

UCDP GED Conflict Polygons Dataset

Codebook

Version 1.1

In force as of 15 March 2011

Mihai Croicu, Uppsala University, UCDP (mihai.croicu@pcr.uu.se)

Ralph Sundberg, Uppsala University, UCDP (ralph.sundberg@pcr.uu.se)

Uppsala Conflict Data Program (UCDP)

Georeferenced Event Dataset Project

www.ucdp.uu.se

This version of the codebook corresponds to the UCDP GED Conflict Polygons dataset version 1.1-2011

Table of Contents

- To cite this dataset:..... 3
- 1. Introduction and purpose 3
- 2. Definition of the UCDP GED polygon 4
- 3. Description of UCDP Polygons 6
 - 3.1. Data used for constructing the UCDP GED polygons 6
 - 3.2. Choice of a Convex Hull..... 8
 - 3.3. Outliers and Special Cases 8
 - 3.4. Construction and specifications of the polygons..... 10
 - 3.4.1. Padding of the polygons 11
 - 3.4.2. Geographical assumptions..... 11
 - 3.4.3. Polygon IDs..... 12
 - 3.4.4 Variables included with each polygon 13
- 4. Data formats for the current release:..... 18
 - 4.1. Notes for the KML version: 18
 - 4.1.2. Technical details for the KML version..... 19
 - 4.2. Notes for the ESRI shape-file: 20
 - 4.3. Notes for the Google Fusion Tables format..... 20
- References: 21
- Appendix 1. Data points in the UCDP GED Point dataset not included in the UCDP GED Conflict Polygons dataset.
- Appendix 2. UCDP Definitions

To cite this dataset:

Please **always** cite when using this dataset: **Croicu, Mihai, Ralph Sundberg (2012), UCDP GED Conflict Polygons dataset version 1.1-2011, UCDP, Department of Peace and Conflict Research, Uppsala University.**

Please do not abbreviate the name of the dataset in any other way but the **UCDP GED Conflict Polygon** dataset.

The most current and sole official version of this dataset, as well as the most current and sole official versions of the UCDP GED Point Dataset are available free of charge at <http://www.ucdp.uu.se/ged>.

1. Introduction and purpose

This codebook describes the UCDP GED Conflict Polygon Dataset, produced by the Uppsala Conflict Data Program (UCDP) as part of the Geo-referenced Event Data (GED) project.

Following the increase in demand from scholars interested in armed conflict to have access to reliable disaggregated data, UCDP decided to convert all of its conflict- and dyad-level calendar-year data into a geo-referenced disaggregated event-data format. This project aims to provide data covering the 1989-2010 time period in all three of the UCDP's categories of organized violence: state-based armed conflict (1), non-state conflict (2) and one-sided violence (3).

This dataset is one of the products released as part of this endeavor, describing 'conflict polygons', i.e. geographical areas that have been affected by each of the three types of organized violence per UCDP's definition. These geographical areas are built in a standardized fashion, using the methodology described below, with the explicit aim of providing a coherent, easily-comparable and standardized dataset.

These areas are provided both in a per-year version, covering the geographical area affected by each and every state-based armed conflict, non-state conflict and one-sided violence in each year it has surpassed the 25 battle related deaths threshold in the UCDP datasets, as well as polygons showing the overall extent (spread) of the conflict throughout the period. The current version of the dataset only covers conflicts taking place in Africa (including North Africa), between 1989 and 2010.

The UCDP GED Conflict Polygon Dataset, while aggregated (the main object in the dataset is the geographical area affected by one incidence (dyad) of violence) is based on fully disaggregated data. The source of the data used for constructing this dataset is the UCDP GED Point dataset (also a part of the

UCDP GED project, available at <http://www.ucdp.uu.se/ged>), covering individual instances of lethal violence occurring at clearly explicit spatially- and temporally-described locations.

Thus, unlike previous attempts at attributing geographical dimensions to conflict (such as Buhaug and Gates, 2002), this dataset is not derived from the classical country-year paradigm, but rather from the actual events of violence taking place as part of an armed conflict. For this dataset, the state is no longer the classical 'black-box' that has been the mainstay of research. As such, polygons may and do spread across country borders, following the actual spread of violence during a conflict.

Please, however, note that while this dataset solely intends to describe the principal areas affected by organized violence. These polygons are not constructed to be perfectly accurate, highly detailed maps of conflict dynamics (such maps could only be constructed through in-depth field studies in each and every individual case), but rather relatively accurate, highly standardized areas, constructed based on a simple, easily reproducible methodology applicable throughout the dataset, perfectly suited for cross-case comparisons, statistical analyses or GIS-based research. Nonetheless, we believe and hope this dataset will, as well, find use amongst scholars interested in cross-case comparisons and field studies, providing a tool for case selection and specification at subnational level.

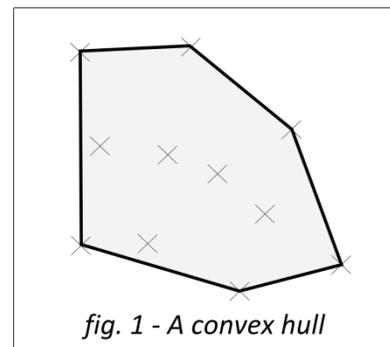
2. Definition of the UCDP GED polygon

The basic unit of the dataset is the *GED polygon*, describing the basic geographical area affected by an episode of organized violence (armed conflict, non-state violence or one-sided violence) in a certain time frame. The UCDP GED polygon is defined as:

"The smallest possible convex geographical area that encompasses the locations of all UCDP GED events in each covered UCDP dyad within a specified time period."

In the context of this definition, a number of elements are further explained:

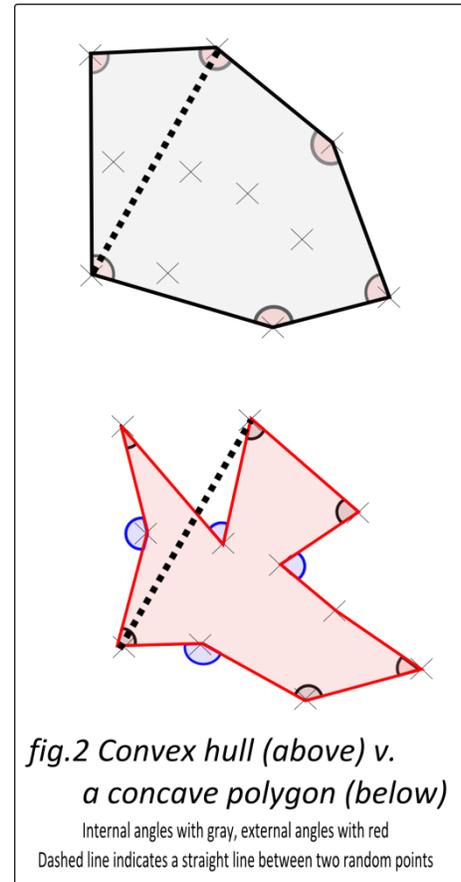
1. The smallest possible convex geographical area is defined as the convex hull constructed with the points in the given set. This shape is unique for each and every set of points given. See **fig. 1** for a graphical explanation of such a polygon.



