Mapping Interstate Territorial Conflict (v. 2):
Codebook and Supplementary Information

Kenneth A. Schultz
Stanford University

Version date: June 25, 2019
This document consists of four parts. Part I describes the files associated with this data release and the variables contained in each. Part II discusses how the area measurements in the state-year and dyad-year data files were constructed and provides descriptive statistics of those files. Part III provides information on changes made from the version 1 release. Part IV provides bibliographic information for the map sources described in the dispute spreadsheet.

I. Files and Variables

This data release consists of five of files:

1. MITC_disputed_regions_v2.shp (and associated files)

This consists of the shapefile containing the disputed regions. Each region is associated with a variable, area_id, that links it to data on territorial disputes. Each region also has a variable, homeland, indicating whether the area was ever part of a dispute involving homeland territory for both states.

2. MITC spreadsheet v2.csv

This spreadsheet set contains information on the territorial disputes identified by Huth and Allee (2002) that can be merged with the spatial data using the area_id variable. Each row consists of a challenger-target pair in a dispute. The following variables are coded:

- dispute_id: A 5-digit identifier for each dispute. It takes the form of RDDDSS, where R=region_id, DDD=disno, and SS=disno_sub. Note that not all observations have unique values of dispute_id if a given dispute is
inherited by a successor dyad (e.g., due to decolonization or state merger).

dispute_id_connect A 5-digit identifier that connects sub-disputes within the same dispute. It is needed for cases in which dispute_id changes within a given dispute due to changes in the scope of the claim. It takes the same form as dispute_id and equals the minimum dispute_id among connected observations.¹

area_id A 5-digit variable that links each dispute to the associated areas in the shapefile. The area_id equals the dispute_id, except when the same area is implicated in more than one dispute, in which case the area_id equals the minimum dispute_id.

region_id Indicates the geographic region of the world in which the dispute took place (1=Africa, 2=Americas, 3=Asia, 4=Europe, 5=Middle East/Northern Africa)

disno The dispute number from the Huth and Allee (2002) data

dinso_sub A sub-dispute number, if relevant. Set to zero if there are no sub-disputes.

period This is a counter for disputes that go through more than one time period. Set to zero if there is only one period.

¹ For example, the claim by Saudi Arabia against Britain/United Arab Emirates has three sub-disputes (506801, 506802, 506803) due to changes in the size of the claim over time; all three receive dispute_id_connect=506801. Not all observations with sub-disputes are connected this way if the sub-disputes arose from having divided a Huth and Allee dispute into separate geographic regions. For example, the Chinese claims on India were broken into three separate sub-disputes (301201, 301202, 301203) because they involve geographically distinct areas; as a result, they have different values of dispute_id_connect. In other words, this identifier is for connecting observations across time, not across challenger-target pairs. The latter can be done using the country codes or the original Huth and Allee dispute numbers (disno).
Challenger The name of the challenger state
ccode_c The Correlates of War (COW) country code of the challenger state.
Target The name of the target state
ccode_t The COW country code of the target state.
start The first year of the dispute. This comes from Huth and Allee (2002) except where changes are noted.
end The year that the dispute ended. This is coded as 8888 if the dispute was ongoing as of 2000.
resolution A variable indicating whether the dispute was resolved at the end of the period and, if so, how:
    0 = dispute not resolved
    1 = challenger dropped claim due to agreement, ruling, or unilateral action
    2 = challenger dropped claim due to loss in war
    3 = target conceded all or part of the disputed territory
    4 = challenger seized the territory
reciprocal Indicates whether or not the dispute was reciprocal in the sense that both states were coded as the challenger over the same piece of territory.
recip_dispute_id In the case of reciprocal disputes, the provides the dispute_id of the other directed dyad in the dispute.
dispute_type Indicates the type of territory involved in the dispute:
    1=Homeland territory for both states
    2=Colonial/dependent territory for both states
3=Homeland territory for the challenger, colonial/dependent territory for the target

4= Homeland territory for the target, colonial/dependent territory for the challenger

islands Indicates that the dispute was solely over off-shore islands

precision The precision code, ranging from 1-5, as discussed in the article.

incomplete Indicates that some areas of disagreement were not mapped.

