defaults are used to supply missing values. Here's how.

1. When the program loads a row from the tab (e.g., an event from the events tab), it looks first in that row for the value of each variable.
2. If a value of one variable is missing, it looks at the defaults row at the bottom of the spreadsheet. In some cases, it copies this value into the variable; in others, it does a

computation based on the default. We will explain shortly what happens in the case of each type of column.

3. If the variable's default is missing, the program then looks at the value assigned to the column's *type* in the defaults tab, and uses that as the default.

The key idea is that defaults supply missing values. If the user specifies a value for a variable, that value overrides any default defined for that variable. If a given column has a default, that overrides the variable type defaults in the **defaults** tab. The values in the **defaults** tab come into play only when the model is missing both a specific variable value, and a column default for that value. Please note that "missing" is not the same as "0". "0" is a value. A missing field has nothing in it.

Here's an example, again from the events tab. The Agency column type includes the seven columns `AgGov`, `AgAO`, `AgRA`, `AgPeo`, `AgVE`, `AgMil`, and `AgEnv`. For each event, these columns define which groups have agency for that event. (Recall from Chapter 1 that a group has agency for an event if its agents can choose to participate that event.) Table 55 is a fragment of the **events** tab for a given event.

Table 55. Example from Ground Truth: events tab.

id	AgGov	AgAO	AgRA	AgPeo	AgVE	AgMil	AgEnv
23	0			0	0		0
defaults	1	0					

In addition, but not shown here, the **defaults** tab says that the overall default for events: Agencies is 1.

Three of the variables are missing. What values will the Agencies variables have when the program runs?

- The modeler has specified `AgGov` as 0. So the program uses 0, even though the default for that variable is 1.
- `AgAO` is missing, so the program looks at the defaults row, finds the variable default 0, and sets `AgAO` to 0.
- `AgRA` is missing, but so is the default for that variable in the defaults row. So the program looks for the default for the variable *type* Agencies in the defaults tab, finds 1, and uses that. The same process assigns 1 to `AgMil`.
- The modeler has set `AgPeo` to 0, so the program uses that. The fact that the variable default is missing and the variable type default is 1 doesn't matter. Any explicit entry is chosen over defaults. Similarly, `AgVE` and `AgEnv` remain at 0, as set by the modeler.

With this mechanism, once a modeler has defined the defaults tab, the row ids, and the variable names for each tab that supports defaults, the model can run.

The remainder of this section discusses each of the tabs, its columns, and the column types to which they belong. It specifies how defaults are processed for each of the seventeen column types, and gives the initial values that we have specified in the defaults tab for each one. Unless otherwise specified, the default value is simply copied to supply missing variables, drawing first from the column defaults, and if these are missing, from the **defaults** tab. See Chapter 2 for detailed definition of each column.

7.1 Tabs with no defaults

Some tabs do not support defaults. These tabs are of two types.

The first type defines abbreviations used elsewhere, and includes the tabs **symbols** and **fixedFeatures**.

The second type contain optional data. These tabs are not necessary for the model to run, but if they are included, the modeler must define them. These tabs include **groupChanges**, **namedAgents**, and **zips**.

In addition, some columns in the other tabs must be defined by the modeler. A common example is the id column. The modeler must tell SCAMP what the events, groups, and regions are.

7.2 events tab

This tab includes five types of columns that have defaults: `Features`, `Probability`, `Agencies`, `TransitTime`, `EffectTime`, and `Stickiness`.

`Features:` Feature columns have names with three parts: the letters "Feat", an abbreviation for a group (from the **groups** tab, `GroupAbbr` column), and an abbreviation for a fixed features (from the **fixedFeatures** tab, column `Abbr`). Feature column can be assigned either a number in [-1, 1] or a symbol from [26 s, m, l], indicating large positive, medium positive, small positive, small negative, medium negative, and large negative values, assigned from the **symbols** tab.

`Probability:` There is only a single column of this type.

`Agencies:` Agency columns have names with two parts: the letters "Ag" and an abbreviation for a group (from the **groups** tab, `GroupAbbr` column).