1.1 Convex refers to the property of the polygon to have all its angles pointing towards the inside of the polygon (see **fig. 2** for a clarification). Further, any two points within a convex polygon can be connected through a straight line that will stay completely inside the polygon. A non-convex polygon will be called a concave polygon for the purpose of this codebook.

For details on the construction of these polygons, refer to paragraph **3** of this codebook, **Description of UCDP GED Polygons**.

2. UCDP GED event: UCDP defines an event as the incidence of the use of armed force by an organised actor against another organized actor, or against civilians, resulting in at least 1 direct death in either the best, low or high estimate categories at a specific location and for a specific temporal duration. For details on this, please consult Sundberg, Lindgren and Pads kocimaite (2012), *UCDP Geo-referenced Event Dataset (GED) Codebook*, version 1.1.
3. UCDP dyad, in this definition refers to all of the dyads part of interstate and intrastate conflicts covered by the UCDP/PRIO Armed Conflict Dataset and the UCDP Dyadic Dataset as well as by all the dyads of non-state violence in the UCDP Non-state Conflict Dataset and by all occurrences of violence against civilians covered in the UCDP One-sided Violence Dataset and coded as taking place in Africa between 1989 and 2010¹.



¹ The UCDP definition of a dyad (UCDP, 2012) is: A dyad is made up of two armed and opposing actors. In state-based conflicts a dyad is defined as two actors, with one or more being the government, that have a stated incompatibility.

In a non-state conflict a dyad is constructed by at least two organised actors, of which none is the government of a state, that oppose each other with arms. In non-state conflicts it is possible for an alliance of non-state actors to enter into a dyad with either an opposing group, or an alliance of opposing groups.

For the purpose of this dataset, for reasons of data usability, brevity and clarity, armed actors (a government or a formally organised non-state group) attacking the civilian population in the one-sided violence category of UCDP organised violence also forms a 'dyad', labeled as 'armed actor – civilians'.

For a dyad to be included in the UCDP data (and thus in this dataset), it must cross the 25 battle-related deaths threshold in a calendar year.

4. Specified time period: This dataset covers the same time period as the UCDP GED dataset version 1.1, that is, 1989-2010. Aggregate polygons containing all the points of a given UCDP dyad and covering the entire period are provided as well as Individual dyad-year polygons are provided, covering every calendar year a UCDP dyad has been active.

Each polygon is a geographical shape that can be over-imposed on a map or globe. The polygons are described by a series of vertexes, i.e. places where two edges of the shape meet (defined by latitude and longitude expressed in decimal degrees) and the lines (edges) connecting the vertices. All polygons are two dimensional (i.e. do not contain altitude data). These polygons are made available in a number of formats for use with various software, such as Geographic Information Systems (GIS), Google Earth, Google Maps etc (see section 3.4 for details).

Please note that the UCDP *GED Polygons* are **not** heat-maps, the surface of the polygon being homogeneous. Thus, using solely the GED Polygons, it is impossible to tell whether the level of violence in one sub-area of a polygon is higher or lower than in another sub-area of the polygon. One may do such analysis by combining the UCDP *GED Polygons* with the compatible UCDP *GED GRID*.

Comparing levels of violence between two or more polygons, as well as calculating violence densities at the level of whole polygons is, however, possible using solely the *UCDP GED Polygon dataset*.

Further, the shape of each polygon is not indicative of actual fighting taking place in each and every point within the polygon. Rather, one can say that all events of fighting in a *UCDP dyad* are contained within and bound by the polygon.

3. Description of UCDP Polygons

This section describes the data used for constructing the *UCDP GED Polygons*, the construction process as well as the format and contents of the resulting polygons.

3.1. Data used for constructing the UCDP GED polygons

The UCDP GED Polygons have been constructed based on version 1.1 of the UCDP GED Event dataset published by the Uppsala Conflict Data Program (UCDP) at Uppsala University.

Just as the UCDP GED Event dataset, the UCDP GED Polygon dataset has a dyad and actor focus, tracing the events of all dyads and actors that have crossed the 25

deaths threshold in any year of the UCDP annual data. The dataset includes all three types of UCDP organised violence: state-based conflict, non-state conflict and one-sided violence. All three conflict categories of the UCDP are mutually exclusive and coded events will therefore also be exclusive and non-overlapping. The data series start in 1989 and events before this calendar year are not included.

The following data is included in the UCDP GED Polygon dataset:

Conflict Type	Period	Actor Inclusion	Event Inclusion	Reference
<i>State-Based</i>	1989-2010	All dyad-years that cross the 25 death threshold and have a stated incompatibility.	All events leading to at least one death.	UCDP/PRIO Armed Conflict Dataset Codebook Version 4-2011
<i>Non-State</i>	1989-2010	All dyad-years that cross the 25 death threshold.	All events leading to at least one death.	UCDP Non-State Conflict Codebook Version 2.3-2010
<i>One-Sided</i>	1989-2010	All actor-years that cross the 25 death threshold.	All events leading to at least one death.	UCDP One-Sided Violence Codebook, Version 1.0-September 28, 2005

For a discussion of biases, data reliability and data validity of the source data, refer to the UCDP GED Codebook v. 1.1-2011 (Sundberg/Lindgren/Padskocimaite, 2012).

To create the polygons, only those GED events covered by a geo-precision score of **1** (location corresponds exactly to the geographical coordinates available), **2** (exact location is known to be located in a 25-km radius from the geographical coordinates in the dataset) and **3** (exact location can only be specified to a second-order administrative division such as a district and municipality) were included.

This was done to maximize the precision of the resulting polygons – accepting points with lower geo-precision (such as locations coded only to the first level administrative division or country level) would significantly increase the margin of error of the resulting polygon (by adding points located too distant from the real location of fighting). This would lead to imprecise polygons of which large parts are actually the result of poor available information and not actual occurrence of violence.

To illustrate this point, the case of Sudan is representative. Using only locations coded to geo-precisions 1, 2 and 3, the distance between the real place of fighting and the points available in the dataset ranges from between 0 and less than 100 km. Using points with geo-precisions 4, this maximum distance increases to over 500 km, while using points with higher geo-precisions leads to a maximum distance of about 2500 km.

