

Dyadic Militarized Interstate Disputes (DYMID2.0) Dataset—Version 2.0¹

Zeev Maoz
University of California, Davis
Davis, CA 95616
zmaoz@ucdavis.edu

I. Rationale

The rationale for generating a dyadic version of the Militarized Interstate Dispute (MID) dataset (Bremer, Jones, and Singer, 1996) is based on the following observations:

1. There is a need for such a version in light of the growing number of studies on international conflict employing the dyadic unit of analysis, and utilizing the MID dataset.
2. The existing versions of the dataset (e.g., the version, which is downloadable through the PSSI Website or the EUGENE Website) consist of two types of observations: (a) *The dispute level*. This level includes general information about the dispute such as its start dates and end dates, the number of participants, its name, the highest level of hostility reached, the outcome of the dispute, etc. (b) *The individual participant level*. This level provides entry and exit dates for each participant, the side it took in the dispute, the highest level of hostility reached, and so forth. Extrapolation from the individual level to the dyadic level by simple computerized means may cause considerable errors. Some of these are enumerated below.
3. In strictly bilateral disputes, it is easy to transform the participation records for a specific dispute into a dyadic record, including the combination of data from the dispute profile record in the dispute dataset. However, performing such a combination on multilateral dispute may cause a great number of errors. Here are some examples of such errors.
 - *Non Valid Dyads*. Some of the dyads in a multilateral dispute may not be actually valid. For example, consider dispute #257, World War I. In this war, there are quite a few states that are listed on opposing sides of the dispute, but never threatened, displayed, let alone used force against each other. Japan, for example, is listed on Side B of this dispute and Bulgaria on side A. However, there is no evidence of a dispute existing between these two states, let alone a war. Another example concerns WWII (dispute #258). By the time Italy and Bulgaria entered the war, Poland had long been under German occupation, so there could be no dispute between Italy or Bulgaria and Poland.
 - *Inaccurate Levels of Hostility*. Take WWII for example (dispute #258). The US and Hungary participated in the war on opposing sides. Both reached the highest level of hostility (War). While both sides did declare war on each other, they did not fight each other. Taking the individual state level of hostility and plugging it directly to the dyadic level misses what happens at that level by a wide margin.

¹ I wish to thank Lesley Terris and Idin Eftekhari for invaluable assistance in preparing and cleaning the original dataset and its updated version.

- *Absurds: A state fighting itself.* France and Russia/Soviet Union are listed on both sides of the dispute in WWII. The former is actually listed twice on Side A (because it mobilized, declared war, and fought Germany between March, 1939 and July 1940), and continued to fight Germany after its liberation in 1944. However, it also did fight the British under the Vichy regime. The later state started the War fighting against Poland in 1939 side by side Germany (in accordance with the secret clauses of the Ribbentrop-Molotov agreement), and ended fighting Germany and its allies after Barbarossa in June 1941. A simple cut and paste from the individual file to a dyadic version would have these states fighting themselves.
 - *Dates of dispute.* Entry and exist of states in multilateral disputes may differ from one dyad to another. State A may enter a dispute at one point in time against state B, and at another point in time against state C. Taking the earliest entry date of the state and assigning it to all of its dyadic combinations in a given dispute may distort the actual duration of the dispute.
 - *Distorted Dispute Outcomes.* A nice example can be given from BdM's *War Trap*. In the appendix, BdM lists all interstate wars and major power disputes with their outcomes. WWII is listed as a war between Poland and Germany with Poland declared the winner. Sounds Funny? Not if we take this straight out of *Resort to Arms* with Poland listed on the winning side of WWII. The same applies to a multitude of other multilateral disputes.
 - *States Switching Sides in Wars/Disputes.* Related to the point of states fighting themselves, we have several cases (nine to be precise in the MID2.10 dataset) in which sides are listed twice in disputes. Typically, these denote situations wherein states switched sides in disputes. In dyadic terms, it implies that old dyads ended at some point of an ongoing disputes (e.g., when the Soviet Union clashed with UK and France in 1939 over its invasion of eastern Poland and the Baltic States, this represents one set of dyadic disputes, and when the Soviet Union fought alongside UK against the Axis as of 1941, this represents another set of dyadic disputes—both sets taking place within the same war).
4. Beyond these problems of extrapolation, there is a need to redefine certain variables at the dyadic level. For example, for some analyses, it makes sense to have an annual record of ongoing MIDs, while for others it is only necessary to look at the start year of the dispute. The minimum and maximum duration variables that are listed at the dispute level are not very useful when it comes to dyadic analyses, as each dyad may have a different duration from each other dyad within the same dispute. Likewise, the outcome and settlement of the dispute may not apply to all dyads.