`TransitTime:` There is only a single column of this type. The value is taken as the parameter of an exponential distribution, which is sampled for the actual value. If a default is given, it is sampled separately for each row.

`EffectTime:` Again, there is only a single column. If a value is present (either in the column, or in the column default, or in the overall defaults), it is taken as the parameter of an exponential distribution, which is sampled for each missing value. If the value is missing entirely, the system will use whatever parameter was given for `TransitTime` for the event in question (which may itself have come from the defaults). Again, it samples the distribution separately for each row.

`Destinations:` There is one column of this type for each group, named "Dest" followed by a group identifier. No default is available. If a cell is empty, that indicates that agents of the designated group do not treat that event as a geospatial event.

`Stickiness:` There is one column of this type for each group, named "Sticky" followed by `GroupAbbr`.

7.3 groups tab

`Preferences:` There are three kinds of preference columns. There is one Preference column for each fixed feature defined in the **fixedFeatures** tab, named "Pref" plus a feature abbreviation from that tab. There is one preference for the goal status (urgency) of each group, with names of

the form `PrefGoalGov` and `PrefGoalMil`. There is one preference for the presence pheromones of each group, with names of the form `PrefPresencePeo`.

`StartLocation`: A single column, specifying where new agents of this group should begin their lives. Defaults can specify only a region within which new agents are to be placed randomly, not features within a region, as can be done with the full spreadsheet. The region `Rany` is available to specify that agents should be randomly scattered over the entire map.

`MvmtDelay`: A single number reflecting the relative speed of movement of agents of each group.

The following columns were originally in the **agents.csv**, but are now included in the **groups** tab of the workbook. **agents.csv** is still supported to enable experimenters to vary some of these parameters (e.g., `NumIndividualAvatars`) without recompiling the entire model.

`IndependentAgency`: A single column indicating whether (1) or not (0) agents of this group can have agency without affiliating with another group; setting this variable to 0 allows definition of neutral agents.

`ReceivesAffiliation`: A single column indicating whether (1) or not (0) agents can affiliate with this group.

`MakesAffiliation`: A single column indicating whether (1) or not (0) agents of this group can affiliate with other groups.

`NumGroupAvatars`: A single column giving the number of avatars representing the group as a whole (typically 1).

`NumIndividualAvatars`: A single column giving the number of individuals to be generated initially for the group.

`AffiliationThreshold`: A single column indicating how close the preference vector of an agent in this group must be to the group preference of another group to form an affiliation. 0 indicates promiscuous affiliation; 1 means that affiliation is impossible.

`PrefVariation`: A single column indicating how much the preferences of individual agents in this group can vary from the group's overall preference.

`UseHGN`: A single column indicating whether agents in this group use the group's HGN. This column, together with the **agents.csv** file (see Chapter 3), allow experimenters to turn an individual group's HGN on and off without removing the HGN from the model.

`GhostsPerShift`: A single column indicating the number of ghosts that an avatar sends out in each shift.

`ShiftsPerGen`: A single column indicating the number of shifts of ghosts that an avatar sends out before making a decision based on their feedback.

`MaxGhostSteps`: A single column indicating the number of steps into the future that an avatar's ghosts explore.

7.4 regions tab

The modeler must define region ids and colors. Defaultable column types are:

`Falloff:` A single column specifying how far away the region is visible.

`DelayDivisors:` One column for each group

Add the final version of **model.xlsx** in the model folder of the model directory.

8 Model Directory

The different model file types go in different folders in the model directory. The contents of the model folder are consumed by the sim. The naming conventions and file formats must be as listed, or the sim will not run.