Thus, for the purpose of creating this dataset, **16942** data points (observations) were extracted from a total of **21860** total data points. These 16942 data points cover **4923** distinct locations (individual latitude/longitude pairs).

3.2. Choice of a Convex Hull

While a concave polygon would have had a smaller surface in most cases than a convex hull, thus containing all observed points in a smaller area, a convex hull was judged to be more appropriate for this dataset. The main reason for this is the unique nature of a convex hull – for any set of points, there is one and only one convex hull, being a direct, fully accurate representation of the data points available.

Creating a concave polygon would have required complex coding decisions for each and every externally-pointing angle, adding bias to each polygon (as each coding decision would have been taken in relation with the knowledge of the coder to the specific country or conflict). This would make such polygons not comparable with each-other, making the resulting dataset less useful for any analysis. Further, even the creation of an algorithmic rule for the creation of such polygons would have had significant bias issues, as the algorithm would only reflect a series of assumptions that would not reflect in any way either the real-life situation on the ground, or the available data.

Further, the choice of a convex hull for the UCDP GED Polygons has the advantage of simpler data replication, as convex hulls are algorithmically created.

3.3. Outliers and Special Cases

In a small number of cases, a limited number of incidents of violence took place at a distant location from the main area of violence. Examples of such events captured by the UCDP GED data are an individual bombing of a subway station in Paris as part of a conflict otherwise exclusively confined to Algeria, an assassination in Addis Abeba as part of a conflict taking place in Egypt, fighting between a group of students in the capital as part of a conflict between two ethnic groups taking place elsewhere in the country etc.

Including these data points in the general algorithm for creating the convex hull would lead to the creation of extremely large polygons, with much of their surface covering zones where no armed conflict

took place, stemming solely from a single data point. Further, such polygons would no longer describe the basic geographical area affected by an episode of organized violence, making these polygons useless for analysis.

Moreover, while these individual outliers provide valuable research opportunities, most of these opportunities stem specifically from their existence as outliers, and not from their presence in an aggregated dataset as the UCDP GED Polygon dataset.

As such, for the purpose of this dataset, the following rule, further referred to as the 20%-5% rule, has been devised for identifying outliers:

An outlier is an event (data point) or a group of events (data points) that:

- 1. If included, would alone account for 20% or more of the total area of the resulting polygon² OR*
- 2. If included, it would drop the total density of the data points in the polygon by more than 20%;*

while at the same time satisfying the following conditions:

- 1. It accounts for 5% or less of the total number of data points in the given UCDP dyad for which the polygon is constructed AND*
- 2. It accounts for 5% or less of the total number of deaths taking place in the polygon for the given UCDP dyad (in the best estimate category).*

Note that outliers are solely determined at the fully aggregated level (dyad/conflict) and not at the dyad-year level. Points considered outliers at the fully aggregated level are considered outliers in the dyad-year version.

For the purpose of this dataset, these outliers are not taken into account when creating the UCDP GED Polygon dataset.

As a further rule, *no more than 5% of the total number of total events in the UCDP GED Point dataset may be eliminated*. If this number is exceeded, events eliminated are those with the largest impact (points accounting for the largest increase in area or decrease in density).

A list of all eliminated points is available as **appendix 1**, attached with this codebook.

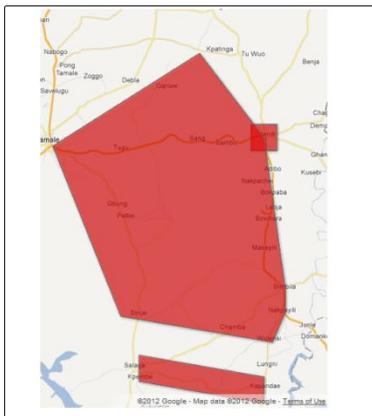
Further, a very limited number of UCDP dyads are treated as special. This was done specifically when the resulted algorithmically-computed polygon would have deviated significantly from the real-world zones

² In practice, for this dataset, the area of the resulting polygon is determined through the usage of the Chamberlain/Duquette (2007) formula ($A = -\frac{R^2}{2} \sum_{i=1}^n (\lambda_{i+1} - \lambda_{i-1}) \cdot \sin \varphi_i$, where λ_i and φ_i are the longitude and latitude of a point i (in radians) and R the radius of Earth. This algorithm is implemented through an in-house solution running over a database of the UCDP GED data points. While the results are spherical approximations deviating from the WGS84 geoid, they are judged to be sufficient for the purpose of determining outliers in the UCDP GED Polygons.

of fighting (i.e. through the existence of two separate zones of fighting, or through pockets of fighting not able to be otherwise considered as outliers according to the 20%-5% rule).

In such cases, data points that were theoretically supposed to be part of the same polygon were split into two or more groups, after a case-by-case analysis, with convex hulls computed for each of the groups. As such, in such cases, a dyad may be represented by two or more separate polygons. These cases are further referred to as *'split dyads'* and are detailed in **appendix 1**.

3.4. Construction and specifications of the polygons



*fig. 3 - Convex hull, point-derived and line-derived polygon example from the dataset
Visualized in Google Maps*

Each polygon is constructed using an in-house software solution based on the classical Quickhull algorithm (Barber/Dobkin/Huhdanpaa, 1996; a reference algorithm was developed by Barber, 1996; another by Westhoff, 2010). For an illustration see **fig. 3**.

A **convex hull** requires at least three distinct locations (latitude/longitude pairs) to be constructed – no polygon can be defined with less (two points define a line, one defines a point). For conflicts or dyads with less than 3 events (data points), the following has been done:

1) If a UCDP dyad contains only one event (data point), a square of 0.1 degrees of latitude by 0.1 degrees of