Territory Brief name of territory

Description A description of the area in dispute. Departures from the Huth and Allee (2002) treatment are noted here, along with other complications in coding

Source of map data A description of the geospatial data used to make the map, including sources when necessary. The bibliography for these sources is located at the end of this document.

3. MITC_state_year.dta

This data set contains annual data for each state involved in a dispute over homeland territory, 1947-2000, with three measures of the amount of areas in dispute. All areas are measured in squared kilometers. Details on these different area measures and how they were calculated follow in section II.

code COW country code

year year
**tot_disputed_area**  
Total area of all disputed regions involving this state. NOTE: This total might be less than the sum of the two variables below because a state can be a challenger and target over the same region. This variable does not double count such regions.

**tot_challenge_area**  
Total area of disputed regions in the state is coded as a challenger

**tot_target_area**  
Total area of disputed regions in the state is coded as a target

**claimed_area**  
Total area of the state’s claimed homeland (see below)

**effective_area**  
The state’s total effective area (see below)

**disputed_homeland**  
The share of the total claimed homeland that is in dispute

**target_share**  
The share of the state’s effective area that is target in disputes

**challenger_share**  
The area of the state’s challenges as a share of its effective area

---

4. MITC_dyad_year.dta

This data set contains annual data for each dyad involved in a dispute over homeland territory, 1947-2000.

**ccode1**  
COW country code of the state with the lower country code

**ccode2**  
COW country code of the state with the higher country code

**dyad**  
A six digit combination of the two country codes

**year**  
year

**dispute_per_dyad**  
Percent of dyadic area in dispute
dispute_per_targetPercent of the target state(s) area in dispute. In the event that both states are coded as targets, the share of each state’s area in dispute was calculated, and the two shares were summed.

dispute_per_lengPercent of the dyadic border by length in dispute. Note that this variable is missing if the states do not share a land border or if they are contiguous only through an overseas enclave (e.g., Britain and Ireland due to Northern Ireland).

island_onlyIndicates that the dispute in the dyad involved only off-shore islands

5. MITC_ddyad_year.dta

This data set contains annual data for each directed dyad (challenger-target pair) involved in a dispute over homeland territory, 1947-2000.

ccode_cCOW country code of the challenger

ccode_tCOW country code of the target

dyadA six digit identifier of the dyad, combining ccode_c and ccode_t with the lower value appearing first

ddyadA six digit identifier of the directed dyad, combining ccode_c and ccode_t with order preserved

yearyear

dispute_per_targetPercent of the target state’s effective area in dispute

island_onlyIndicates that the challenger’s claim involved only off-shore islands
II. Measuring Dispute Areas

For each state in each year, I calculated three quantities that reflect the amount of homeland territory that was disputed. The first measures the total area in dispute as a percentage of the state’s total claimed area, where total claimed area combines both regions that that the state controls and regions that it claims but does not control. Thus, we can think of this quantity, disputed homeland, as measuring how much of the state’s claimed homeland was contested, with the claimed homeland corresponding to the territory it would have if all its claims were met.

Second, I calculated the area of territory subject to claims by other states as a percentage of the state’s total effective area. As described below, total effective area captures how much (homeland) territory was in the state’s de facto control in each year. Thus, target share captures the extent to which area under the state’s control was targeted by other states, and it can be thought of as a measure of territorial threat. Finally, challenger share measures the total area of territory that the state was demanding from others as a share of it total effective area. This variables captures the state’s territorial ambitions: i.e., how much it desires relative to what it already has. Note that, while disputed homeland and target share are logically constrained to fall between zero and one, challenger share is not, since a state can issue a challenge for more territory than it currently possesses. At the dyadic and directed dyadic level, the data report the share of the effective dyadic territory and the share of the target state’s effective area that was in dispute, respectively. Descriptive statistics for all these variables can be found in Tables A1-A3.