II. Structure of the Dyadic dataset

1. *Records:* Each record constitutes a dyadic dispute year. This implies the following things:
 - a. *Short Bilateral Disputes:* If a dispute had only one state on each side and if it started and ended within the same calendar year, it gets only one record.

- b. *Long Bilateral Disputes*: If a dispute had only one state on each side, but started in a given year and ended in another, it will get a record for each year the dispute was underway. Thus, if a dispute started on December 21, 1838 and ended on January 12, 1841, it will get four records: 1838, 1839, 1840, and 1841. Two variables allow tracing the number of records (or years) for a dispute: the YEAR variable, and the DURINDX variable. They are described below.
 - c. *Short Multilateral Disputes*: If a dispute had more than one state on a given side of the dispute, but started and ended within the same year, it gets one record for every *valid* dyad in the dispute. If the dispute had m initiators and n targets, it may—but does not necessarily—have as many as $m \times n$ records. We attempted to include only dyads for which there is evidence of actual exchange of military hostility acts.
 - d. *Long Multilateral Disputes*. Dispute with multiple participants on one or both sides that started and ended in different years get one record per valid dyad per calendar year of dispute. Records are based on actual entry and exit of individual states into the dispute. The start date for a given dyadic dispute within a long multilateral dispute is the last date of entry of the states comprising the dyad. The end date is the first date of exit of the states comprising the dyad. Suppose a dispute had two states (i and j) on side A and two states (k and l) on side B, and covered three years from the first to the last incident. Suppose that state j on side A entered a year after the dispute broke out, and state l on side B left a year before the dispute ended. Thus we have the following dyads: (1) i and k , three records, one for each year; (2) i and l , two records, one for the first year and one for the second year; (3) j and k , two records, one for the second year and one for the third year; (4) j and l , one record—for the second year only.
2. *Levels of hostility and highest military confrontation action (MCA)*. These are adjusted such that we have the specific level of hostility and highest MCA in a specific dyad as directed by states towards each other. Thus if state A fought a war against state D but states B and C participated in the disputes only at levels of mobilization (show of force), then the AD dyad will have war as its MCA, but AC, BD, and BC will have mobilization as the highest MCA. If there is evidence that A actually used force against C then the AC dyad will have use of (troops, ships, planes) as the highest MCA. In addition, each state's highest MCA is also specified.
 3. *Fatality Levels*. Adjusted to reflect the actual level of hostility of the dyad. Again, due to considerable missing information, this variable should be used with great deal of caution.
 4. *Dyadic adjustment of outcomes and settlements*. In some cases, where we had information about specific outcomes of dyadic disputes, which were part of multilateral disputes, we fit these variables to characterize the dyadic outcome and dyadic form of settlement. This coding is highly incomplete, however, and should be used with caution.
 5. *New Dyadic Disputes in case of switching*. As noted above, states who switched sides in the course of a dispute alter their dyadic participation. Suppose states A, D, and F

fought against states B and C, in the course of a multilateral dispute. Suppose there is evidence of actual exchange of hostility between all possible dyads. Thus, the dispute starts with dyads AB, AC, DB, DC, FC, FC. Suppose that in the midst of the dispute state A switched to the BC coalition. Thus, at the point of switching, the dyads AB, AC, are determined as terminated, and a new set of dyads is added (AD, AF). However, the latter set of dyads is now defined as a new dispute. It gets a new number. This applies almost exclusively to WWII.