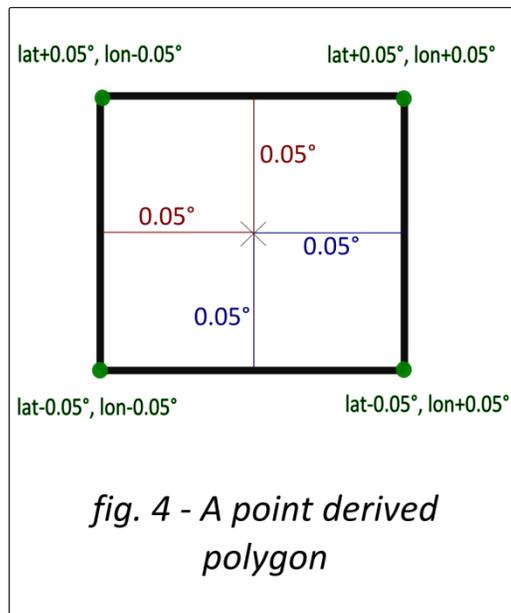


fig. 4 - A point derived polygon

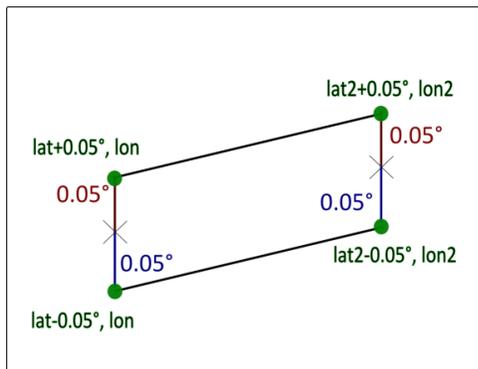


fig. 5 - A line derived polygon (not to scale)

longitude is constructed around the point. The point will lie exactly in the center of the rectangle (at the intersection of its diagonals). This case is further referred to as a *"single-point derived polygon"*. The points defining the polygon, in a clockwise direction

from the top-left are (lat+0.05, lon-0.05);

(lat+0.05, lon+0.05); (lat-0.05, lon+0.05); (lat-0.05, lon-0.05), where lat and lon are the coordinates of the point. See **fig. 4** for a graphical explanation.

- II) If a dyad contains two events (data points), the resulting line is padded with 0.1 degrees of latitude (0.05 degrees of latitude on both sides of the line) in order to create a rectangle. Given lat, lon as the coordinates of the first data point and lat2, lon2 as the coordinates of the second data point, the coordinates of the resulting polygon will be (lat+0.5, lon); (lat-0.5, lon); (lat2-0.5, lon2); (lat2+0.5, lon2)³. See **fig. 5** for a graphical explanation. Polygons generated in this way are referred to as “**line derived polygons**”.

3.4.1. Padding of the polygons

No padding whatsoever has been used when constructing the polygons – vertices of the polygon consist exclusively of UCDP GED events (data points), and data points may and will exist on the edges of the polygon.

It is up to the users of the data to add padding to the polygons according to their needs. While such padding may be useful for various research purposes (it can be expected that, e.g., the area affected by fighting extends beyond the punctual location of actual violence), it is up to the users to add it according to their own needs. While this is easily accomplished in most GIS software, it is beyond the scope of this manual to explain in any detail. Please consult the manual of your GIS solution for the practical way to achieve this.

3.4.2. Geographical assumptions

UCDP does not claim authority with concern to either geographic features (coast lines, terrain, relief etc.) or geo-political features (state borders, ownership of disputed territories etc.). As such, the *UCDP conflict polygons* make no assumption with regards to such features and are constructed ignoring such features (only taking into account the exact locations of violent *events*). The only assumption of geographic nature taken into account in the creation of the dataset is the use of the WGS84 reference system for determining these polygons.

³ In one case, if lon=lon2, these transformations would have no effect, the resulting shape also being a line following the North South direction. This case was taken into account by the software solution designed to generate the dataset, with no cases identified for the current version.

Therefore, some parts of individual *conflict shapes* may either cross shorelines, thus overlapping various bodies of water, or may cross national borders even if a conflict has been, in practice, taking place in a single country. It is up to the user to intersect UCDPs polygons with their desired shoreline data or, if needed for analysis, to make sure polygons are confined within a single country by intersecting the dataset with one of the many available authoritative sources of geographical information on boundaries⁴.

Please note that UCDP provides a variable within each polygon containing the names and ISO country codes of countries where the actual violence described in the polygon took place.

3.4.3. Polygon IDs

Each polygon is defined by a `PolygonID`, a unique identifier assigned to each individual polygon based on the information it was generated on. The format of this ID is as follows:

Polygon type-Type of violence-Dyad ID-Construction mode-Year

Polygon type:

PLY: A *proper convex hull* (see section 3.4), constructed out of 3 or more data points.

LNE: A *line derived polygon* (see section 3.4), constructed out of 2 data points.

PNT: A *point derived polygon* (see section 3.4), constructed out of a single data point.

Type of violence:

Reflects the type of conflict for which the polygon is constructed:

1. State-based violence
2. Non-state violence
3. One-sided violence

Dyad ID:

⁴ Some of the projects that provide GIS data regarding boundaries of sovereign states and disputed territories include Natural Earth (www.naturalearthdata.com), the GADM Project (www.gadm.org), CShapes (also providing historical data), (<http://nils.weidmann.ws/projects/cshapes>), DIVA-GIS (<http://www.diva-gis.org/Data>) etc. Similarly, shorelines can be obtained from a variety of sources, including the USGS (http://pubs.usgs.gov/of/2006/1187/basemaps/continents/continent_faq.htm), Natural Earth, GADM or from ESRI. UCDP does not particularly endorse or recommend any such source of data.

Croicu, M. C. and Sundberg, R. (2012), UCDP GED Polygon Dataset codebook, version 1.1

UCDP dyad ID code for state based dyad. (Themnér/Wallensteen, 2011)

UCDP conflict ID code for non-state dyad. (Pettersson, 2011)

UCDP actor ID code for the one-sided violence actor. (Sundberg, 2009)

Note that Conflict, dyad and actor IDs all correspond to the IDs used by the UCDP in the datasets that deal with the respective type of organized violence, as well as the IDs listed in the UCDP Actor dataset (UCDP, 2011).

Thus, the Dyad ID should be read only together with the type of violence variable, as there may be separate Dyad ID 100 for type of violence 1 (state-based) and a Dyad ID 100 for type of violence 2 (non-state conflict).

If the data points that were theoretically supposed to be part of the same polygon were split into two or more groups resulting in two or more separate polygons (see section 3.3), a letter is appended to the dyad ID, starting from A, for each of the resulting groups and thus for each of the resulting polygons (e.g. 313A, 313B, 313C, etc.).

Construction mode:

E, standing for 'Exact' if **Polygon type** is PLY.

A, standing for 'Approximation' is used if **Polygon type** is LNE or PNT (i.e. derived from less than three points).

Year:

The calendar year covered by the polygon. Only dyad-years where at least one data point exists are covered by the dataset.

For aggregated polygons, covering the entire 1989-2010 period, this part of the PolygonID is missing.

Examples of PolygonIDs:

PLY-1-55-E: Proper convex hull (PLY), covering a state based dyad (-1-), UCDP dyad ID 55, exact construction method, aggregated polygon that covers the entire 1989-2010 period.

PNT-2-103-A-2000: Point derived-polygon (PNT), covering a non-stat dyad (-2-), UCDP dyad ID 103, polygon derived from less than three points. Yearly polygon, covering the period spanning between 2000-01-01 and 2000-12-31.