With one exception, areas of disputed regions were calculated based on the mapped polygons. Because much of the land border dispute between China and Vietnam could not be mapped, I substituted in the widely reported figure of 227 square kilometers in dispute. The calculation of undisputed area was based on the country sizes in the DoS map. Since this map
reflects 2013/14 borders, polygons for several states that no longer exist (e.g., North and South Vietnam, North and South Yemen, the Soviet Union, East and West Germany, Yugoslavia, and Ethiopia and Sudan prior to their breakups) had to be created by the author. In addition, corrections were made for states that experienced large boundary changes (e.g., Jordan prior to its renunciation of the West Bank). Even so, a challenge arises due to the fact that the borders in the DoS base map do not necessarily the de facto or de jure borders of the state in earlier years. Today’s borders may include areas that were formerly not part of the state but became so due to the resolution of a claim. Similarly, a dispute that leads to a territorial loss means that today’s borders do not include territory that once did fall under the state’s control. Thus, the challenge is to create a time-varying measure of area under a state’s control from a map of current borders and the location and timing of territorial disputes.

The approach taken here leverages the coding of challengers and targets in disputes. By Huth and Allee’s (2002) coding rules, a state is considered the challenger if it lays claim to territory that is not under its control, and a state is a target whenever there is a claim to territory not in its control. This makes it possible to identify (a) areas that are now part of the state but which were not at an earlier time and (b) areas that used to be part of the state but which are not now. The former are regions that overlap with the current state territory but over which the state was coded as a challenger in an earlier dispute; the latter are regions that do not overlap with the state’s current territory but over which the state was coded as targeted in earlier dispute. The only remaining complication are areas in which the status quo at the time of the dispute was not established, so the state was coded as both the challenger and target over those areas. These areas are not counted as being under the state’s control at the time of the dispute.
Figure A1 should help clarify. Let the square region labeled A denote the boundaries of state A as depicted in the current map. Let the oval shaped regions labeled C and T denote regions over which state A was coded as the challenger and target, respectively, in time $t$. Let the region labeled CT denote a region for which state A was coded as both a challenger and a target at time $t$—that is, a region that was subject to a reciprocal dispute where there was no clear delineation of the status quo. We can then define two quantities. The first is the total claimed area of the state in time $t$. This quantity captures all of the area that the state controls and all area that it claims but does not control, shown in grey in panel (a). Formally,

$$\text{Total Claimed Area} = \text{Area}(A \cup T \cup C \cup CT).$$

The second quantity is the area effectively under A’s control in time $t$, shown in grey in panel (b). This quantity captures the current area plus the area of regions in which it was coded as target minus the area overlapping regions in which it was coded as challenger or in which control was unclear. Formally,

$$\text{Effective Area} = \text{Area}(A \cup T) - \text{Area}(A \cap C) - \text{Area}(A \cap CT) = \text{Total Claimed Area} - \text{Area}(C \cup CT).$$

When calculating area shares at the dyadic levels, the effective areas of the two states were summed. The area of any territory that was involved in a reciprocal dispute between the states—and thus under neither’s control but still part of the dyadic territory—was added to this sum.
Figure A1. Calculating the Claimed and Effective Area of States

(a) Claimed Area
(b) Effective Area

A = Area of state on current base map
T = Area over which state is target in dispute
C = Area over which state is challenger in a dispute
CT = Area over which state is both challenger and target
## Table A1. Descriptive Statistics of State-Year data (MITC_state_year.dta)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Stand. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccode</td>
<td>3869</td>
<td>469.2</td>
<td>258.3</td>
<td>2</td>
<td>935</td>
</tr>
<tr>
<td>year</td>
<td>3869</td>
<td>1976.1</td>
<td>14.9</td>
<td>1947</td>
<td>2000</td>
</tr>
<tr>
<td>tot_disputed_area</td>
<td>3869</td>
<td>68038.2</td>
<td>198841.5</td>
<td>.023</td>
<td>2502293</td>
</tr>
<tr>
<td>tot_challenge_area</td>
<td>3869</td>
<td>41834.6</td>
<td>179029.3</td>
<td>0</td>
<td>2502293</td>
</tr>
<tr>
<td>tot_target_area</td>
<td>3869</td>
<td>27951.3</td>
<td>83350.4</td>
<td>0</td>
<td>1079422</td>
</tr>
<tr>
<td>claimed_area</td>
<td>3869</td>
<td>1320687</td>
<td>3104246</td>
<td>484.5</td>
<td>2.22e+07</td>
</tr>
<tr>
<td>effective_area</td>
<td>3869</td>
<td>1278852</td>
<td>3097510</td>
<td>483.4</td>
<td>2.22e+07</td>
</tr>
<tr>
<td>disputed_homedland</td>
<td>3869</td>
<td>.13</td>
<td>.23</td>
<td>3.30e-08</td>
<td>1</td>
</tr>
<tr>
<td>target_share</td>
<td>3869</td>
<td>.076</td>
<td>.21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>challenge_share</td>
<td>3869</td>
<td>.11</td>
<td>.37</td>
<td>0</td>
<td>4.68</td>
</tr>
</tbody>
</table>