6. *Changes in coding of some dispute records from MID2.10 to DYMID1.1.* As a result of research done to convert the MID2.10 to DYMID1.1, and—more importantly—as a result of research following comments by colleagues,² the coding of some disputes has changed. You may not find a one-to-one correspondence between the dyadic disputes listed herein and the MID2.10 dataset. For details, see the enclosed comments file.
7. *New Variables.* I added several new variables that were necessitated by the coding of data at the dyadic level, and generation of multiple records per dispute. These are listed and explained below.
8. *Download the file:* Click the file name to download it ([dyadmid1.1](#)). This is an Excel 2000 file. If you have problems reading it send email to lesley@post.tau.ac.il.

II. Dyadic MID Codebook

Following is the variable list and origin of the variables (MID3.02 is the original value of the variable in the MID 3.02 version; DYMID1.1 is an indication that this variable has been changed in some cases here.

Variable	Variable Label	Version	Value Labels	Comments
DISNO	Dispute Number	MID3.02	—	
STATEA	COW Number of state A in dyad	MID3.02	—	
NAMEA	Abbreviated name of state A in dyad	MID3.02	—	
STATEB	COW Number of state B in dyad	MID3.02	—	
NAMEB	Abbreviated name of state B in dyad	MID3.02	—	

² I wish to thank Sturart Bremer, Paul Hensel, Jim Ray, Scott Bennett, and Mike Ward for their comments and suggestions. Responsibility for final coding decisions is mine alone. All comments and coding decisions are given in the [midcomment.doc](#) file.

Variable	Variable Label	Version	Value Labels	Comments
STRTDAY	Start day of dyadic dispute	DYMID1.1	—	Adjusted for dyad (last day of entry of one of the states)
STRTMNTH	Start month of dyadic dispute	DYMID1.1	—	Same as above
STRTYR	Start year of dyadic dispute	DYMID1.1	—	Same as above
YEAR	Year of dispute	DYMID1.1	—	Year dispute is underway (differs from strtyr for years after first year; differs from endyear for years before last year)
OUTCOME	Outcome of dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Changed to reflect the actual dyadic outcome. Listed only for last year of dispute.
SETTLMNT	Settlement type of dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Changed to reflect the actual dyadic settlement. Listed only for last year of dispute.
FATLEV	Fatality level of dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Changed to reflect the actual fatality level. Listed only for last year of dispute. Use with caution.
HIGHACT	Highest military confrontation action in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Reflects highest MCA reached by members of dyad toward each other.
HIHOST	Highest level of hostility in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Adjusted as above for MCAs.
RECIP	Reciprocated dispute?	MID3.02	-1: Yes 0: No	Not adjusted, may not reflect actual dyadic dis-

Variable	Variable Label	Version	Value Labels	Comments
				pute.
NOINIT	Number of states on Side A	DYMID1.1	—	Adjusted to reflect number of valid actors on side A. May not reflect actual valid dyads
NOTARG	Number of states on side B	DYMID1.1	—	Adjusted to reflect number of valid actors on side B. May not reflect actual valid dyads
STDAYA	Start day state A	MID3.02	—	
STMNTHA	Start month state A	MID3.02	—	
STYEARA	Start year state A	MID3.02	—	
ENDDAYA	End day state A	MID3.02	—	
ENDMNTHA	End month state A	MID3.02	—	
ENDYEARA	End year state A	MID3.02	—	
SIDEAA	State A on initiator's side?	MID3.02	1: Yes 0: No	
REVSTATA	State A revisionist?	MID3.02	1: Yes 0: No	
REVTYPEA	Type of revision sought by state A	MID3.02	See file format.txt in MID 2.1 dataset package	May not fit actual dyad. Use with extreme caution.
FATLEVA	Level of fatality incurred by state A in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Reflects actual dyadic dispute. Missing data. Use with caution. Listed only for last year of dyadic dispute.
HIGHMCAA	Highest MCA initiated by state A	DYMID1.1	See file format.txt in MID 2.1 dataset	Reflects actual MCA initiated by state A towards state B in dyadic dispute. Use with caution. See