3.4.4 Variables included with each polygon

Variable Name	Content	Type
Polygon_ID	The unique Polygon ID given in accordance to the rules and containing the data set forward in section 3.4.3 Polygon IDs of this codebook.	Formatted string (AAA-N-NNNN-A-[-NNNN])
Type_of_violence	Reflects the type of conflict: 1. State-based 2. Non-state 3. One-sided	Integer Range: 1-3.
Dyad_ID	UCDP dyad ID code for the state-based dyad that this polygon represents. UCDP conflict ID code for the non-state dyad that this polygon represents.. UCDP actor ID code for the one-sided violence actor that this polygon represents. For alliances between non-state actors in non-state conflicts or one-sided violence the same system is used as in the UCDP GED Dataset. Note that Conflict, dyad and actor IDs all correspond to the IDs used by the UCDP in the datasets that deal with the respective type of organized violence. Thus, the Dyad ID should be read only together with the type of violence variable, as there may be separate Dyad ID 100 for type of violence 1 (state-based) and a Dyad ID 100 for type of violence 2 (non-state conflict). If the data points that were theoretically supposed to be part of the same polygon were split into two or more groups resulting in two or more separate polygons (see section 3.3), a letter is appended to the dyad ID, starting from A, for each of the resulting groups and thus for each of the resulting polygons (e.g. 313A, 313B, 313C, etc.).	Integer Range 1-599999
Dyad_name	Name of either the dyad (state-based and non-state conflicts) or the actor (one-sided violence).	Formatted string 'Side_A' - 'Side_B'
Split_dyad	Indicates whether a dyad is 'split' (see section 3.3 Outliers and Special Cases), i.e. more than one polygon is used for the representation of this dyad.	Boolean Represented by integer
Conflict_ID	Conflict ID that corresponds to the conflict name (below).	Integer Range 1-599999
Conflict_name	Name of the conflict to which the dyad belongs. For non-state conflicts and one-sided violence this corresponds to the dyad name.	String length: <500 char.
Side_A_ID	The unique ID of Side A in the dyad represented by the polygon, taken from the UCDP Actor Dataset.	Integer Range 1-599999

	<p>In state-based conflicts always the ID of a government. In one-sided violence always the ID of the perpetrating party.</p> <p>For alliances between non-state actors in non-state conflicts or one-sided violence the same system is used as in the UCDP GED Dataset.</p>	
Side_B_ID	<p>The unique ID of Side B in the dyad represented by the polygon, taken from the UCDP Actor Dataset.</p> <p>In state-based conflicts always the ID of a rebel movement or rivalling government. In one-sided violence the value is always 9999, the value representing <i>civilians</i>.</p> <p>For alliances between non-state actors in non-state conflicts or one-sided violence the same system is used as in the UCDP GED Dataset.</p>	<p>Integer Range 1-599999</p>
Side_A	<p>The name of Side A in the dyad represented by the polygon, taken from the UCDP Actor Dataset.</p> <p>In state-based conflicts always a government. In one-sided violence always the perpetrating party.</p>	<p>String length: <500 char.</p>
Side_B	<p>The name of Side B in the dyad represented by the polygon, taken from the UCDP Actor Dataset.</p> <p>In state-based conflicts always the rebel movement or rivalling government. In one-sided violence always <i>civilians</i>.</p>	<p>String length: <500 char.</p>
Year	<p>The calendar year when the violence covered in the polygon took place.</p> <p>As polygons are available both at both dyad-year level and at aggregated dyad level (covering the entire 1989-2010 period), aggregated polygons do not contain any value in this field.</p>	<p>Year Range: 1989-2010</p>
No_of_events	<p>The number of <i>UCDP GED events</i> taken into account for the construction of the polygon.</p> <p>See section 3.1. Data used for constructing the UCDP GED polygons for details on the data selection process.</p>	<p>Integer Range: 1-9999</p>
Best_est	<p>The UCDP best estimate of casualties for the polygon. It is the sum of the best estimates in all of the events taken into account for the construction of the polygon.</p> <p>In effect, the estimate is the aggregated most reliable numbers for all incidents of organized violence covered in the polygon. If different reports provide different estimates, an examination is made as to what source is most reliable. If no such distinction can be made, UCDP as a rule includes the lower figure given.</p>	<p>Integer Range:1-9999999</p>

	<p>Please note that due to the selection process used for the creation of the polygons that removes a significant number of events coded with low geographical accuracy scores (see 3.1. Data used for constructing the UCDP GED polygons) , for a polygon to be included in the UCDP GED dataset it need not pass the 25-deaths threshold needed for its respected UCDP dyad to be included in the UCDP data.</p> <p>Moreover, due to the same reasons, the best estimate need not be the same as (and, in most cases is different from) the one given in the UCDP Battle Related Deaths dataset, the UCDP Non-state Conflict dataset or the UCDP One-sided Violence dataset.</p> <p>Thus, while the dyad that is represented by the polygon has over 25 battle-related deaths in a calendar year (so as to pass the UCDP inclusion threshold), the polygon need only have at least 1 battle related death in this category to warrant inclusion in this dataset.</p>	
<p>Low_est</p>	<p>The UCDP low estimate of casualties for the polygon.</p> <p>In effect, it is the aggregated low estimates for all incidents of organized violence covered in a polygon. If different reports provide different estimates and a higher estimate is considered more or equally reliable, the low estimate is also reported if deemed reasonable.</p> <p>Same considerations apply as in the case of the best_est variable.</p>	<p>Integer Range:0-9999999 Low_est<=Best_est</p>
<p>High_est</p>	<p>The UCDP high estimate of casualties for the polygon.</p> <p>In effect, it is the aggregated high estimates for all UCDP GED events (incidents of organized violence) covered in a polygon. If different reports provide different estimates and a lower estimate is considered more or equally reliable, the high estimate is also reported if deemed reasonable. If there are incidents when there is some uncertainty about which party has been involved, these may also be included in the high estimate.</p> <p>Same considerations apply as in the case of the high_est variable.</p>	<p>Integer Range:1-9999999 High_est>=Best_est</p>
<p>Onset_date</p>	<p>Date of the chronologically first UCDP GED event (first event of organized violence) taken into account in the creation of the polygon.</p>	<p>Date Format: YYYY-MM-DD</p>