## Table A2. Descriptive Statistics of Dyad-Year data (MITC_dyad_year.dta)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Stand. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccode1</td>
<td>2633</td>
<td>505.5</td>
<td>238.2</td>
<td>2</td>
<td>900</td>
</tr>
<tr>
<td>ccode2</td>
<td>2633</td>
<td>554.8</td>
<td>242.8</td>
<td>20</td>
<td>910</td>
</tr>
<tr>
<td>dyad</td>
<td>2633</td>
<td>506044.7</td>
<td>238390.7</td>
<td>2020</td>
<td>900910</td>
</tr>
<tr>
<td>year</td>
<td>2633</td>
<td>1975.5</td>
<td>14.8</td>
<td>1947</td>
<td>2000</td>
</tr>
<tr>
<td>dispute_per_dyad</td>
<td>2633</td>
<td>.038</td>
<td>.098</td>
<td>1.70e-08</td>
<td>.71</td>
</tr>
<tr>
<td>dispute_per_target</td>
<td>2633</td>
<td>.11</td>
<td>.24</td>
<td>3.30e-08</td>
<td>1.14</td>
</tr>
<tr>
<td>dispute_per_leng</td>
<td>2102</td>
<td>.38</td>
<td>.38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>island_only</td>
<td>2633</td>
<td>.19</td>
<td>.39</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

## Table A3. Descriptive Statistics of Directed Dyad-Year data (MITC_ddyad_year.dta)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Stand. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ccode_c</td>
<td>3263</td>
<td>546.5</td>
<td>237.5</td>
<td>2</td>
<td>910</td>
</tr>
<tr>
<td>ccode_t</td>
<td>3263</td>
<td>532.9</td>
<td>240.3</td>
<td>2</td>
<td>900</td>
</tr>
<tr>
<td>dyad</td>
<td>3263</td>
<td>517743.3</td>
<td>235669.8</td>
<td>2020</td>
<td>900910</td>
</tr>
<tr>
<td>ddyad</td>
<td>3263</td>
<td>547061.5</td>
<td>237675.5</td>
<td>2020</td>
<td>910900</td>
</tr>
<tr>
<td>year</td>
<td>3263</td>
<td>1974.7</td>
<td>15</td>
<td>1947</td>
<td>2000</td>
</tr>
<tr>
<td>island_only</td>
<td>3263</td>
<td>.18</td>
<td>.39</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>dispute_per_target</td>
<td>3263</td>
<td>.088</td>
<td>.22</td>
<td>3.30e-08</td>
<td>1</td>
</tr>
</tbody>
</table>
### III. Changes from Version 1