<modelName>

geo (constructed by the SCAMP simulator)

- map.xcf

model

- CEG.xml
- <group>.hgn.xml
- map.ora
- model.xlsx

source

- <modelName>.cmap
- <group>.hgn.cmap
- <modelName>.xcf
- codebook.doc

Acknowledgements

The development of SCAMP was funded by the Defense Advanced Research Projects Agency (DARPA), under Cooperative Agreement HR00111820003. The content of this paper does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

1. *CmapTools Knowledge Modeling Kit*. Florida Institute for Human and Machine Cognition.
2. *CmapTools FAQs*. 2020 [cited 2020; Available from: https://cmap.ihmc.us/docs/cmaptools-faqs.
3. Kahneman, D. and A. Tversky, *The simulation heuristic*. 1981, Stanford Univ CA Dept of Psychology.
4. *The Bible*. New International Version. Vol. Joshua 11.
5. *Mapping Militant Organizations: Hay'at Tahrir al-Sham (Formerly Jabhat al-Nusra)*. 2018 June 2018 [cited 2018; Available from: http://web.stanford.edu/group/mappingmilitants/cgi-bin/groups/view/493#basic-info.
6. Davidson, E.J., *Evaluative Reasoning*, in *Methodological Briefs: Impact Evaluation 4*. 2014, UNICEF Office of Research: Florence.
7. Kimball, S. and P. Mattis, *GNU Image Manipulation Program (GIMP)*.
8. *Land Processes Distributed Active Archive Center (DAAC), USGS Earth Resource Observation and Science (EROS) Center, Global Data Explorer (GDEx)*. 2018 07/10/2018 [cited 2018; Available from: https://gdex.cr.usgs.gov/gdex/.
9. *The World Factbook: Syria*. 2018 August 20, 2018 8/24/18]; Available from: https://www.cia.gov/library/publications/the-world-factbook/geos/sy.html.

10. Monterie, M. *Road numbering systems: Syria*. 6/05/19]; Available from: https://sites.google.com/site/roadnumberingsystems/home/countries/syria.
11. *Syria - Agricultural land (% of land area)*. 2019 6/05/19]; Available from: https://tradingeconomics.com/syria/agricultural-land-percent-of-land-area-wb-data.html.
12. *QuickFacts: Lexington-Fayette, Kentucky*. 6/05/19]; QuickFacts provides statistics for all states and counties, and for cities and towns with a population of 5,000 or more.]. Available from: https://www.census.gov/quickfacts/fact/table/lexingtonfayettekentucky/LND110210.
13. *QuickFacts: Chicago city, Illinois*. 6/05/19]; QuickFacts provides statistics for all states and counties, and for cities and towns with a population of 5,000 or more.]. Available from: https://www.census.gov/quickfacts/fact/table/chicagocityillinois/LND110210.
14. *UNHCR Factsheet-Zaatari Refugee Camp*. 2018 6/05/19]; Factsheet]. Available from: https://data2.unhcr.org/en/documents/download/62004.
15. *UAE-funded Camp for Refugees Opens in Jordan*. Gulf News, 2013.
16. International, G., *The Energy Situation in the Dadaab Refugee Camps, Kenya*, S. Okello, Editor. 2016: London.
17. *Ifo Camp Profile*. 2015; Camp Snapshot]. Available from: https://reliefweb.int/sites/reliefweb.int/files/resources/IfoCampProfile_DadaabKenyaAugust2015.pdf.
18. *Syria Dynamic Monitoring Report*. 2017, Dynamo Information Management Unit. p. 118.
19. *Turkey-Syria Border Crossing Status*. 2018, United Nations Office for the Coordination of Humanitarian Affairs.
20. *Syrian Airfields*. 2018 4/15/18 6/06/19]; Available from: https://www.globalsecurity.org/military/world/syria/airfields.htm.
21. Rollins, T. *The unravelling of Syria's Eastern Ghouta*. Al Jazeera, 2016.
22. *Final De-escalation Zones Agreed on in Astana*. Al Jazeera, 2017.
23. Solomon, E., R. Kwong, and S. Bernard *Inside Isis Inc: The journey of a barrel of oil*. Financial Times, 2016.
24. *GoogleEarth Pro*. 2019.
25. Scholem, G.G., *Origins of the Kabbalah*. 1990: Princeton University Press.
26. *Google Coin Toss*. [cited 2019; Available from: https://www.google.com/search?q=flip+a+coin.
27. *Random Sequence Generator*. 2019 6/05/19]; Available from: https://www.random.org/sequences/?min=1&max=118&col=1&format=html&rnd=new.