	<p>For UCDP GED events spanning multiple days, the end dates of the events have been taken into consideration.</p> <p>Please note that due to the selection process used for the creation of the polygons (see 3.1. Data used for constructing the UCDP GED polygons), a significant number of events coded with low geographical accuracy scores were eliminated in the construction of this dataset. This will make onset dates different from those given in the UCDP Dyadic Dataset, UCDP Non-state Actor Dataset or UCDP One-sided Violence Dataset. Thus, the onset date is valid solely for the polygon, and not for the dyad it represents.</p> <p>This variable is only coded for dyad-aggregated polygons and not for dyad-year polygons.</p>	
Onset_date_25	<p>Date when a number of 25 deaths has been reached in the polygon in the best estimate of casualties category.</p> <p>For UCDP GED events spanning multiple days, the end dates of the respective events have been taken into consideration.</p> <p>Please note that due to the selection process used for the creation of the polygons that removes a significant number of events with low geographical accuracy scores (see 3.1. Data used for constructing the UCDP GED polygons), a number of events were eliminated in the construction of this dataset. This will result in the fact that some polygons will never exceed the 25 battle-related deaths threshold.</p> <p>Further, this will mean that onset dates in this dataset are valid solely for polygons and not for the dyad represented by the polygons. Moreover, this will make a significant number of onset dates different from the ones coded in the UCDP datasets.</p> <p>See also the entry for best_est in this table.</p> <p>This variable is only coded for dyad-aggregated polygons and not for dyad-year polygons.</p>	<p>Date Format: YYYY-MM-DD</p>
Last_activity_date	<p>Date of the chronologically last UCDP GED event (last incident of organized violence) taken into account in the creation of the polygon.</p> <p>For UCDP GED events spanning multiple days, the end dates of the respective events have been taken into consideration.</p>	<p>Date Format: YYYY-MM-DD</p>

Please note that due to the selection process used for the creation of the polygons (see **3.1. Data used for constructing the UCDP GED polygons**), a significant number of events coded with low geographical accuracy scores were eliminated in the construction of this dataset. This will make termination dates different from those in the UCDP Conflict Termination dataset.

Thus, the last activity date is valid solely for the polygon, and not for the dyad it represents.

This variable is only coded for dyad-aggregated polygons and not for dyad-year polygons.

4. Data formats for the current release:

For this version, data is made available as an ESRI shapefile for use with popular geographic information systems (GIS), as a kml file for use with Google Earth and other xml-capable software, and as a Google Fusion Tables id for cloud-based computing. **All versions have the same citation requirements and usage policy.**

4.1. Notes for the KML version:

The KML file is ready to use with any supporting Google product (Google Earth, Google Maps, Google Maps API etc.). Further, it is sufficiently clean XML (the file is generated via Google Fusion Tables), to be usable with any XML parser.

Two separate files are provided, one for the aggregated (1989-2010) polygons, the other for the yearly polygons.

In the aggregated polygons file, data is split in three sub-folders, one for each of the three types of violence.

In the yearly polygons file, the data is split in yearly folders, each covering one calendar year. Further, each folder contains three sub-folders, one for each of the three types of violence

If you only want to use this file in conjunction with a Google software (such as Google Earth), there is no need to read any further.

4.1.2. Technical details for the KML version.

Each individual polygon is contained within its own <Placemark>, following the same naming convention used by Google's Fusion Tables (each <name> tag associated to the <Placemark> contains the following value <![CDATA[**PID**]]> where **PID** is the **Polygon ID** of the respective polygon).

Each polygon is described clockwise with the edge of the convex hull being the outer boundary of the KML polygon. Each node has an altitude of 0.0.

Please note that, per KML's specification, the format for geo-data is reversed from the traditional order, following the 'longitude', 'latitude', 'altitude' pattern.

The variables included with each polygon (see section 3.4.4 of the codebook) are contained in the <ExtendedData> tag of each <Placemark>, with the following format:

```
<ExtendedData>
<Data name="VARIABLE_NAME">
  <value>
    VARIABLE_VALUE
  </value>
</Data>
...
</ExtendedData>
```

e.g.:

```
...
<Data name="dyad_name">
<value>Government of Algeria - AIS</value>
</Data>
<Data name="low_est">
<value>1322</value>
</Data>
...
```

Moreover, each polygon has a 'Google description bubble' that contains the following information in textual form:

```
<description>
<![CDATA[<div class="googft-info-window" style="font-family:sans-serif">
<b>Name:</b>Polygon_ID<br>
<b>type_of_violence:</b>Type_of_violence<br>
<b>dyad_ID:</b>Dyad_id<br>
<b>dyad_name:</b>Dyad_name<br>
<b>low_est:</b>low_est<br>
<b>best_est:</b>best_est<br>
<b>high_est:</b>high_est<br>
<b>date_first:</b>onset_date<br>
<b>date_first_25:</b>onset_date_25
</div>]]>
</description>
```

4.2. Notes for the ESRI shape-file:

The reference system that should be used with the shape-file is WGS84.

4.3. Notes for the Google Fusion Tables format

Only the aggregated polygons (1989-2012) are currently released in the Google Fusion Tables Format.

The table ID is: **3212021**

The encrypted table ID is: **11Ilg2HHDnftXsi0y2f_7KfJdOJM0Da6IFXxfVEpc**

Note that the data **is not** open source and **is not** in the public domain. If you use UCDP GED Data in **any form including Google Fusion tables**, the same citation requirements exist. **Always** quote Uppsala Conflict Data Program, Croicu, M.C., R. Sundberg (2012), UCDP GED Conflict Polygons Codebook v.1.1.

References:

Barber, C.B., Dobkin, D.P., and Huhdanpaa, H.T. (1996), The Quickhull algorithm for convex hulls, in *ACM Transactions on Mathematical Software*, 22(4):469-483.

Barber, C.B. (1996; 2003; 2012), Qhull reference implementation, <http://www.qhull.org/>

Buhaug, H. and Gates S. (2002), The Geography of Civil War, in *Journal of Peace Research*, 39(4):417-433.

Chamberlain, R.G., Duquette, W.H. (2007), Some Algorithms for Polygons on a Sphere, NASA JPL, Publication 07-3.

Pettersson, T. (2011), Non-state Conflicts 1989-2010 – Global and Regional Patterns in Pettersson, Therése and Lotta Themnér Eds. (2011) *States in Armed Conflict 2010*, Department of Peace and Conflict Research, Uppsala University.

Sundberg, R. (2009), Revisiting One-sided Violence – A Global and Regional Analysis, in Harbom, Lotta and Ralph Sundberg, *States in Armed Conflict 2008*, Department of Peace and Conflict Research, Uppsala University.

Sundberg, R., Lindgren, M. and Padsokocimaite, A. (2010), UCDP GED Codebook version 1.1-2011, Department of Peace and Conflict Research, Uppsala: Uppsala University.