<table>
<thead>
<tr>
<th>Dispute ID</th>
<th>Challenger</th>
<th>Target</th>
<th>Territory</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>104400</td>
<td>Somalia</td>
<td>Britain</td>
<td>Kenya’s Northern Frontier District</td>
<td>Map revised to reflect the fact that Somalia’s claim on Kenya pre-independence corresponded to the former Northern Frontier District (NFD). Precision code revised to 2.</td>
</tr>
<tr>
<td>104400</td>
<td>Somalia</td>
<td>Kenya</td>
<td>Somali region of Kenya (Somalia-Kenya)</td>
<td>Upon independence, Kenya reorganized the NFD, and the Somali-inhabited region became the North Eastern Province. Somalia did not recognize the reorganization and based its claim on self-determination for the Somali people. Thus, post-independence, the disputed area corresponds to the ethnic settlement area. Precision code revised to 3.</td>
</tr>
<tr>
<td>301900</td>
<td>China</td>
<td>Kazakhstan</td>
<td>China-Kazakhstan border</td>
<td>Maps revised based on additional information and sources. Precision code revised to 3 to reflect uncertainty/incompleteness.</td>
</tr>
<tr>
<td>302000</td>
<td>China</td>
<td>Kyrgyzstan</td>
<td>China-Kyrgyzstan border</td>
<td>Maps revised based on additional information and sources. Precision code revised to 3 to reflect uncertainty/incompleteness.</td>
</tr>
<tr>
<td>302600</td>
<td>China</td>
<td>Tajikistan</td>
<td>China-Tajikistan border</td>
<td>Maps revised based on additional information and sources. Precision code revised to 3 to reflect uncertainty/incompleteness.</td>
</tr>
<tr>
<td>302401</td>
<td>China</td>
<td>Soviet Union</td>
<td>China-USSR border in Central Asia</td>
<td>Maps revised based on additional information and sources. Precision code revised to 3 to reflect uncertainty/incompleteness.</td>
</tr>
<tr>
<td>302200</td>
<td>China</td>
<td>Pakistan</td>
<td>China-Pakistan border</td>
<td>Map revised based on additional information about the claims and new sources. Precision code revised to 2.5.</td>
</tr>
<tr>
<td>303000</td>
<td>India</td>
<td>China</td>
<td>Aksai Chin (post-1963)</td>
<td>Map revised based on new sources. Precision code revised to 1.</td>
</tr>
<tr>
<td>303201</td>
<td>India</td>
<td>Pakistan</td>
<td>Pakistani-controlled Kashmir plus Siachen</td>
<td>Precision code revised to 1 based on more accurate map of India claim on Shaksgam valley.</td>
</tr>
<tr>
<td>Code</td>
<td>Country 1</td>
<td>Country 2</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>303202</td>
<td>India</td>
<td>Pakistan</td>
<td>Pakistani-controlled Kashmir plus Siachen glacier (India-Pakistan, 1949-63) This sub-dispute now terminates in 1984 to reflect Indian seizure of Siachen glacier (thus removing this regions from India’s challenge).</td>
<td></td>
</tr>
<tr>
<td>303203</td>
<td>India</td>
<td>Pakistan</td>
<td>Pakistani-controlled Kashmir without Siachen glacier (India-Pakistan, post-84) New sub-dispute created to reflect Indian seizure of Siachen glacier in 1984.</td>
<td></td>
</tr>
<tr>
<td>30130</td>
<td>China</td>
<td>Burma</td>
<td>China-Burma border Maps were revised based on additional information and new sources.</td>
<td></td>
</tr>
<tr>
<td>501901</td>
<td>Egypt</td>
<td>Sudan</td>
<td>Hala’ib Triangle and Wadi Halfa Salient (Egypt-Sudan) Original dispute 501900 was split into three sub-disputes to reflect the fact that Egypt seized the Hala’ib Triangle in 1995. Thus, Egypt’s challenge over the Triangle ends then, and its only remaining challenge is over the Wadi Halfa Salient. After 1995, Sudan becomes the challenger over the Triangle.</td>
<td></td>
</tr>
<tr>
<td>501902</td>
<td>Egypt</td>
<td>Sudan</td>
<td>Wadi Halfa Salient This new sub-dispute covers the Egyptian claim to the Wad Halfa Salient after 1995.</td>
<td></td>
</tr>
<tr>
<td>501903</td>
<td>Sudan</td>
<td>Egypt</td>
<td>Hala’ib Triangle This new sub-dispute covers Sudan’s claim to the Hala’ib Triangle after 1995.</td>
<td></td>
</tr>
</tbody>
</table>
IV. Additional References for Map Data