Appendix: Tabs, columns, format and definitions for model.xlsx

Tab	Column	Format	Definition
event	id	A positive integer [1,1000].	The id number used to identify the event in the CEG
	Event		The full text of the CEG event label
	Feat<Gp><Ff>	"Feat" plus GroupAbbr from the groups tab and Abbr from the fixedFeatures tab	Numerical characteristics of an event that define how attractive or repulsive that event is to a given group.
	Probability		
	Ag<Gp>	"Ag" plus GroupAbbr from the groups tab	Agency is eligibility to participate in an event.
	TransitTime		The duration of the event
	EffectTime		How long the effects of that event persist
	Region	R#_plus text name of region	A named area on the geospatial map

Tab	Column	Format	Definition
	Dest<Gp>	Dest plus GroupAbbr from the groups tab	A goal feature or destination for each group that has agency for a geospatial event
	Sticky	"Sticky" plus GroupAbbr from the groups tab	An event that can be immediately repeated
symbols	symbol	Uppercase and lowercase letters, and 0	Scoring categories
	value	A positive or negative integer with up to 1 decimal place	The nominal values associated with the symbols
fixedFeatures	id	A positive integer [0,10]	The id number of the feature
	Feat	Uppercase and/or lowercase letters, as well as numbers, but it should not include punctuation or spaces.	The full feature name
	Abbr	Uppercase and/or lowercase letters, as well as numbers, but it should not include punctuation or spaces.	The abbreviation for the feature.
groupChanges	Order	Google Coin Toss [26] to intersperse rows for two groups Random Sequence Generator [27] to intersperse rows for more than two groups	The sequence of opposing rows to keep the populations of opposing groups as constant as possible over the course of several rows
	Trigger	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) where group changes can be initiated
	TriggerText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for Trigger This is an optional field for the convenience of the modeling team.
	FromGroup	id from the groups tab of model.xlsx	The group(s) from which agents can be drawn or a list of avatar ids
	FromGroupText	Group from the groups tab of model.xlsx	Full text for FromGroup This is an optional field for the convenience of the modeling team.
	FromLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the agents are drawn
	FromLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for FromLoc This is an optional field for the convenience of the modeling team
	FromLimit	A positive integer value or * (meaning no limit)	The maximum number of potentially affected agents
	ToGroup	id from the groups tab of model.xlsx	The group(s) to which an agent will be switched
	ToGroupText	Group from the groups tab of model.xlsx	Full text for ToGroup This is an optional field for the convenience of the modeling team.

Tab	Column	Format	Definition
	ToLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) to which an agent can be moved
	ToLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for ToLoc This is an optional field for the convenience of the modeling team.
	ConditionalGroup	id from the groups tab of model.xlsx	The group(s) that are required to be present for any agents to be affected
	ConditionalGroupText	Group from the groups tab of model.xlsx	Full text for ConditionalGroup This is an optional field for the convenience of the modeling team.
	ConditionalLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) where Conditional Groups have to be located for any agents to be affected
	ConditionalLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for ConditionalLoc This is an optional field for the convenience of the modeling team.
	OLDProb	A positive integer [0,1]	The base Probability used in the prior sim run
	Prob	A positive integer [0,1]	The base Probability required for an agent to be affected
	Promoters	id from the groups tab of model.xlsx	The group(s) that increase the likelihood that the group change will occur
	PromotersText	Group from the groups tab of model.xlsx	Full text for Promoters This is an optional field for the convenience of the modeling team.
	PromoterLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the Promoters are drawn
	PromoterLocText	Event from the events tab of model.xlsx Name from the regions tab of model.xlsx	Full text for PromoterLoc This is an optional field for the convenience of the modeling team.
	Blockers	id from the groups tab of model.xlsx	The group(s) that decrease the likelihood that the group change will occur
	BlockersText	Group from the groups tab of model.xlsx	Full text for Blockers This is an optional field for the convenience of the modeling team.
	BlockerLoc	id from the events tab of model.xlsx id from the regions tab of model.xlsx	The location(s) from which the Blockers are drawn
	BlockerLocText	Event from the events tab of model.xlsx	Full text for BlockerLoc