Themnér, L. and Wallensteen, P. (2011), *Armed Conflict, 1946-2010*, in *Journal of Peace Research* 48(4).

UCDP (2011), UCDP Actor Dataset 2.1-2011, Uppsala Conflict Data Program, www.ucdp.uu.se, Uppsala University".

Westhoff, J. (2010), Calculate a convex hull - The QuickHull algorithm, http://westhoffswelt.de/blog/0040_quickhull_introduction_and_php_implementation.html/

Appendix 1.

1. Points not taken into account in the creation of the UCDP GED Conflict Polygons Dataset

Polygon	Points not taken into Account for the construction of the UCDP GED Conflict Polygons Dataset	Observations
PLY-1-3-E	ALG-1998-1-3-30	
PLY-1-12-E	BUI-1995-1-12-22	
PLY-1-13-E	BUI-1997-1-13-3	
PLY-1-27-E	CHA-1999-1-27-6.1	
PLY-1-41-E	DRC-1998-1-41-4	
PLY-1-89-E	IVO-2002-1-89-7	
PLY-1-93-E	LBR-1990-1-93-50	
PLY-1-128-E	RWA-2009-1-128-47	
PLY-1-129-E	SEN-2000-1-129-6	
PLY-1-131-E	SIE-1998-1-131-9	
PLY-1-190-E	ANG-2002-1-190-3	
PLY-1-211-E	SOM-1996-1-211-10	
PLY-1-241-E	EGY-1995-1-241-37	
PLY-1-433-E	SUD-2008-1-433-3	
PLY-1-434-E	SUD-2008-1-434-24 SUD-2007-1-434-6	
PLY-1-631-E	SUD-2006-1-631-9	This point does not fulfill one criterion for elimination in the codebook (<i>no more than 5% of the total number of events may be eliminated</i>) but was nonetheless eliminated as the best fatality estimate for that point was 0 deaths.
PLY-1-641-E	SUD-1997-1-641-3.1	
PLY-1-645-E	SUD-1999-1-645-10	
PLY-1-770-E	SOM-2009-1-770-138 SOM-2008-1-770-51 SOM-2008-1-770-52	
PLY-2-86-E	NIG-1997-2-86-10.1 NIG-1997-2-86-10.2	One of the two points does not fulfill one criterion for

		elimination in the codebook (<i>no more than 5% of the total number of events may be eliminated</i>) but was nonetheless eliminated as the best fatality estimate for that point was 0 deaths.
PLY-2-307-E	SOM-1996-2-307-35	
PLY-2-360-E	NIG-2004-2-360-24	
PLY-3-461-E	TOG-2005-3-461-99 TOG-2005-3-461-100	
PLY-3-501-E	KEN-2007-3-501-12	
PLY-3-560-E	SAF-1989-3-560-12	
PLY-3-625-E	SUD-1990-3-625-3 SUD-2007-3-625-1 SUD-2007-3-625-2 SUD-2000-3-625-9	
PLY-3-1268-E	DRC-1997-3-1268-23 DRC-1997-3-1268-76	
PLY-3-1270-E	DRC-2001-3-1270-9	
PLY-3-1278-E	BUI-2006-3-1278-9	
PLY-3-1280-E	BUI-1996-3-1280-1	
PLY-3-1336-E	UGA-2002-3-1336-1 UGA-2009-3-1336-166	
PLY-3-1337-E	UGA-1998-3-1337-6 UGA-1998-3-1337-25	
PLY-3-1381-E	SEN-1998-3-1381-8	
PLY-3-1390-E	ALG-1999-3-1390-19 ALG-1995-3-1390-1	One of the two points does not fulfill one criterion for elimination in the codebook (<i>no more than 5% of the total number of events may be eliminated</i>) but was nonetheless eliminated as the best fatality estimate for that point was 0 deaths.
PLY-3-1421-E	ANG-2000-3-1421-57 ANG-2000-3-1421-53 ANG-1994-3-1421-18	
PLY-3-1496-1	LBR-1994-3-1496-14	
PLY-3-1505-1	DRC-2002-3-1505-8	
PLY-3-1565-1	KEN-2008-3-1565-3	

16.942 points extracted from the UCDP GED Point Dataset version 1.1. **52** of these points were not taken into account for the construction of the UCDP GED Polygon Dataset version 1.1, representing **0.31%** of the total data points extracted from the UCDP GED.

If they were to have been taken into account, they would have increased the cumulated area of the polygons (including those where no change was made) by approximately **8.5%**.

2. Dyads that were *split* (i.e. dyads resulting in the creation of more than one polygon)

Type of violence	Dyad_ID	Number of resulting polygons	Observations
1 (state-based)	2	2	Two separate conflict areas, one near Algiers, one close to the desert town of El Oued.
1 (state-based)	4	2	A large majority of incidents of violence took place in the densely populated areas of northern Algeria, making a compact, high density conflict zone. However, apx. 40 events took place in the Sahara, creating a very large, low density area. This produced two separate zones (one small/high density and one large/low density). As this phenomenon was difficult to attribute to differences in activity, but rather to differing population densities, it was decided to keep the events in two separate conflict zones.
1 (state-based)	49	2	Events in the capital city area were separated from those in main conflict zone (located in a different part of the country).
1 (state-based)	189	2	Events in the capital city area were separated from those in main conflict zone (located in a different part of the country).
1 (state-based)	650	2	Events took place in two separate parts of the country, located nearly 1000 km apart.
1 (state-based)	767	2	Events in the capital city area were separated from those in main conflict zone (located in a different part of the country).
2 (non-state)	4	2	Events took place mainly in and near urban locations, in various areas of the country. This dyad covers electoral violence.
2 (non-state)	5	2	Events took place mainly in and near urban locations, in various areas of the country. This dyad covers electoral violence.
2 (non-state)	79	2	Events in and near the capital city area were separated from those in main conflict zone (located in a different part of the country).