Variable	Variable Label	Version	Value Labels	Comments
			package	problem file in this package.
HIGHOSTA	Highest level of hostility reached by state A in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Reflects actual HIHOST initiated by state A toward state B in dyadic dispute. Use with caution. See problem file in this package.
ORIGNATA	State A originator (initiator) of dispute?	MID3.02	-1: Yes 0: No	
STDAYB	Start day state B	MID3.02	—	
STMNTHB	Start month state B	MID3.02	—	
STYEARB	Start year state B	MID3.02	—	
ENDDAYB	End day state B	MID3.02	—	
ENDMNTHB	End month state B	MID3.02	—	
ENDYEARB	End year state B	MID3.02	—	
SIDEAB	State B on initiator's side?	MID3.02	1: Yes 0: No	
REVSTATB	State B revisionist?	MID3.02	1: Yes 0: No	
REVTYPEA	Type of revision sought by state A	MID3.02	See file format.txt in MID 2.1 dataset package	
FATLEVB	Level of fatality incurred by state B in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Same as FATLEVA
HIGHMCAB	Highest MCA	DYMID1.1	See file	Same as HIGHMCAA

Variable	Variable Label	Version	Value Labels	Comments
	initiated by state B		format.txt in MID 2.1 dataset package	
HIGHOSTB	Highest level of hostility reached by state B in dyadic dispute	DYMID1.1	See file format.txt in MID 2.1 dataset package	Same as HIGHOSTB
ORIGNATB	State B originator (initiator) of dispute?	MID3.02	-1: Yes 0: No	
ROLEA	Role of state A in dyadic dispute	DYMID1.1	1: Primary Initiator 2: Joiner on initiator side 3: Primary target 4: Joiner on target side	See problem file in this package.
ROLEB	Role of state B in dyadic dispute	DYMID1.1	1: Primary Initiator 2: Joiner on initiator side 3: Primary target 4: Joiner on target side	
WAR	Is this dispute a COW war dyad	DYMID1.1	1: Yes 0: No	Adjusted to reflect actual warring dyads in multilateral disputes.
DURINDX	Duration index	DYMID1.1	1: First year of dispute 2: Second year of dispute...	Those who want to use only dispute outbreaks should select durindx=1
DURATION	No. of dispute days during year	DYMID1.1		For disputes crossing a calendar year, each record reflects no. of dispute days during the

Variable	Variable Label	Version	Value Labels	Comments
				present calendar year.
CUMDURAT		DYMID2.0		Cumulative number of days from the start date of the MID to the present (either the end of the year of MIDs crossing over to next year or end date—if in present year)
RECIPROC	Two-sided confrontation	DYMID1.1	0: Only one side 1: Both sides	Denotes “reciprocated” disputes, that is differentiates between disputes in which only one side threatened, displayed, or used force, and disputes in which both sides committed actual hostile actions.

9. *Problems.* There are quite a few problems in some of the variables and disputes that we were unable to resolve at the present time. We enclose a file of the problems we detected. This file [Problems.xls](#) lists the disputes with their associated problems. There are numerous other problems we have probably not detected and we encourage you to send us any observations and critiques you have or any errors you detect in the dataset. In addition, if anyone has any information (preferably the original code sheets) for the disputes listed in the **PROBLEMS.XLS** file, please send us mail.

10. Final word of caution. As with any kind of dataset, and given the historical and geopolitical scope of the MID dataset, problems, errors, omissions, and other queries are bound to exist. Data development is an interactive process. We encourage users of this dataset to: (a) look and feel the data, (b) examine them critically, (c) use with caution, and (d) report any issues and problems that they find. We—at our end—promise to look at these things and fix the data if necessary. As always, we are committed to transparency of data development and freedom of use of these data.