Tab	Column	Format	Definition
		Name from the regions tab of model.xlsx	This is an optional field for the convenience of the modeling team.
groups	id	A positive integer [0,10].	The unique number assigned to the group.
	Group	Uppercase and/or lowercase letters, as well as numbers, but it should not include punctuation or spaces.	The full group name
	GroupAbbr	Uppercase and/or lowercase letters, as well as numbers, but it should not include punctuation or spaces.	The abbreviation for Group.
	Pref<Ff>	"Pref" plus abbr from the fixedFeatures tab	A measure of how much a group cares about that feature
	PrefGoal<Gp>	"PrefGoal" plus GroupAbbr from the groups tab	A measure of a group's attitude towards achieving its goal and how much each group cares about other groups achieving their goals
	PrefPresence<Gp>	"PrefPresence" plus GroupAbbr from the groups tab	A measure of homophily or how much a group wants to be around other groups.
	StartLocation1	id from regions tab	The location of a group when SCAMP initializes
	Feature1Name	Name from regions tab	
	Pct1	A positive integer with up to 1 decimal place	The proportion of agents in a group at StartLocation(n)
	StartLocation2	id from regions tab	The location of a group when SCAMP initializes
	Feature2Name	Name from regions tab	
	Pct2	A positive integer with up to 1 decimal place	The proportion of agents in a group at StartLocation(n)
	StartLocation3	id from regions tab	The location of a group when SCAMP initializes
	Feature3Name	Name from regions tab	
	Pct3	A positive integer with up to 1 decimal place	The proportion of agents in a group at StartLocation(n)
	MvmtDelay		The movement speed for each group, which is the time it takes an agent to traverse one tile.
	IndependentAgency		
	ReceivesAffiliation		
	MakesAffiliation		
	GhostsPerShift		
	ShiftsPerGen		
	MaxGhostSteps		
	GroupAvatars		
	IndividualAvatars		
	AffiliationThreshold		
	PrefVariation		
	UseHGN		
defaults	type	See Section 7	
	ColumnType		

Tab	Column	Format	Definition
	Default		
namedAgents	Name		The individual agent name
	Leader?		Group leader
	Influencer?		Group influencer
	NumBackups		The number of backups needed for the agent in case it dies or moves to another group
	Notes		Descriptive information about the agent
regions	id		R plus a unique 3-digit number assigned to the region.
	Name	String	A named area on the geospatial map
	Layer	Layer from Codebook	The GIMP layer for the region
	Color	Color from Codebook	The color assigned to the region in GIMP
	Falloff		A measure of the visibility of a region
	DelayDiv	"DelayDiv" plus GroupAbbr from the groups tab	The amount by which the delay for an agent moving through the region is divided
zips	id	id from the groups tab	The unique number assigned to the group.
	Group	GroupAbbr from the groups tab	The abbreviation for Group.
	Goal ID	A positive integer [2000,3000].	The number used to identify the goal in the HGN.
	Goal Text		The full text of the HGN goal label
	Event ID	id from the events tab	The number used to identify the event in the CEG.
	Event Text	Event from the events tab	The full text of the CEG event label
	Relation		The type of relation or zip. A *support* zip asserts that the occurrence of the event supports the achievement of the marginal goal. A *block* zip asserts that the occurrence of the event blocks the achievement of the marginal goal.
TLHR calculation	#	id from the events tab	The id number used to identify the event in the CEG
	Name	Event from the events tab	The full text of the CEG event label
	env		Events agented by the Environment group
	TT	TransitTime from the events tab	The duration of the event

Tab	Column	Format	Definition
	ET	EffectTime from the events tab	How long the effects of that event persist
	mean_median	"ENV mean", "ALL mean", "ENV median", "ALL median"	Mean or median for Environment events Mean or median for all events
	transitTime	ENV mean calculation, ALL mean calculation, ENV median calculation, ALL median calculation	Mean or median TransitTime
	effectTime	ENV mean calculation, ALL mean calculation, ENV median calculation, ALL median calculation	Mean or median EffectTime