2 (non-state)	193	2	Events taking place in a large town (Dire Dawa), located at significant distance from the main conflict zone (Oromiya) were separated from those in main conflict zone
2 (non-state)	333	2	Events in and near the capital city area (Khartoum, Omdurman and Umm Dawm) were separated from those in main conflict zone (located in South Sudan).
2 (non-state)	372	3	Three distinct and not inter-related conflict zones were clearly reported in both period media and later accounts of the conflict. Each of these zones had distinct patterns and led to different outcomes of the conflict. The first zone is the town of Nouadhibou, Mauritania, the second zone the capital city of Mauritania, Nouakchott and its surroundings, the third zone is centered around two main cities of Senegal (Dakar and Touba), extending north to the Senegalese – Mauritanian border.
2 (non-state)	433	2	Events in and near the main city (Abidjan) were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	437	2	Events in and near the main city (Abidjan) were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	482	2	Events in and near the capital city area were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	483	3	Three distinct areas of violence against civilians can be distinguished in Chad – the first one, in the south of the country, in the area bordering the Central African Republic, near the towns of Doba and Mondou, the second one, in the Abeche region, bordering Sudan, and the third one, in the capital city, N'Djamena.
3 (one-sided)	540	2	Events in Cabinda are separated from events in metropolitan Angola.
3 (one-sided)	1380	2	Events in and near the capital city area were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	1524	2	Events in and near one of the country's important cities (Cape Town) area were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	1557	2	Events in and near the capital city area were separated from those in main conflict zone (located in a different part of the country).
3 (one-sided)	1578	4	Events in Uganda's capital city (Kampala), those in Puntaland and Somaliland were separated from those taking place in the main conflict area in Somalia.

21 dyads were split (**6** representing state-based conflicts, **7** representing non-state conflict and **8** representing one-sided violence) out of a total of **505** dyads for which polygons were constructed (**115** for state-based conflicts, **287** for non-state conflicts and **103** for one-sided violence).

Thus, **4.16%** of the total number of dyads was split, resulting in more than 1 polygon (**5.22%** of all state-based dyads, **2.44%** of all non-state dyads and **7.77%** of all one-sided violence dyads).

Appendix 2.

UCDP Definitions

This is a shortened, much abridged version of the UCDP Definitions available at <http://www.pcr.uu.se/research/ucdp/definitions/> . Note that some modifications specific solely to the GED project were made. These modifications are labeled below as *comments*.

Actor:

(state-based, non-state, one-sided)

A state or a non-state formally organised or organised group.

A formally organised group is any non-governmental group of people having announced a name for their group and using armed force.

An organised group is, in non-state conflicts only, any group who does not have an announced name, but who uses armed force and whose violent activity meets at least one of the following organisational requirements: there must be a clear pattern of incidents which are connected, or there must be evidence that violence was planned in advance.

State based Armed Conflict (state-based conflict, Armed-conflict):

(state-based)

An armed conflict is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year. A “state-based conflict” is referred to as opposed to “non-state conflict”, in which none of the warring parties is a government.

Non-state conflict

(non-state)

The use of armed force between two organised armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year.

One-sided violence

(one-sided)

The use of armed force by the government of a state or by a formally organised group against civilians which results in at least 25 deaths in a year. Extrajudicial killings in government facilities are excluded.

UCDP Dyad

(state-based, non-state, one-sided)

A dyad is made up of two armed and opposing actors. In state-based conflicts a dyad is defined as two actors, with one or more being the government, that have a stated incompatibility.

In a non-state conflict a dyad is constructed by at least two organised actors, of which none is the government of a state, that oppose each other with arms. In non-state conflicts it is possible for an

alliance of non-state actors to enter into a dyad with either an opposing group, or an alliance of opposing groups.

Comment specific to GED data:

For the purpose of the GED datasets, when studying one-sided event, armed actors (a government or a formally organised non-state group) attacking the civilian population are also referred to as 'dyads'. Other UCDP datasets refer to them as 'actors'.

Deaths

(state-based, non-state, one-sided):

For state-based violence and non-state violence it is the same as **battle-related deaths**:

Counted as battle-related deaths is armed conflict behaviour between warring parties in a conflict dyad, be it state-based or non-state. In state-based conflicts the violence must be directly related to the incompatibility, i.e. carried out with the purpose of realising the goal of the incompatibility and result in deaths. In non-state conflicts the violence does not have to be related to an incompatibility (since incompatibilities are not used in such conflicts), but has to take place between warring parties and result in deaths.

For one-sided violence, this is referred to the number of deaths inflicted by the one-sided actor on civilians.

Estimate, of deaths

(state based, non-state, one-sided)

Best estimate: the aggregated most reliable numbers for all incidents of organized violence during one event. If different reports provide different estimates, an examination is made as to what source is most reliable. If no such distinction can be made, UCDP as a rule includes the lower figure given.

Low estimate: the aggregated low estimates for all incidents of organized violence during one event. If different reports provide different estimates and a higher estimate is considered more or equally reliable, the low estimate is also reported if deemed reasonable.

High estimate: the aggregated high estimates for all incidents of organized violence during one event. If different reports provide different estimates and a lower estimate is considered more or equally reliable, the high estimate is also reported if deemed reasonable. If there are incidents when there is some uncertainty about which party has been involved, these may also be included in the high estimate

Dyad active

(state-based, non-state, one-sided)

State-based, and non-state dyads are active when violence between their constituent parts causes at least 25 battle-related deaths in one calendar year.

One-sided dyads are active when violence perpetrated by an actor against civilians causes at least 25 deaths in one calendar year.

Incompatibility

(state-based)

The stated (in writing or verbally) generally incompatible positions. As one country can experience several conflicts; we need a way to differentiate between them. The incompatibility criterion is only applicable to state-based conflicts and is not a prerequisite for non-state conflicts and one-sided violence. Incompatibility can be either over Government or Territory. There can only be one incompatibility over Government in a given year, but there can be several territorial conflicts. States and

parties may have several incompatibilities with several states simultaneously. This is not a problem; we are counting incompatibilities concerning at least one particular state, and it makes no difference whether the opposition organisations are unique to each situation or if one group roams across all states in the world. The primary entity is the state, then the incompatibility, and only after that the opposition organisation. Also, it should be noted that one military situation might include several incompatibilities.

Region

(state-based, non-state and one-sided)

In all UCDP datasets including the GED datasets and the data in the UCDP Conflict Encyclopedia (UCDP database) states are divided into five regions: Africa, Asia, Americas, Europe and Middle East.

Side

(state-based, non-state and one-sided)

In an intrastate conflict the government and its allies are always fighting on side A in a conflict, the opposition organisations and their allies are fighting on side B.

In an interstate conflict side A is the government that comes first in alphabetical order and its allies, side B is the other government and its allies.

In a non-state conflict side A is the organised group that comes first in alphabetical order, side B is the other actor.

In one-sided violence, there is only one actor. Side A thus designate the state or the formally organised group that target civilians.

State

(state-based, non-state and one-sided)

A state is either an internationally recognised sovereign government controlling a specified territory, or an internationally unrecognised government controlling a specified territory whose sovereignty is not disputed by another internationally recognised sovereign government previously controlling the same territory. Basically coincides with the list of UN member states, with the addition of a few non-members such as Taiwan. These are states both de jure and de facto. If we are dealing with a non-recognised entity, or a de facto state, it can meet the criteria of a state as defined here, if no other state claims that territory - given that the other state once used to control the entity.